10-8-2024

Genomic analysis of virulence loci that enable growth of Plasmopara viticola on resistant Vitis spp.

Sarah Teeter
srt4632@rit.edu

RIT

Genomic analysis of virulence loci that enable growth of *Plasmopara viticola* on resistant *Vitis* spp.

by

Sarah Teeter

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Bioinformatics

Thomas H. Gosnell School of Life Sciences
College of Science

Rochester Institute of Technology
Rochester, NY
October 8, 2024

Table of Contents

Rochester Institute of Technology
Thomas H. Gosnell School of Life Sciences
Bioinformatics Program

To: Head, Thomas H. Gosnell School of Life Sciences

The undersigned state that Sarah Teeter, a candidate for the Master of Science degree in Bioinformatics, has submitted her thesis and has satisfactorily defended it.

This completes the requirements for the Master of Science degree in Bioinformatics at Rochester Institute of Technology.

Thesis committee members:

Name	**Date**
Michael V. Osier, Ph.D. Thesis Advisor	
Eli J. Borrego, Ph.D.	
Lance Cadle-Davidson, Ph.D.	

Acknowledgements

I would like to thank my thesis committee, Dr. Michael Osier, Dr. Lance Cadle-Davidson, and Dr. Eli Borrego, for their invaluable knowledge of bioinformatics and plant biology. I wish to also thank them for their insightful comments and constructive criticism, which greatly improved the quality of this thesis. An additional thank you goes to Anna Underhill, Kas Deys, Cheng Zou, and the entire Cadle-Davidson team at the USDA at Cornell Agritech for helping with wet lab preparations and guidance. And finally, a massive thank you to the USDA at Cornell Agritech for providing me with the funds and means to carry out this research.

Competing Interests

The writer claims no conflict of interest.

Abstract

Plasmopara viticola is considered one of the most damaging grapevine pathogens, costing growers millions of dollars in losses. Despite the discovery of resistance to *P. viticola* (RPV) loci in grapevines to combat this pathogen, recent research shows that *P. viticola* can rapidly adapt virulence on some RPV loci, reducing their effectiveness. This thesis explored two strategies to understand how *P. viticola* adapted to two resistance loci of moderate strength, RPV3.1 and RPV10.3, by identification and characterization of *P. viticola* genomic regions showing a high selective sweep. Using sequenced natural infections from field samples, multiple tools and custom R scripts were applied to detect high selective sweep signals, objectively select regions of interest, and determine what genes and proteins are present. This pipeline of analysis revealed distinct sweep regions and patterns across the RPV resistant samples tested. While no selective sweep was detected for *P. viticola* collected from vines with only RPV3.1, *P. viticola* from RPV10.3 showed a strong selective sweep on scaffold 6, spanning 42 genes. These genes included a homologue of a known RxLR effector protein critical in host-pathogen interactions (PVIT_0003606) and a hypothetical protein predicted by SignalP to be secreted (PVIT_0003584). Similarly, vines containing both RPV3.1 and RPV10.3 showed strong selective sweeps on scaffolds 3, 11, and 6. The strong selective sweep on scaffold 3 spanned 21 genes, including two homologues of known avirulence proteins (PVIT_0002214 and PVIT_0002215). The sweep on scaffold 11 corresponded with

another hypothetical protein predicted by SignalP to be secreted (PVIT_0006058). Comparative genomic analysis of scaffold 6 between samples revealed that RPV10.3 and RPV3.1+10.3 grown samples contain the same genes. This substantial overlap suggests a conserved selective sweep on scaffold 6 when RPV10.3 is present in the host vine. The findings suggest regions within *P. viticola* that are responsible for overcoming resistance and how these differ between RPVs as well as providing insight into possible future evolutionary responses of *P. viticola.*

Introduction

Effects of Downy Mildew on the Grape Industry

Considered one of the most damaging grapevine pathogens, *P. viticola*, an oomycete that is notorious for causing grapevine downy mildew, can result in yield losses of up to 75% in grapevines during humid conditions. German vineyards saw a reduction in output by nearly 33% from 1907 to 1916 while French vineyards saw 70% of grapevines destroyed in 1915. (Koledenkova et al., 2022). In the 1960s, Texas saw $2.5 million in losses and up to 90% in field incidence. Then in 1979, Canada and northeastern U.S growers lost $200 million (Koledenkova et al., 2022). All these losses were attributed to *P. viticola*, and despite the creation of the Bordeaux mixture in 1885 that harnessed the antifungal properties of copper sulfate, *P. viticola* continues to have a significant economic impact. Fungicides designed to manage this pathogen account for 54% of the $1.2 billion global fungicide market specifically targeting oomycete pathogens (Taylor and Cook, 2018).

Chemical Fungicide Use

Management of *P. viticola* is reliant on fungicide that is applied at 7-10 day intervals if the weather is wet and 10-14 day intervals if the weather is dry from mid-April to mid-September. Specifically, growers in Georgia are estimated to apply 17 applications per growing season (Campbell et al., 2021). The frequency of these applications provides selection pressure on *P. viticola,* which leads to an increase in chance of fungicide resistance within the pathogen population (Campbell et al., 2021). Fungicides such as mancozeb and captan and newer active

ingredients such as azoxystrobin and mandipropamid are all used. However, while mancozeb and captan remain relatively effective against *P. viticola,* single-site mode of action fungicides such as quinone outside inhibitor (QoI) (azoxystrobin is an example) rapidly become ineffective. Resistance to QoI fungicides emerged after four years on the market in both France and Italy. Additionally, pyraclostrobin poses a high risk for development of resistance (Campbell et al., 2021).

History and Origins

The cultivated grapevine (*Vitis vinifera*) emerged from the domestication of the wild Eurasian species *V. sylvestris* in Transcaucasia. The oomycete attributed to causing downy mildew in grapes, *Plasmopara viticola* was first collected in northeastern America in 1834 and subsequently observed for the first time in Bordeaux, France in 1878. Due to this lack of co-evolution, *P. viticola* has threatened viticulture since the middle of the 19th century (Rouxel et al., 2012). It is believed that this introduction can be attributed to the French grafting vines onto American rootstocks to prevent phylloxera, a grapevine pest of American origin that appeared in Europe. From France, the pathogen spread to Northern Italy in 1879 and the Mosel region of Germany in 1880 (Gobbin et al., 2007). *P. viticola* crossed over the Mediterranean to Algeria in 1881(Taylor et al., 2019). The pathogen continued its trajectory and spread to Eastern Europe such as Turkey and Greece around 1887 and South Africa around 1907. Continuing from Eastern Europe, the pathogen reached Australia around 1917, New Zealand around 1926, and Tasmania in 1959 (Taylor et al., 2019) as is shown in Figure 1C. From Australia, the pathogen arrived in Argentina in 1920 (Fontaine et al., 2021). Most recently, *P. viticola* was detected in South Korea

in 2018 (Kim et al., 2019) and the Russian Far East in 2022 (Nityagovsky et al., 2024). Figure 1A shows the overall movements while Figure 1B and 1C shown more specific movements within Australia and Europe as described previously. After Fontaine et al. (2013) analyzed spatial pattern of genetic diversity in *P. viticola*, they concluded that the European *P. viticola* populations were less diverse than the North American populations, which suggested a founder effect during the introduction of this pathogen to Europe. As of recent, cryptic-specialized species have been identified in North American populations of *P. viticola*, perhaps due to co-evolution of clades on North American resistant wild species.

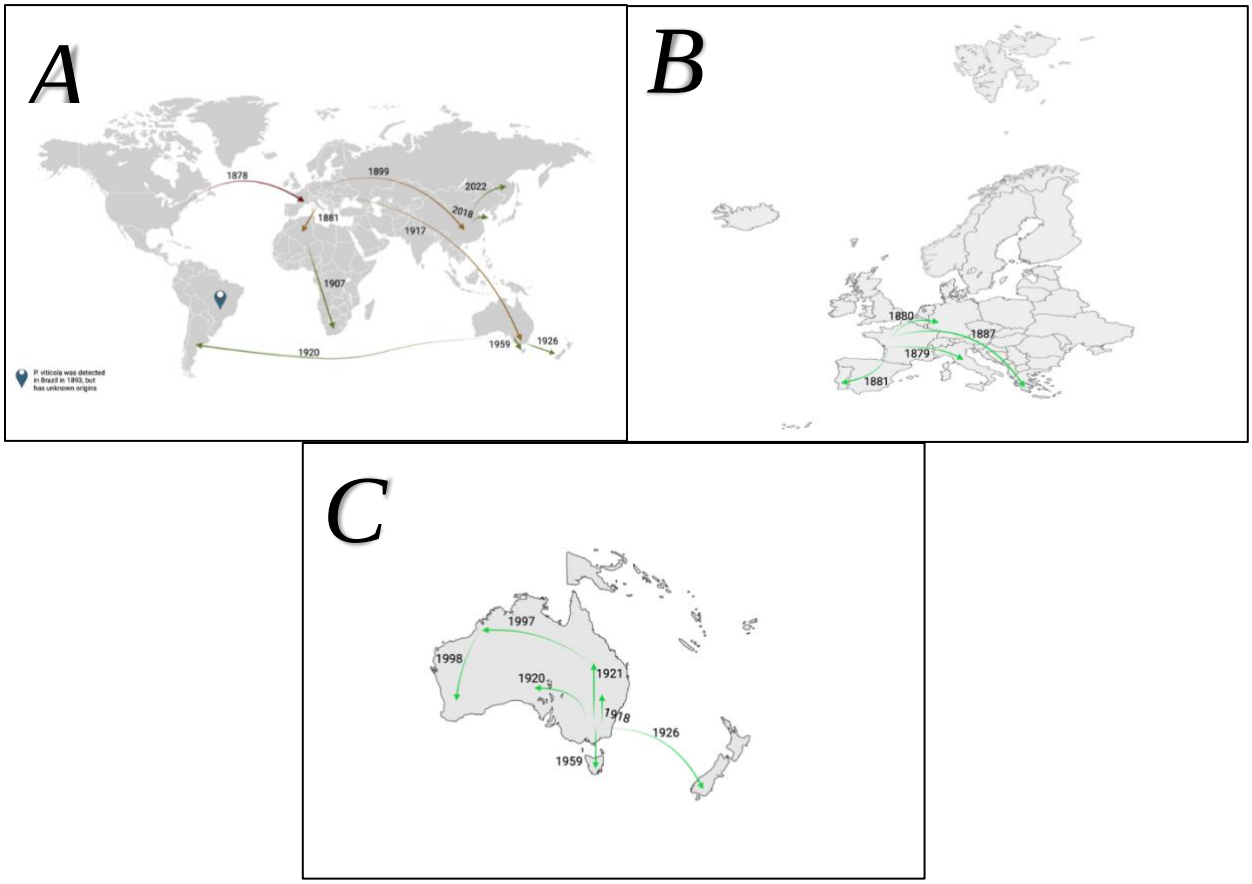

Figure 1: The historical geographical trajectory of *Plasmopara viticola*, based on reports of grapevine downy mildew. A) This is a concise overview of detection of *P. viticola* starting in the United States, moving to Europe, and subsequently to Africa, Asia, and Australia. B) A more detailed map of the path throughout Europe. C) A more detailed map of the path through Australia starting in Rutherglen, Victoria and subsequently moving to New Zealand, Tasmania, and westward. The arrows show time sequence of detection, and do not imply movement from point A directly to point B. (This figure was created with BioRender.com).

Cryptic Species

Cryptic species refer to species that are not clearly distinguishable morphologically, resulting in two or more taxa grouped under a single name. These species are useful in providing novel insight into patterns and processes of biodiversity (Wei et al., 2022). There are currently five identified *P. viticola* cryptic species in North America. They are *P. viticola* f. sp. *riparia (PvR), P. viticola* f. sp. *aestivalis (PvA), P. viticola* f. sp. *vinifera (PvV), P. viticola* f. sp. *quinquefolia (PvQ),* and *P. viticola* f. sp. *vulpina (PvVU).* While both *PvA and PvV* infect *Vitis vinifera, PvA* is more abundant in vineyards that contain *V. vinifera* than *PvV* (Rouxel et al., 2014). Previously Rouxel et al. (2012) concluded the host range expansion of *P. viticola* from wild species to cultivated species. They determined *PvR* began in *V. riparia* and expanded to a hybrid of Chancellor; *PvA* began in *V. aestivalis* and expanded to *V. labrusca* 'Niagara' and *V. vinifera*; *PvV* has unclear wild species origins but has since expanded to *V. vinifera*; and finally, *PvQ* began in *Parthenocissus quinquefolia* and has yet to expand to a cultivated host (Rouxel et al. 2012). *PvVU* was the most recently determined cryptic species and has only been observed in *V. vulpina,* which is its original host.

Polymorphic Sites (SNPs) for Cryptic Species Identification

To distinguish between cryptic species, certain single nucleotide polymorphisms (SNP) have been confirmed as having the ability to differentiate between the species (Rouxel et al., 2012). *PvR* is the most divergent taxa for each of these sequences, with 36 of the 67 of the polymorphisms (53.7%) being unique to the *PvR* cryptic species. These SNPs can be visualized in Table 1 to see the diagnostic nucleotide change. This information was gathered by doing a

multi-sequence alignment using datasets provided by Rouxel et al. *PvVU* only had strains present for the ITS region, and therefore, only has the specified SNPs in that region.

ITS region

Location (bp)	37	38	53	56	70	75	94	102	172	173
riparia	T	A	C	T	T	C	T	T	T	G
aestivalis	C	T	T	C	G	C	T	T	C	G
vinifera	C	T	T	C	T	C	A	A	C	G
quinquefolia	C	T	T	C	T	T	T	T	C	G
vulpina	C	T	T	C	T	C	T	T	C	A

β-tubulin region

Location	1	22	30	33	36	69	78	84	90	91	111	114	153	154	162	171	174	175	183
riparia	G	C	C	G	T	T	T	T	C	T	T	T	G	T	A	G	A	C	T
aestivalis	A	T	T	A	C	T	C	C	C	T	C	C	G	C	G	A	A	T	T
vinifera	A	C	T	A	C	A	C	C	C	T	C	C	G	T	G	A	A	T	C
quinquefolia	A	C	T	A	A	T	C	C	T	C	C	C	A	T	G	A	G	T	T

β-tubulin region (continued)

Location	216	246	273	288	297	345	411	414	417	471
riparia	C	T	A	T	A	T	T/C	T	G	C
aestivalis	T	C	T	C	G	C	A	T	T	T
vinifera	T	T	A	C	G	C	A	T	G	T
quinquefolia	T	T	A	C	G	C	G	C	G	T

Table 1: The SNPs identified between the different cryptic species for four different regions: Internal Transcribed (ITS) region, 28S region, β-tubulin region, and actin region.

28S region

Location	58	76	132	134	137	200	374	377	408	443	457	606
riparia	A	A	A	C	T	T	C	G	G	G	C	T
aestivalis	A	G	G	C	C	T	T	A	T	T	G	G
vinifera	G	G	A	C	T	C	C	A	T	C	A	G
quinquefolia	A	G	A	T	T	T	C	A	T	T	A	A

Actin region

Location	15	33	51	60	67	68	117	150	174	183	207	216	217	243	267	279	333	378
riparia	C	C	T	A	T	C	A	T	T	A/C	A/G	G	T	T	C/T	T	G	C
aestivalis	A	T	G	G	T	C	G	C	G	T	T	G	C	C	G	G	T	A
vinifera	A	T	T	G	T	C	G	T	G	T	C	G	T	T	G	T	T	A
quinquefolia	A	T	T	G	A	G	G	T	G	T	T	A	T	T	G	T	T	A

Table 1 (continued)

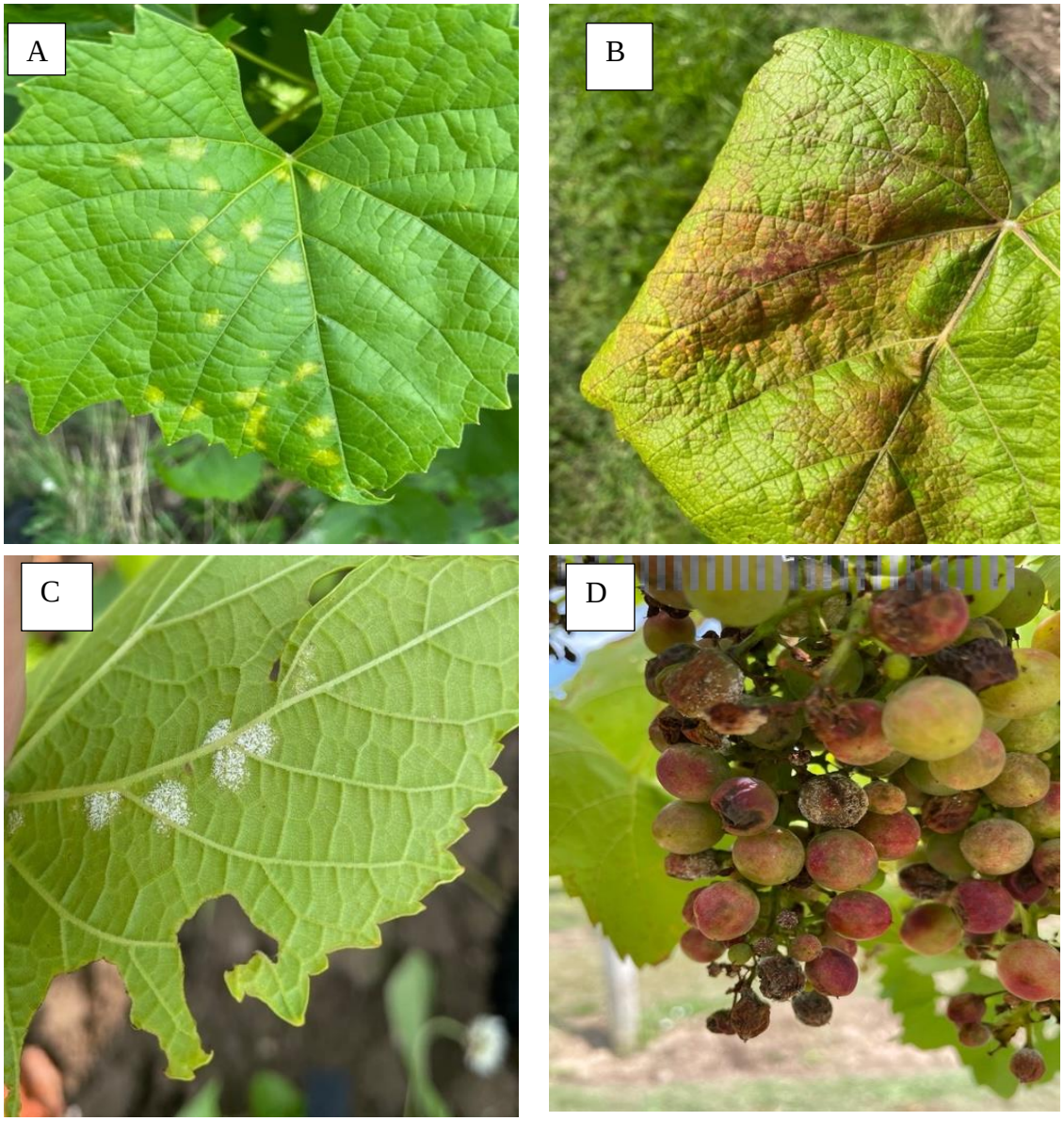

Figure 2: Symptoms and signs of grapevine downy mildew on leaves. A) The beginning stage of *P. viticola*. The oil spots appear on the adaxial side of a leaf of a *Vitis amurensis* hybrid. B) Another beginning stage of *P viticola*, but the oil spot appears a dark reddish brown. This leaf is from the hybrid 'Steuben'. C) The sporulation of *P. viticola* on the abaxial side of a *V. amurensis* leaf. After optimal conditions are met, the oomycete begins to sporulate in fluffy white lesions. D) The white berries were infected with *P. viticola* soon after flowering. Several discolored and shriveled berries were infected subsequently, but too late for sporulation, as their stomata converted to lenticels. Some of the berries were not infected and have developed resistance as they mature.

Symptoms of Grapevine Downy Mildew

The first symptoms of *P. viticola* typically appear 3 to 4 days after infection but can be extended to 5 to 7 days depending on weather conditions and susceptibility of the host. The

initial symptoms are yellow “oil” spots that appear on the adaxial side of the leaf as shown in Figure 2(A). Young oil spots may be surrounded by a thin, brown halo that disappears as the spot matures. These oil spots can also present as a reddish brown on certain darker grape varieties as shown in Figure 2(B). On the abaxial side of the leaf, directly underneath the oil spots, white sporangia will appear in lesions as shown in Figure 2(C). Although the leaves are the most common area of infection, the berries may also become infected. The young berries, similar to the leaves, will present the white sporangia before becoming discolored and necrotic as shown in Figure 2(D) (Jones and McManus, 2017). These fruit infections are particularly devastating to the yield and quality of grape harvest. However, as the berries mature, they develop ontogenic resistance mostly likely caused by the non-functionality of the stomata after their conversion to lenticels (Kennelly et al., 2005). Berries typically become fully resistant within 2 to 3 weeks after blooming. The stem of the grapevine may also be infected and remains susceptible up to two months after bloom (Jones and McManus, 2017).

Life Cycle of P. viticola

P. viticola is an obligately, biotrophic oomycete organism that causes downy mildew on grapevines. *P. viticola* is a polycyclic pathogen that causes both primary and secondary infections (Maddalena, Russo, and Toffolatti, 2021. The oospores are considered the main survival structure of this pathogen. They are large (25-50 µm diameter) thick-walled sexual spores that form within leaf tissues. The process of infection begins with the overwintering of oospores on the soil surface and fallen leaves from the previous growing season, which produces an inoculum for the following growing season. At the tip of the germ tube, the oospore differentiates a macrosporangium where the infection spores, zoospores, are formed (Maddalena,

Russo, and Toffolatti, 2021). The inoculum is dispersed by wind and rain, with free water allowing the swimming zoospores to infect susceptible grapevine tissue through the stomata (Brischetto et al., 2021). The joining of the sexual gametangia (oogonium and antheridium) forms the oospores, and a single antheridial nucleus passes through a fertilization tube into the oosphere. The new oospore with then begin maturing. As this process is active, the antheridial and oogonial nuclei fuse and the multi-layered cell wall increases in thickness for protection (Vercesi et al., 1999). Studies have shown that the maturation period is completed over variable time-period, relying most on the time required by the oospores to reach the peak in the germination percentage at the optimal temperature of 20 ˚C. For most of the population, this process will take up to 5 months to complete, yet many oospores do not complete this process and die (Vercesi et al., 1999). Secondary infections are caused by sporangia produced on *P. viticola* lesions. Under favorable weather conditions, the pathogen emerges from the stomata and forms sporangia-bearing sporangiophores (Brischetto et al., 2021). Ideal conditions for sporangial sporulation are as follows: darkness (as sporulation does not occur in sunlight), a minimum of 4 hours with ≥ 98% relative humidity, and ≥ 19˚C ambient temperature (Brischetto et al., 2021). The number of sporangia produced by a lesion will depend on the weather conditions and the lesion age. Once matured, the sporangia will detach themselves from sporangiophores and disperse through rainfall/water droplets and wind (Turrà and Pietro, 2015). A concise life cycle is shown in Figure 3.

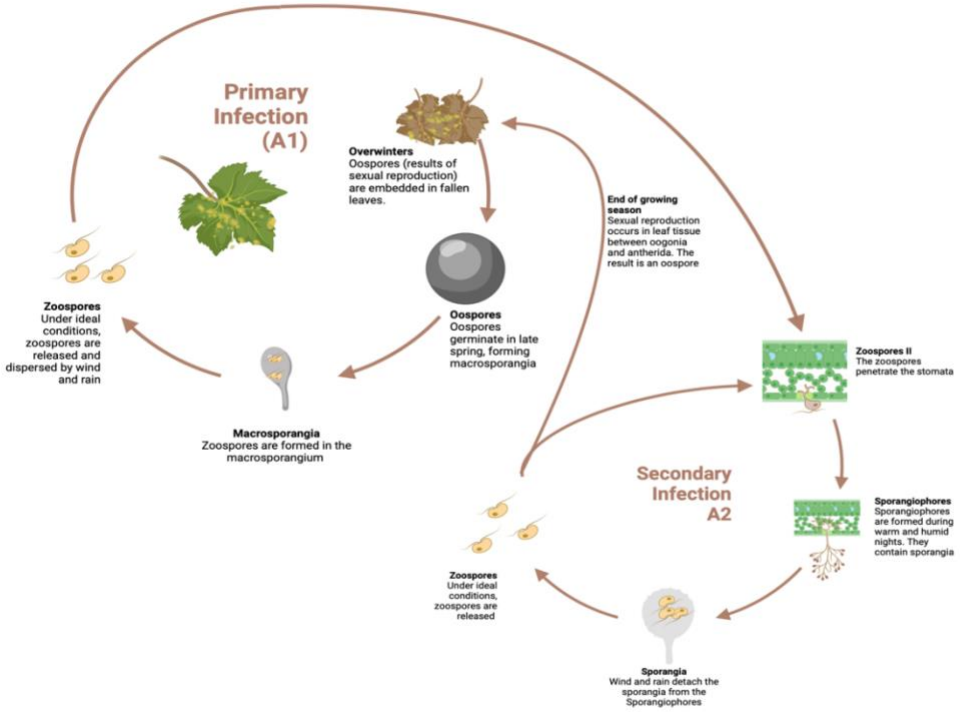

Figure 3: The basic life cycle of *P. viticola.* The primary infection involves oospores, the products of sexual reproduction, which occurs only once in the pathogen's lifecycle. The secondary infections happen many times through the growing season (summer) through asexual reproduction. (This figure was created with BioRender.com).

Host Resistance

Quantitative trait locus (QTL) mapping involves the division of progeny into groups according to inherited genotypes and phenotypes, and then applying a comparison to identify if there is a significant association between the traits and the allelic variants. As of August 2022, 33 loci for resistance to *P. viticola* have been identified (Possamai and Wiedermann-Merdinoglu, 2022) in *Vitis* spp: Resistance to *P. viticola* (RPV)1 and RPV2 in *V. rotundifolia;* RPV3.1 and RPV19 in *V. rupestris;* RPV4, RPV7, RPV11, RPV17, RPV18, RPV20, and RPV21 in an unspecified American species; RPV5, RPV6, RPV9, and RPV13 in *V. riparia;* RPV8, RPV10.3, RPV12, RPV22, RPV23, RPV24, RPV24, and RPV26 in *V. amurensis;* RPV14 in *V. cinerea;* RPV15 and RPV16 in *V. piasezkii;* and RPV27 in *V. aestivalis* (Sargolzaei et al., 2020). RPV1 is

found on chromosome 12 and is a nuclear-binding site of leucine-rich receptors (NB-LRR). RPV3.1 is found on the distal part of chromosome 18 (Bellin et al. 2009) and associated with the biosynthesis of stilbenes. RPV10 locus originated from the Asian grape species *V. amurensis* and has been mapped to chromosome 9 (Wingerter et al. 2021), however, the mechanism for this locus remains under characterized.

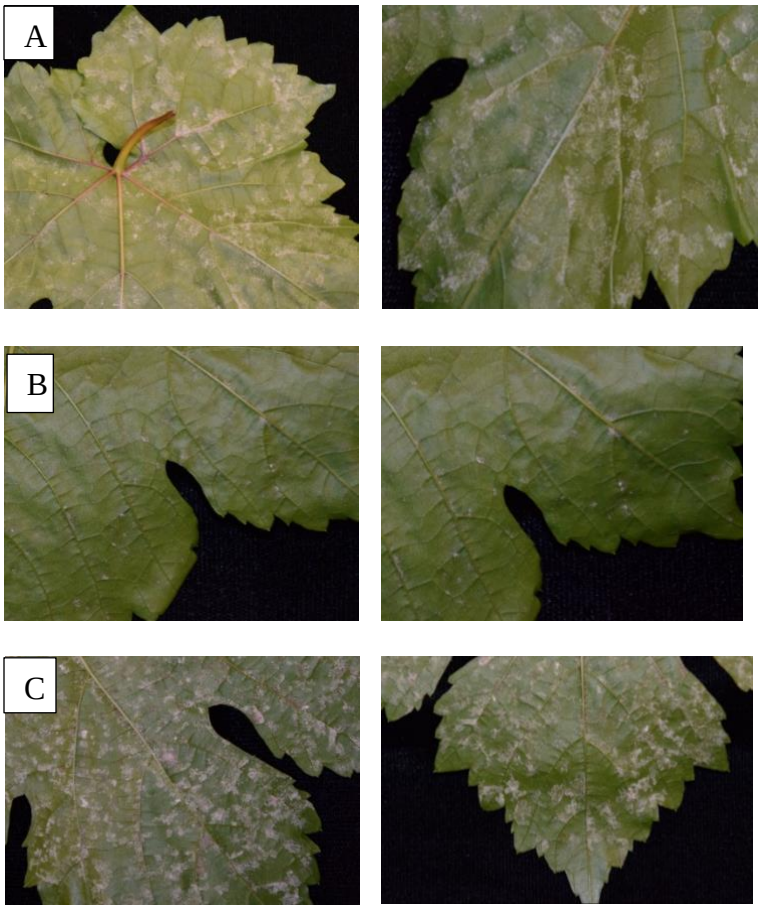

Figure 4: Images of *P. viticola* on leaves containing different RPVs. For (A) left and right, this leaf contains no known RPVs. For (B) left and right, this leaf contains RPV3.1. For (C) left and right, this leaf contains RPV10.3. For (D) left and right, this leaf contains both RPV3.1 and RPV10.3.

Figure 4 (continued)

The RPVs within the leaves provide quantitative variation in the disease severity phenotype of the grape to *P. viticola* (Figure 4). Genetic variability in unknown loci also impacts disease severity. For a leaf that has no known RPVs, the sporangia are clustered together. There are densely populated sporulating lesions and less densely sporulating lesions interspersed across the leaf. For a leaf that contains RPV3.1, there is a distinct lack of clustering. The few sporangia present are dispersed across the leaf instead. For a leaf containing RPV10.3, the sporangia are densely cluster together to the point of clusters combining with each other to form one large patch of sporangia. Finally, a leaf that contains RPV3.1 and RPV10.3 shows few sporangia dispersed across the leaf as well as large patches of the sporangia clustering together. Because RPV3.1 and RPV10.3 quantitatively shift the mean severity toward more resistance amid the variable genetic background of other undetected loci impacting disease severity, substantial variation can be seen among resistance phenotypes give the presence or absence of each resistance locus.

Preliminary Suppositions and Implications (Hypothesis)

Given their mixed mode of reproduction (primarily clonal with a single sexual cycle per year), *P. viticola* strains that are able to grow on a resistant grapevine will show a selective sweep and the resulting increased linkage disequilibrium (fixed alleles) at one locus responsible

for virulence on that resistant grapevine and random assortment of alleles on other chromosomes and away from the virulence locus. Different host resistance genes will cause selective sweeps around different virulence loci.

Materials and Methods

Overview

This experiment had three different aspects as highlighted in Figure 5. The first objective was to assmeble a reference genome from a single spore collected from Western New York and sent out for Paci Bio sequncing. This was hypothesized to better encapsulates the diversity present here than the most recent genome assemble of an isolate from Blanquefort, France. The second objective was to detect selective sweep signals to characterize the biology of virulence and selection on different Resistance to *P. viticola* (RPVs) after natural infection in the field over the course of one growing season, which would provide about five generations of *P. viticola*. This was to determine the regions under selections, and predict what genes and proteins were present in those locations that may be attributed to the host's resistance. The third aspect looked to mimic the field infections by serial passaging a mixed population of *P. viticola* in the lab for at least five generations.

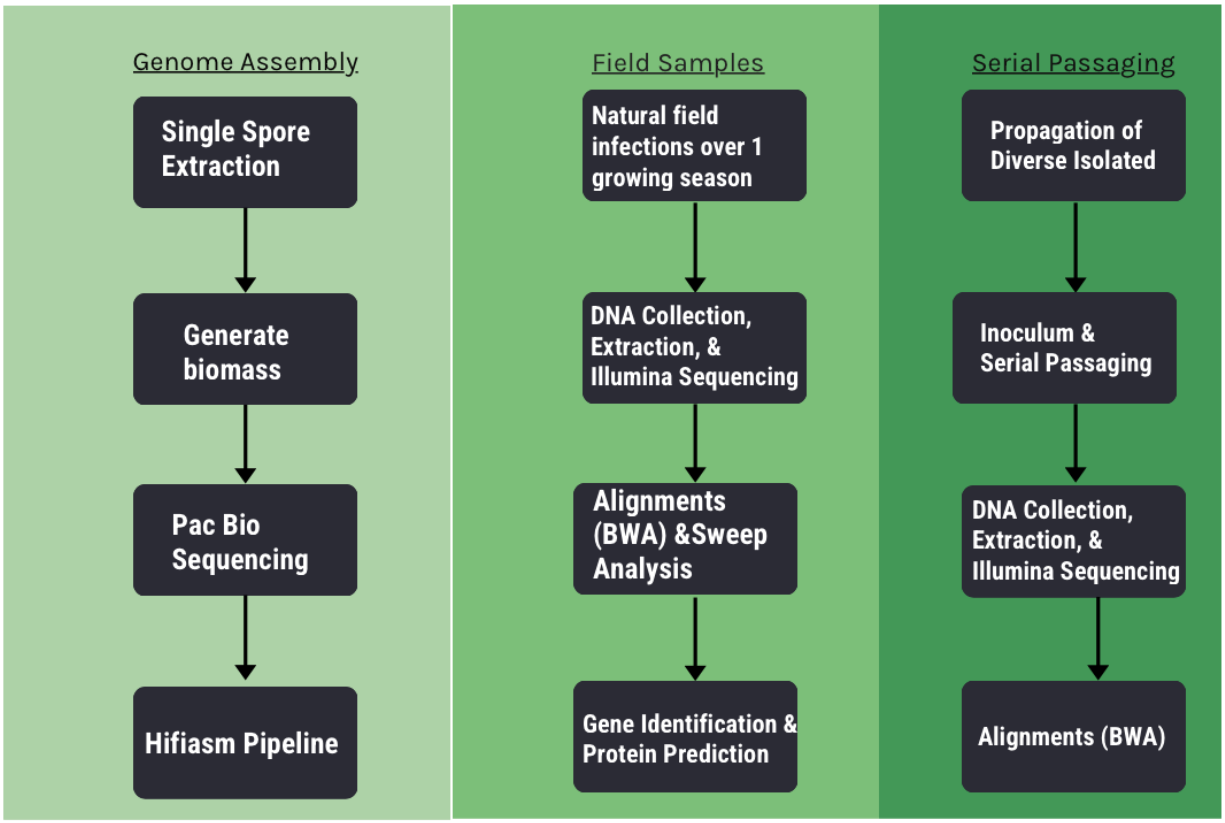

Figure 5: A flow chart visualizing a brief overview of the steps taken for each of the three different parts of this thesis research.

Single Spore Isolations (Genome Assembly)

56 leaves were collected from around the New York Finger Lakes Region from different *Vitis vinifera* cultivars and multiple *Vitis* species. The 56 leaves containing the *P. viticola* spores were placed under a microscope, and fine tip tweezers were used to extract a single sporophore from the leaf. Two susceptible, Chardonnay leaves were placed abaxial side up on a wet paper towel in a petri dish. The sporophore was then placed in a 10 µL droplet of nanopore water on the Chardonnay leaf. This process was repeated seven more times for the same leaf and subsequently for the remaining 55 leaves. The petri dishes were placed in a dark cabinet overnight and moved to an incubator the following morning. After 6 days, the leaves should have grown mini colonies of *P. viticola* sporangial on the surface. The same procedure was repeated

to assure that each clusters contained a single, genetically identical colony. 6 days after the second transfer, a leaf disc containing a cluster of *P. viticola* was punched out and placed in a 1.5 mL tube. Once sporulation was observed, the leaf discs were dried at room temperature for 8 hours and frozen at -20 ˚C. 500 µl of water was added to the tube containing the DeChaunac leaf disk to create a sporangial suspension. This suspension was pipetted in small drops onto Chardonnay leaves. Once sporulation was observed, the sporangia were washed off the leaf surface with water. This mixture was transferred to a spray bottle and sprayed onto new Chardonnay leaves. To increase the biomass, the above rinse and spray procedure was replicated, and the resulting Chardonnay leaves were placed in crisper boxes with moist paper towels. The sporangia were then collected using a cotton swab dipped into Type 1 water. One cotton swab could be used for one leaf's sporangia. The swab was then rinsed into a 2 mL microfuge tube filled with 0.5 mL of Type 1 water. After this collection, the samples were frozen at -80 ˚C to await transport to University of Delaware for PacBio sequencing.

Hifiasm Pipeline (Genome Assembly)

The Pac Bio FastQ file containing the long reads was put through bbmap (Bushnell, 2014) using the reformat.sh built in bash script to convert the FastQ to Fasta format. The fasta file was then used as the input file for bbmap using the readlength.sh built in script to generate a histogram for read lengths with a bin size of 100. This histogram was then put through a custom R script that generated a bar plot of read lengths. Using the fasta file, K-mer Counting (KMC) (Kokot et al., 2017) was applied to count the k-mers from the sequencing reads with a k-mer size of 21 as this is the size Dussert et al. used when generating INRA-PV221 (Dussert et al., 2019). The kmer count file was then transformed into a histogram to visualize. GenomeScope (Vurture

et al., 2017) was then applied to analyze the k-mer histogram. Hifiasm (Cheng et al., 2021) was used to assemble long reads from the original FastaQ file. Custom awk commands were then used to extract the contig sequences from the gfa files generated from hifiasm and reading this information into Fasta files. These Fasta files were then ran through Busco (Simão et al., 2015) using the Alveolata, stramenopiles, and eukaryote datasets as the lineage dataset to assess completeness. These fasta files were also put through a custom python script that measured the N50 and N90 statistics as well as total length. This concluded the genome assembly and the field sample analysis was started.

Disease Ratings for Field Infections (Field Samples)

The following method was applied to all 19 plants in this experiment. 10 leaves were located on the plant. If more than half of the leaf was covered in *P. viticola,* and this was evident in all 10 of the leaves, the plant received a disease rating of 0.5. If this same parameter was met, and 10 clean leaves were unable to be located, the plant was evaluated by counting the number of clean leaves able to be located multiplied by 0.05 and subtracted from 1. For *P. viticola* infected leaves, if the lesions were small and covered only a small section of the leaf, a disease rating between 0.01 and 0.09 was provided. After the final disease rating, the leaves were then collected for sequencing.

DNA Collection and Extraction (field samples)

4 leaves were selected per plant within a collection. If the leaves did not have an adequate amount of downy mildew, the number of leaves used was increased. It was observed that using four leaves was the least number of leaves to use on a plant containing no RPVs and

RPV3.1+RPV10.3 (which has the least number of lesions), and still produce a high, yet similar amount of DNA. An Eppendorf tube was then filled with 300 µL of CTAB extraction solution. A separate Eppendorf tube was filled with CTAB extraction solution as well. Using a Q-tip, one end was dipped in one of the full Eppendorf tubes and was rolled over all areas containing spores on the leaf. In the other tube, the "spored" end of the Q-tip was twisted around. Failure to soak the Q-tip before collecting the spores will result in the absorption of the 300 µL of extraction buffer. If this does happen, another 300 µL was added and continued twisting the Q-tip around. It is still possible to collect DNA. Once all four leaves had their spores removed, the tubes were placed in the centrifuge and spun for one minute at 12,000 rpm. If DNA has formed into a pellet but does not remain stuck to the side of the tube, centrifuge for another minute maintaining the 12,000 rpm speed. After the centrifugation was completed, the extraction buffer was emptied from the tube. 100 µL of isopropyl alcohol was added to dry out the excess extraction buffer to ensure ice does not form over the pellet because of excess liquid. After the isopropyl alcohol was added, the tube was centrifuged for 15 seconds at 12,000 rpm. Any more than 100 µl or longer than 15-20 seconds resulted in the pellet dissolving slightly or become dislodged. If the pellet does become dislodged, 100 more µl of extraction buffer was added and centrifuged with the same parameters (15-20 seconds at 12,000 rpm) but allowing the tube to sit opened for 3 minutes to evaporate the excess liquid. After the time is complete, two Spex beads were placed into the tubes containing the DNA pellet. Once tube beads were inserted, all tubes were placed on ice until the DNA collection was complete. Once DNA collection was completed, the tubes were put into a freezer safe container and placed in the -80 ˚C freezer until DNA extraction. For DNA extraction, the spore pellet was thawed and placed in 500 µl of CTAB buffer before being transferred to a microfuge tube. The mixture was incubated for 15 minutes at 55˚C in

recirculating water bath. Following the incubation, the mixture was centrifuged at 12000g for 5 minutes. The supernatant was transferred to clean microfuge tubes. To each tube, 250 µl of Chloroform: Isoamyl Alcohol (24:1) was added and mixed by inversion. After thoroughly mixed, the tubes were then centrifuged at 14000 rpm for 1 minute. The upper aqueous phase was transferred to a clean microfuge tube were 50 µl of 7.5 M Ammonium Acetate was added. 500 µl of ice-cold absolute ethanol was then added. The tube was then inverted slowly several times to precipitate the DNA. After precipitating the DNA, the DNA was then pipetted off by slowly rotating a tip in the cold solution. The DNA stuck to the pipette. To wash the DNA, the precipitate was transferred to a microfuge tube containing 500 µl of ice cold 70% ethanol and mixed using careful inversion. After washing the DNA, the tube was centrifuged for one minute at 14000 rpm. The supernatant was disposed while the DNA pellet was allowed to air dry for 15 minutes. The DNA was then resuspended in sterile Dnase free water (between 50-400 µl). The amount of water depended on the amount of DNA isolated. 10 µl/ml of RNaseA was added prior to the addition of DNA to the water to remove RNA. After resuspension, the DNA was incubated for 20 minutes at 65˚C to deactivate Dnase and stored at 4 ˚C. To access the DNA quality, agarose (1% solution) gel electrophoresis was used. A 10 µl 1 kb ladder and 5 µl sample + 5 µl water or 10 µl sample + 2 µl 6x loading buffer was added to the wells. The gel was run for 30 minutes at 100 V and imaged. DNA was isolated and Illumina libraries prepared for each sample using the Illumina DNA Prep (M) Tagmentation kit, followed by sequencing on an Illumina NovaSeq X. The reads were returned in FastQ format, and the adapter sequences were trimmed off using Trim Galore!, and then heterozygosity was ready to be assessed.

Heterozygosity within Field Samples

Jellyfish count was applied to the 17 field samples using a kmer size of 25 and a hash table size of 6250000000 (to account for the memory available) (Marçais and Kingsford, 2011). Jellyfish histo was then applied to obtain the .histo format that GenomeScope requires. GenomeScope was then run using a kmer size of 25 and a -p of 2 to indicate these samples are diploid (Vurture et al., 2017).

Alignments (Field Sample)

After confirming that the samples were majority *PvA*, FastQC (Andrews, 2012) was then applied to the 17 paired-end sample reads. Once cleaned, they were then aligned to the reference genome INRA_Pvit_2 (GCA_001695595.3) using BWA-MEM as the aligner with default parameters (Li, Heng, and Durbin, 2009). Once the alignment was complete, the number of mapped reads versus the number of unmapped reads was accessed using Samtools flagstat (Danecek, Petr, et al., 2020). This was then visualized with a custom R script utilizing the library ggplot2 and is shown in Figure 10.

Selective Sweep Pipeline (Field Samples)

Using the output file from the samtools (Danecek, Petr, et al., 2020). filtering, freebayes (Garrisone, Erik, and Marth, 2012) was applied for variant calling using default parameters. The variant vcf files had to be zipped and indexed using bgzip and tabix respectively. These variant files, which contained variants from strains growing on the same RPVs, were then merged using bcftools merge (Danecek, Petr, et al., 2020). with default parameters. The haplotypes were then calculated using Beagle (Browning et al., 2021). If a genotype was ambiguous (marked by ./.), it was filtered out using bcftools view (Danecek, Petr, et al., 2020).

The selective sweep regions were predicted using SweeD-P-C (Pavlidis, Pavlos, et al., 2013) with a grid size of 694. SweeD uses theoretical models based on population size changes and exponential growth without the need of computing a site frequency spectrum (SFS) for the individual genomes. If this is not provided, it will estimate based on the population data, and choose the neutral SFS model based on these estimations. This helps account for various demographic events that could cause deviation from the neutral SFS. It also scans broad regions to determine if the deviations are isolated or part of a broader pattern. If a region significantly diverges from the neutral SFS, then the probability it is a "sweep model" is divided by the probability that it follows the neutral model is calculated and reported only if the sweep model is higher. The neutral model expects a distribution that's heavily weighted towards low frequency variants due to the random effects of genetic drift. The selective sweep will then result in an SFS that has fewer intermediate frequencies and an excess of high-frequency alleles as the beneficial mutation and its linked sites rapidly increase in frequency (Nielsen et al., 2005). The sweep data from across the genome was visualized in Q-Q plots (Wickham, 2009) that can be seen in Figures 12-14. The x-axis contains the log values of no rpv *P. viticola* samples functioning as the control while the y axis contains the samples grown on vines containing the different RPVs. Because each of the quantiles represents 16.7% of the data, the range of quantiles in which the divergence was located was multiplied by 16.7% to get the lower and upper percentile for that divergence. Once the sweep data had been filtered, the filtered file's range was compared to Dussert et al. (Dussert et al., 2018) gene annotation range for the reference genome to determine what genes are present in the high likelihood sweep region. After determining the upper range of likelihood scores based on breaks in the data as shown in Figures 16 and 17, Augustus (Stanke et al., 2006) was run on the scaffolds of those "upper" scores. The gene's start and end position was

then compared with those found by Augustus (Stanke et al., 2006) to find regions of overlap. From there, the protein sequences were extracted and put through blastp (Altschul et al., 2007). This entire pipeline is shown in Figure 6. The top hit was recorded, which is where the predicted proteins came from in Tables 4 and 5. SignalP (Almargo et al., 2019) was used to determine if signal peptides were present by selecting Eukarya for the organism group and the long output for format. EffectorP (Sperschneider & Dodds, 2022) and SecretomeP (Bendtsen et al., 2004) was used to determine what type of effect, cytoplasmic or apoplastic, an identified RxLR effector was as well as if the predicted effector is secreted without having a signal peptide identified by SignalP. For EffectorP, EffectorP-fungi 3.0 was used as this included oomycete. For SecretomeP, mammalian was selected as there was not an option for fungi or oomycetes. The final aspect, serial passaging, was now ready to be analyzed.

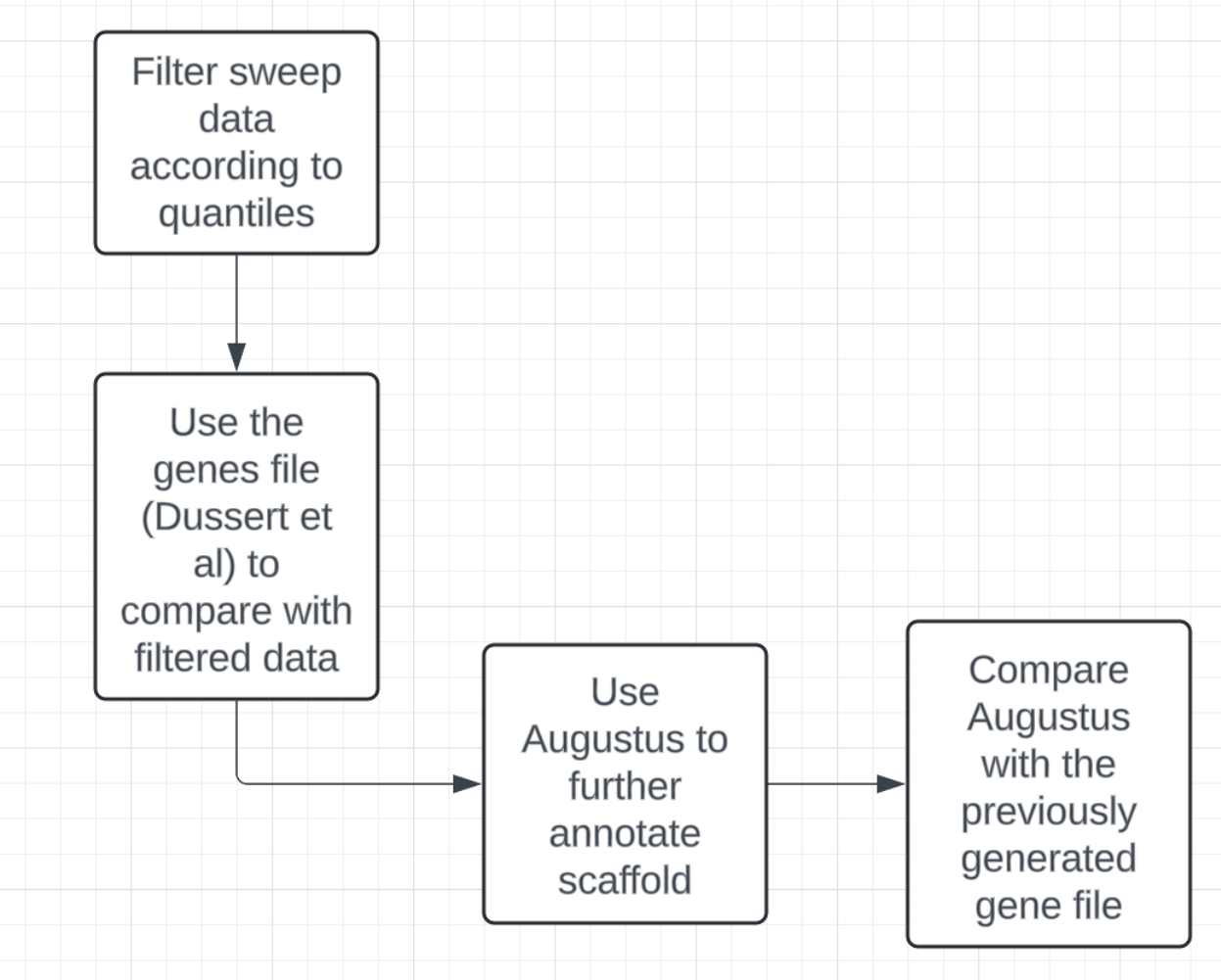

Figure 6: A flow chart outlining how the regions of high likelihood sweep regions were identified using Augustus (Stanke et al., 2006).

Propagation (Serial Passaging)

5 greenwood cuttings were selected from 20 plants used in both the laboratory and field experiments discussed later in this section. All cuttings were taken from full sibling plants resulting from the cross-hybridization NY84.0101.03 x *V. amurensis* '588634'. The cuttings were dipped into 1:5 Dip N Grow rooting hormones before being placed in a soil cone and moved to a mist bed. Once roots were visible, the cuttings were transferred into appropriately sized pots and placed in the greenhouse to finish growing. The table below details how many plants were available. 5 cuttings were taken to propagate to have an adequate number of leaves

as well as to repeat the serial passaging a second time to corroborate results/data from the first serial passaging. While these plants were propagating, inoculum was prepared.

Plant Label	RPV	# of cuttings tried	# of that grew
9-54	**None**	**5**	**2**
10-14	None	5	0
10-70	**None**	**5**	**5**
11-39	**None**	**5**	**3**
12-25	**None**	**5**	**3**
9-25	RPV3.1	5	0
9-37	RPV3.1	5	0
10-46	**RPV3.1**	**5**	**2**
11-64	RPV3.1	5	0
12-58	**RPV3.1**	**5**	**4**
9-28	**RPV10.3**	**5**	**2**
10-77	**RPV10.3**	**5**	**1**
11-57	**RPV10.3**	**5**	**5**
11-65	RPV10.3	5	0
12-70	**RPV10.3**	**5**	**2**
9-20	RPV3.1+RVP10.3	5	0
11-2	**RPV3.1+RVP10.3**	**5**	**1**
11-9	**RPV3.1+RVP10.3**	**5**	**2**
10-28	**RPV3.1+RVP10.3**	**5**	**3**
12-40	RPV3.1+RVP10.3	5	0

Table 2: The plant IDs along with the RPV that each cutting contained and how many of the cuttings survived are reported in the table above. The plants in bold were used in the serial passaging experiment.

Inoculum (Serial Passaging)

Before the end of the growing season, leaves containing *P. viticola* from each plant for each RPV was collected from the field. The leaves were separated based upon what RPV they contained. The leaves containing no RPVs were placed in a funnel connected to a glass jar and sprayed with distilled water to collect the spores in the jar below. Once all the leaves for this locus were washed, the spores were then sprayed onto four susceptible Thompson Seedless leaves. Younger, susceptible leaves were avoided when selecting leaves to infect. This was repeated for four susceptible Thompson Seedless leaves receiving inoculum per each of the other three loci. The leaves were placed in a cabinet overnight to sporulate and moved to an incubator the following morning. This inoculum was transferred to fresh, susceptible leaves every 6 days.

Once the propagated plants contained an adequate number of leaves each, the spores from each of the separate leaves were washed off together and sprayed onto each of the leaves selected from the plants in Table 2.

Serial Passaging

The propagated plants in Table 2 were grown until an adequate number of leaves could be extracted for the two repeated experiments. A single leaf was collected from each plant and placed on a wet paper towel at the bottom of a petri dish. The Thompson Seedless leaves, containing the downy mildew spores for each of the RPV were combined into a bulk inoculum by spraying each of the leaves with distilled water. This bulk inoculum, representing the diversity of resistance loci, vines, and pathogen diversity recovered from the field experiment, was then used to infect all the leaves from the propagated grapevine leaves using a spray bottle to mist the abaxial surface with the newly made inoculum. One-quarter of the inoculum was saved for DNA sequencing by being stored in a -20˚C freezer. The petri dishes were sealed with parafilm and placed in a dark space for 12 hours to allow for sporulation. After the 12 hours were up, the spore suspension was dried by dabbing with a Kimwipe, and petri dishes were then transferred to a 20˚C incubator to allow for ideal temperature to induce further sporulation. Two days later, a second experiment following the same steps as outlined previously was carried out. The spores were allowed to grow for 5-6 days before being transferred onto a fresh leaf for the next generation. The individual inoculum was frozen in a -20˚C freezer. Once the Thompson Seedless from both experiments finished at the 10^{th} transfer, the experiments were stopped despite no other leaves being alive for their 10^{th} transfer. DNA isolation and Illumina sequencing on an Illumina NovaSeq X was completed as described above for field samples.

Results

Genome Assembly Stats

Raw reads from PacBio sequencing were assessed using bbmaps to estimate the read length distribution (Figure 7). The peak was at 7,950 bp with a right skewed distribution and standard deviation of 1,843 bp. These read lengths were similar to the Dussert et al. (2019) average read length of 8,072 bp. Genome Scope was used to predict the layout of the assembly (Figure 8). The predicted genome length was 155,100,514 bp, but only 65.4% were unique, which could imply duplications, incomplete assemblies, or polyploid genomes. The kcov was on the lower side with a score of 33.6 meaning that on average, a k-mer of 21 was observed 33.6 times. Due to the relatively low coverage, the assembly may have problems assembling repetitive or complex regions, leading to fragmented contigs and larger and more frequent scaffold gaps. Furthermore, the assembler may incorrectly collapse short, repetitive sequences into a single contig despite their presence in multiple regions of the genome, resulting in errors in the assembly.

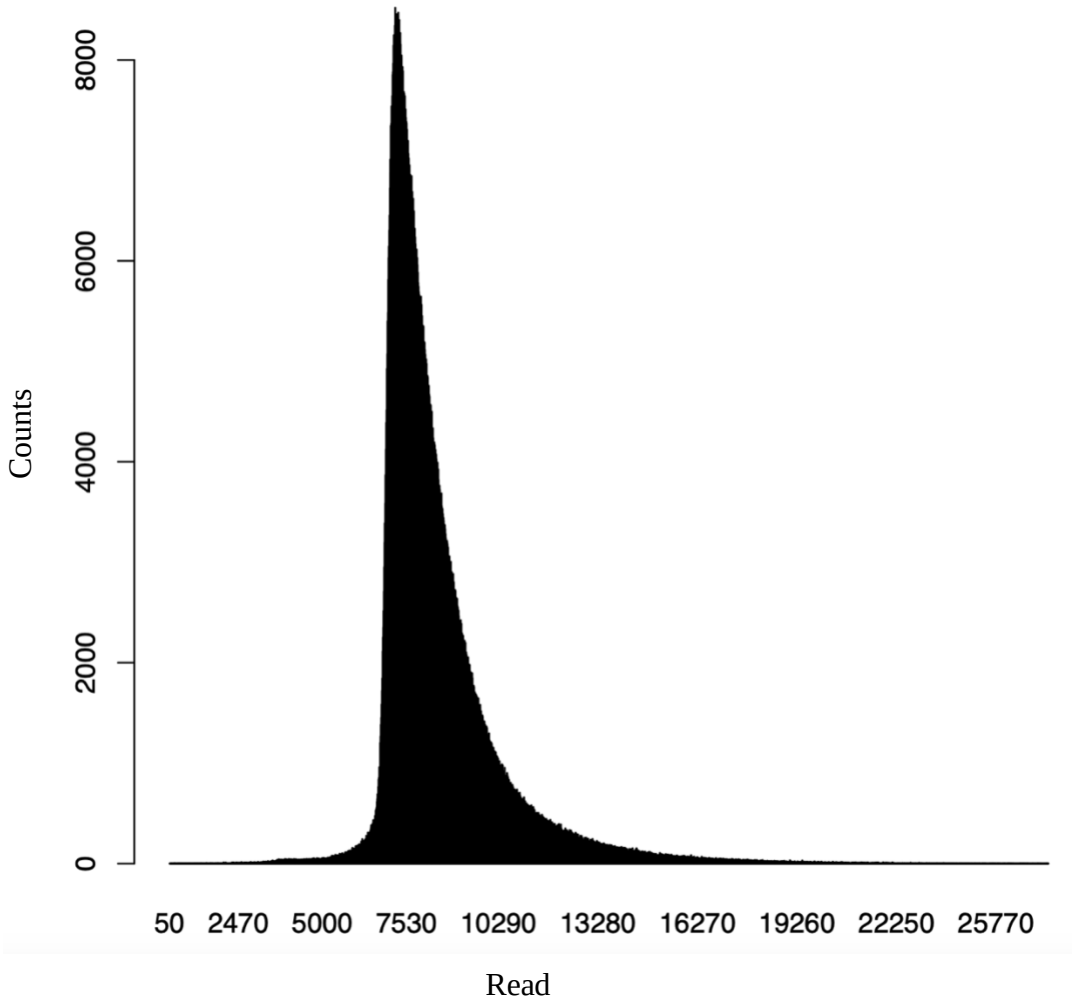

Figure 7: Frequency of read lengths for the PacBio long reads.

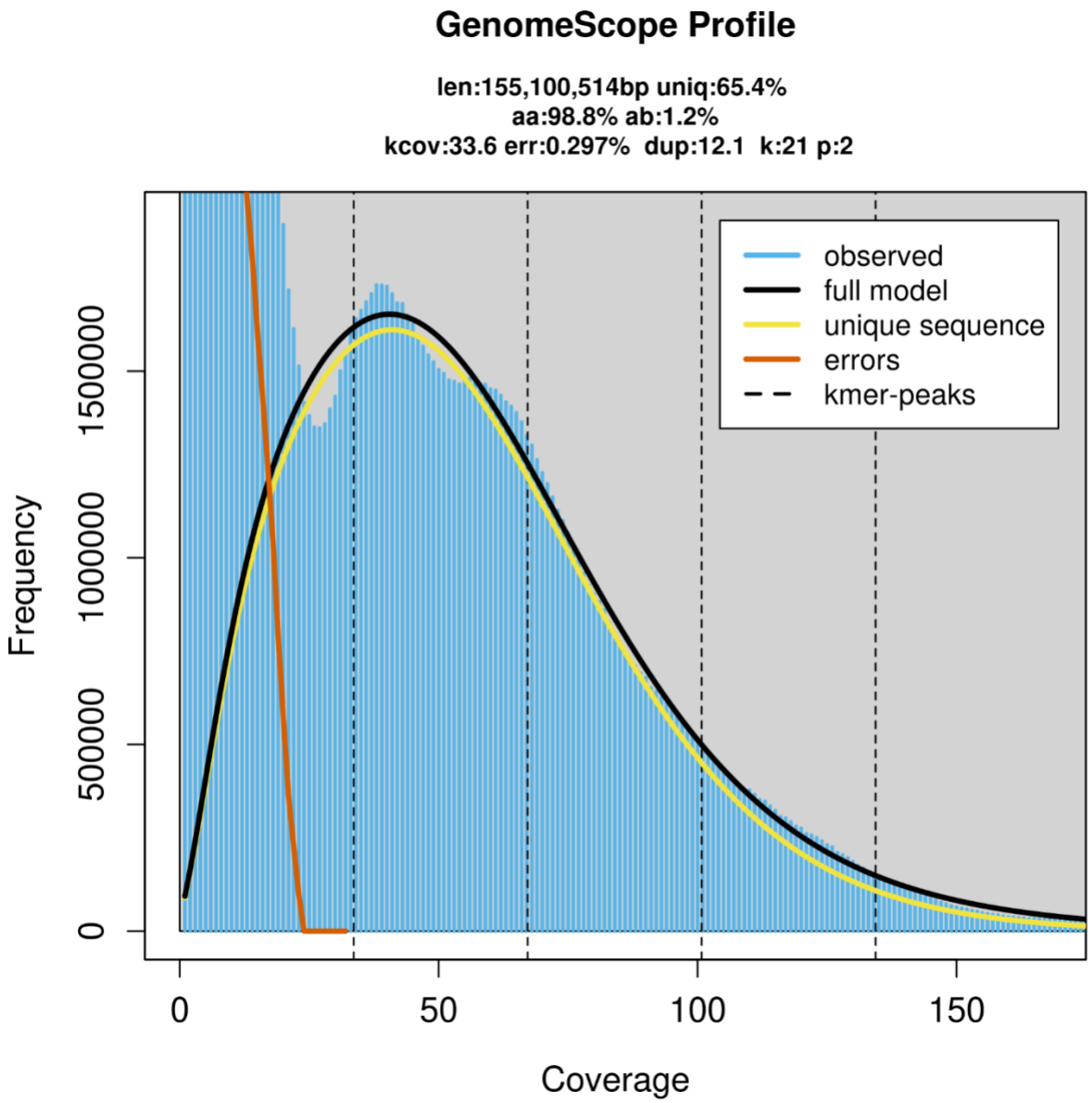

Figure 8: The genome scope analysis and predictions for the Pac Bio long reads.

Statistics describing the Hifiasm assembly indicated a relatively high level of completeness based on BUSCO gene representation, but with a high degree of fragmentation (Table 3). While the N50 is relatively high (2.3 Mb to 3.0 Mb), suggesting large contigs, the N90 is smaller (100 kb to 531 kb), suggesting that there are many smaller contigs, causing the assembly to be very fragmented. BUSCO (Simão et al., 2015) was used to access for complete. *P. viticola* was part of three different datasets, alveolate, stramenopiles, and eukaryote. Although the BUSCO scores were all above 90%, indicated high completeness, these scores were derived from high duplication rates (73.7% duplication for Alveolata genes), indicating that this assembly was either highly fragmented or multiple haplotypes of the same region were present as separate contigs, so to attempt to address these issues, a different –hom-cov of 70 was selected when running the Hifiasm assembler (Table 3). The haplotype 1's length was decreased by 108 Mb and haplotype 2 was decreased by 56 Mb when using this new -hom-cov. RepeatMasker (Tarailo-Graovac and Chen, 2009) was also attempted to mask repetitive sequences to focus the analysis on the unique regions of the genome. This resulted in no change of the genome length nor the BUSCO completeness.

	Haplotype 1	Haplotype 2
Total Length	278,494,004 bp	256,636,789 bp
N50	2,364,878 bp	3,050,775 bp
N90	100,202 bp	530,945 bp
BUSCO (Alveolata)	96.5% (S: 22.8% D:73.7%)	96.2% (S:22.8% D:73.7)
BUSCO (Stramenopiles)	100% (S:19.5% D:80.5%)	100% (S:20% D:80%)
BUSCO (Eukaryota)	94.5%(S:19.6% D:74.9%)	93.8% (S:18.7% D:75.1%)

Table 3: Stats on the two haplotypes of the draft diploid assembly.

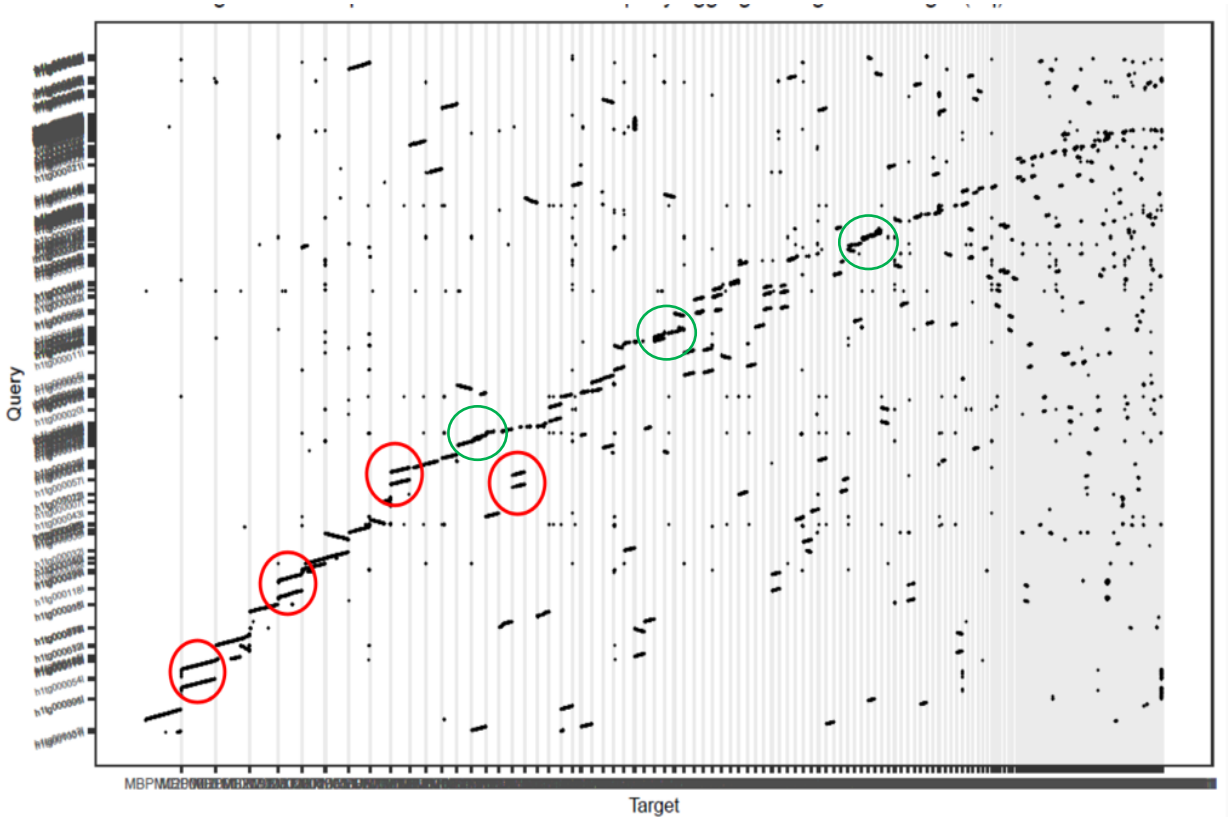

Figure 9: This assembly (query) compared to Dussert et al genome (target). Circled in green are examples of areas with good alignment. Areas circled in red are areas of concern due to possible duplications.

Although the assembly showed a high BUSCO completeness score, a majority of the completeness derived from duplications. Comparatively, Dusset et al. (2021) obtained a BUSCO score of 95.7% with a duplication rate of 1.7% whereas this assembly had an average BUSCO score of 97% and an average duplication rate of 76.4%. The next concern was to determine the similarities or differences between this assembly and Dussert et al. 2021 to determine areas that may have been misassembled, duplicated, or incomplete. Figure 9 shows how well the assembly aligned to the previous reference assembly (Dussert et al., 2021). Ideally, a genome dot plot comparing two assemblies should be a near perfect diagonal line showing 1:1 matches of assembled genome sequences. In comparing our assembly with Dussert et al. 2021 in Figure 9, some well-matched examples are circled in green. This indicated the areas that aligned well with the reference genome. The areas with points outside of the diagonal suggested differences

between this assembly and the Dussert et al. assembly as a result of insertions, deletions, or rearrangements. Some areas of possible duplication, a ploidy assembly, or misassembly were represented by red circles.

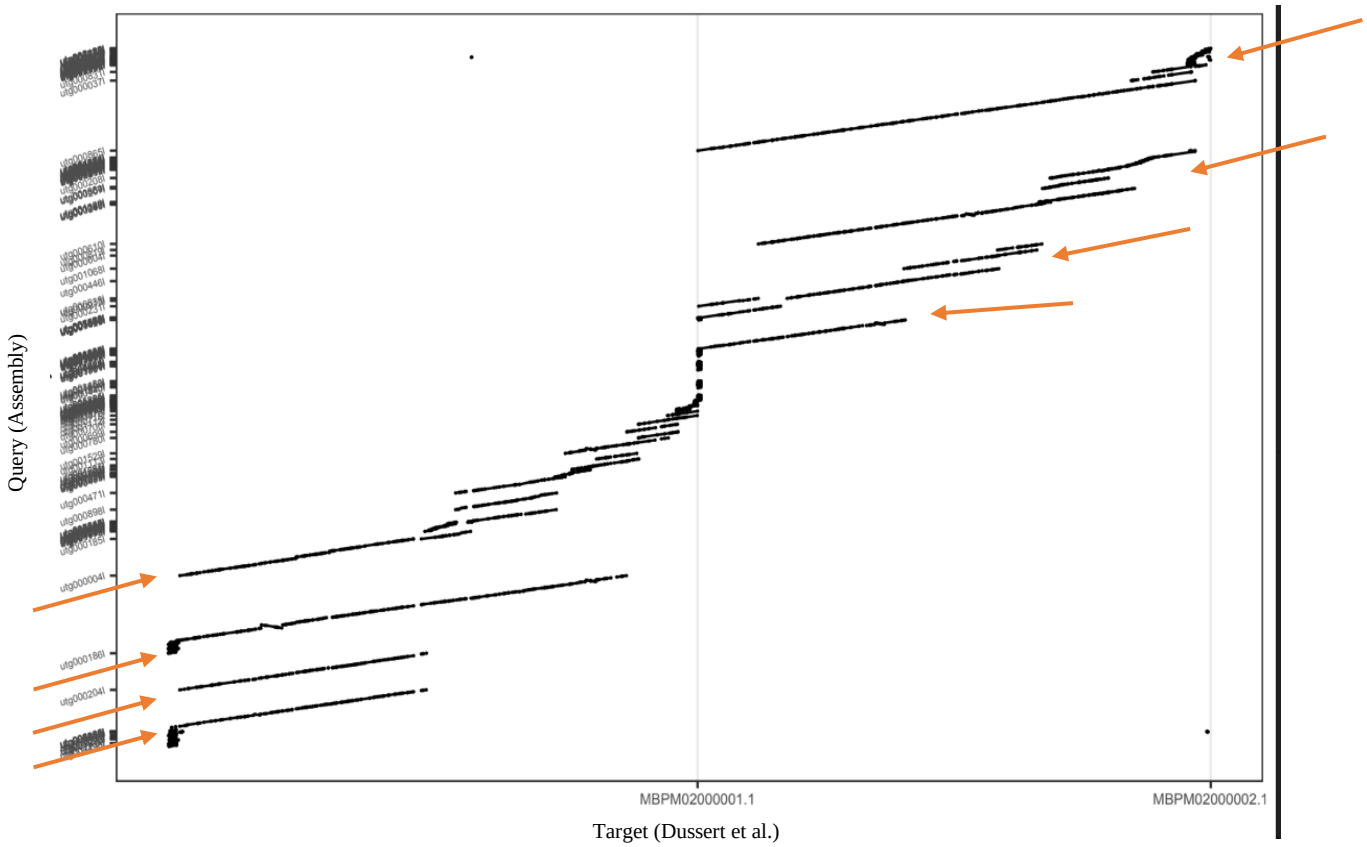

Figure 10: The assembly was aligned with the longest scaffolds from the Dussert et al. reference genome. This is the resulting alignments. The orange arrows represent the four different haplotypes found.

In an ordered dot-plot (Figure 10), there were four distinct haplotypes that were aligning to the two longest scaffolds. The bottom two haplotypes appeared to be truncated and ultimately only cover half of the genome likely due to the insufficient confidence to extend the contig beyond certain boundaries. Therefore, these two haplotypes reached the bounds and could not continue. Because of this, instead of using an assembly pipeline for diploid organisms, these reads will have to be assembled through a pipeline designed for tetraploid organisms and were therefore not used as the reference for the field samples alignment.

Disease Progression across Susceptible vs Resistant Samples

Disease severity for the 19 field plants was assessed over one growing season, and the final ratings were visualized in a box plot (Figure 11). This was estimated to be around five generations of *P. viticola* on the assumption they had a consistent six-day sporulation latent period. The susceptible checks (no RPV) had the most sporulation except for plant 10_70, which had a final percentage of 8%. Due to these plants being F1 progeny of highly heterozygous parents subject to natural infection in the field, there are multiple factors that could be affecting this, such as the dispersion of spores not reaching this plant, neighboring vines being resistant resulting in lower local inoculum, or uncontrolled background genetic effects.

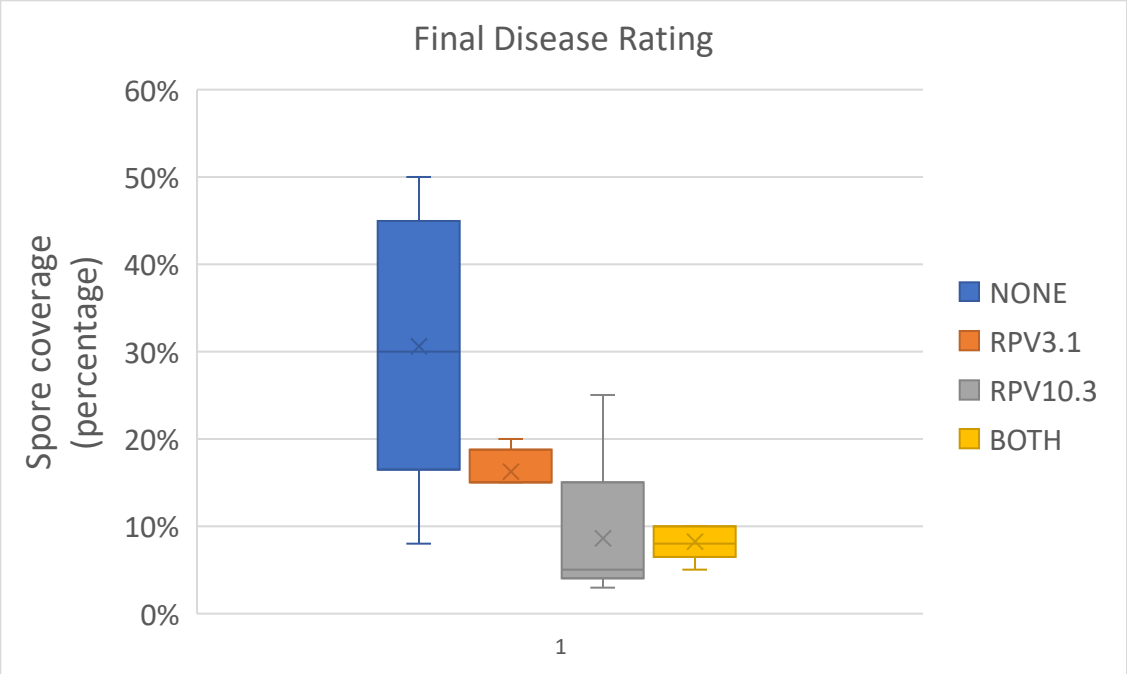

Figure 11: The final disease rating of *P. viticola* infection at the end of the 2023 growing season in the four host resistances.

Illumina library quality control

Using samples that were included in the final disease rating, two of the 19 samples (10-38 and 11-74) had insufficient Illumina genomic DNA library concentrations, 0.501 ng/µl and 3.2 ng/µl respectively, to yield high quality sequencing results (Table 4). These *P. viticola* samples grew on grapevines containing both RPV3.1 and RPV10.3 and were eliminated from further analysis. This resulted in five samples for no RPV, four samples for RPV3.1, five samples for RPV10.3, and three samples for RPV3.1+10.3.

Sample Name	DNA conc. (ng/ul)	[library] (ng/µl)
9-27	10.5	21.2
9-28	6.6	16.7
9-54	31.4	32.6
10-11	90.2	29.4
10-15	0.376	9.240
10-24	123.0	27.8
10-28	7.38	22.8
10-38	4.58	0.501
10-70	32.4	29.6
11-39	40.0	26.4
11-57	17.0	16.5
11-74	3.72	3.2
12-6	2.14	6.7
12-12	5.8	17.1
12-24	11.4	20.4
12-25	64.4	26.4
12-58	31.4	20.2
12-70	34.4	23.8
12-81	0.231	14.4

Table 4: The DNA concentration for each sample prepared by using Illumina DNA Prep (M) Tagmentation kit. The samples highlighted in light red were excluded due to the low concentration of the library.

Heterozygosity of the Field Samples

To generate a baseline understanding of the genetic makeup of the 17 remaining samples, heterozygosity, genome length, and repeat percentages were assessed (Figure 12). All the samples for no RPVs and RPV10.3 were 2.0% or below heterozygous while two samples from RPV3.1 and RPV3.1+RPV10.3 were above 2.0% heterozygous. While there was no significant correlation between heterozygosity and RPVs nor heterozygosity and susceptible (no RPVs)/resistant (RPV3, RPV10, Both) groups, there was a moderate, negative correlation (R = -0.45) between heterozygosity and estimated genome length as well as heterozygosity and

repeated sequence percentage (R =-0.49). With respect to the correlation between heterozygosity and genome length, this could be because of factors diminishing the impact of new mutations, reducing heterozygosity or this could be reflecting an evolutionary trend of an increased genome is associated with a decreased genetic variation due to selective sweeps. With respect to the correlation between heterozygosity and repeat percentage, this correlation suggested that regions with high repeat percentages were subjected to selection pressures that maintain the genetic diversity, or the accumulation of repeats could be associated with mechanisms that suppress mutation rates in surrounding regions.

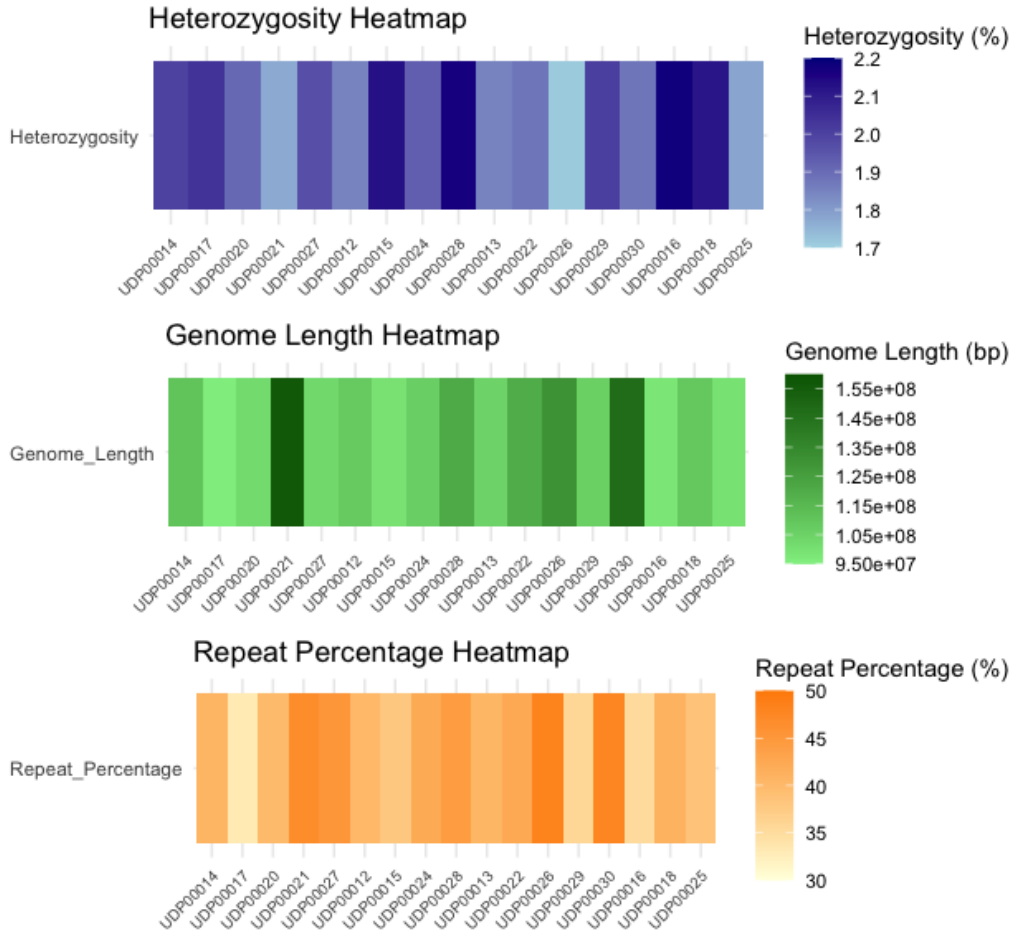

Figure 12: The heatmaps of heterozygosity, genome length, and repeated sequence percentage across the field sample population. The first five samples contained no RPVs (UDP0014, UDP00017, UDP00020, UDP00021, and UDP00027). The next four contained RPV 3.1 (UDP00012, UDP00015, UDP00024, and UDP00028). The next five contained RPV10.3 (UDP00013, UDP00022, UDP00026, UDP00029, and UDP00030). The final three samples contained both RPV3.1 and RPV10.3 (UDP00016, UDP00018, and UDP00025).

Field sample Cryptic Species Composition

The 17 Illumina samples were processed starting with trimming the adapters using trim galore! (Krueger, 2019). These trimmed reads were then aligned to the *PvA* ITS and 28S region using STAR aligner (Dobin et al., 2012) using the default parameters to determine the composition of cryptic species in each sample read (Table 5). Samtools (Danecek, Petr, et al., 2020) was used to eliminate alignment samples that had a quality score of less than 60. To determine the cryptic species, each species had a sequence of 20 nucleotides that included the SNPs identified in Table 1. For the ITS region, *PvA* sequence was GCAGCTAATGGATTCCTATC, *PvR* was TAACTGACTTTATTGTCGGT, *PvV* was ATAGCATGGAATTAATTCCG, and *PvQ* was TTGATTTCTATCATAGTGAA. For the 28S region, *PvA* was CCGTTCGTCCCCAAGTTGCT and ATTGGCGAGTGTATGCGTGC, *PvR* was GGTCAGTATGAGCACTTGGG, TGACGAGCGTGTGCGTGCGT, and AGTGGCCTTTTGGCTGCGCTC, *PvV* was GTAGTCTATGGAAGCGTGGT, GCGCAAAGCAGGTGGTAAA, and TGTGCTTGCCGGTGCCCTGT, and *PvQ* was CCGTTCATTCCTAAGTTGCT and GGATTTGGATCTCCGTGTGC.

ITS region

Plant #	Comp. #	Aestivalis	Riparia	Vinifera	Quinquefolia
9-27	00012	✓			8 instances
9-28	**00013**	**✓**	**✓**		**2 instances**
9-54	00014	✓			8 instances
10-11	00015	✓			
10-15	00016	✓			2 instances
10-24	00017	✓			
10-28	00018	✓			4 instances
10-70	00020	✓			14 instances
11-39	**00021**	**✓**	**✓**		**1 instance**
11-57	**00022**	**✓**	**✓**		**2 instances**
12-6	00024	✓			
12-12	00025	✓			5 instances
12-24	**00026**	**✓**	**✓**		
12-25	00027	✓			1 instance
12-58	00028	✓			
12-70	00029	✓			14 instances
12-81	00030	✓			9 instances

Table 5: In bold are the *P. viticola* samples that contained an admixture of both *PvA* and *PvR*. The check marks in the ITS region table indicate that the sample has a high number of instances of those identifiers.

28S Region

Plant #	Comp. #	Aestivalis	Riparia	Vinifera	Quinquefolia
9-27	00012	19,142 instances		4 instances	2 instances
9-28	**00013**	**28,480 instances**	**923 instances**	**1 instances**	**8 instances**
9-54	00014	17,919 instances	1 instance	3 instances	3 instances
10-11	00015	25,596 instances	1 instance	3 instances	
10-15	00016	37,581 instances	40 instances	9 instances	2 instances
10-24	00017	45,181 instances	5 instances	3 instances	5 instances
10-28	00018	22,898 instances	3 instances	5 instances	4 instances
10-70	00020	31,123 instances	4 instances	15 instances	4 instances
11-39	**00021**	**13,966 instances**	**369 instances**	**8 instances**	
11-57	**00022**	**9,862 instances**	**454 instances**	**5 instances**	**8 instances**
12-6	00024	15,831 instances	1 instance	5 instances	3 instances
12-12	00025	24,584 instances	4 instances	8 instances	5 instances
12-24	**00026**	**19,119 instances**	**1,410 instances**	**5 instances**	**1 instance**
12-25	00027	30,392 instances	195 instances	16 instances	5 instances
12-58	00028	12,930 instances		5 instances	1 instance
12-70	00029	42,073 instances		20 instances	23 instances
12-81	00030	27,866 instances	10 instances	5 instances	2 instances

Table 5 (continued)

Mapped vs. Unmapped Reads (Field Samples)

Figure 13 shows the unmapped reads compared to the mapped reads. UDP00021 has the highest unmapped reads with 43.36% of reads being unmapped. The first 200 unmapped read sequences with a size of 150 bp were extracted and put through BLAST. When using Blastn (Altschul et al., 2007) using the ‘nt’ database, the top two results for this sample were a predicted *Mercenaria mercenaria* (hard clam) mRNA (XM_045328816.1) with an 88.1% identity across a

total length of 42 bp and Danio rerio (zebrafish) genome assembly of chromosome 17 with a 100% identity across a total length of 24 bp – both shorter than the 150 bp read length. When blastn (Altschul et al., 2007) was performed on UDP00015 that had the least number of unmapped reads, the top two results were *Pterostichus madidus* (black clock beetle) with a 100% identity at a total length of 23 bp and *Pipistrellus pipistrellus* (common pipistrelle) with a 96.6% identity and a total length of 29 bp. Due to these results not aligning with *P. viticola* nor any of its relatives, it can be hypothesized that the most recent genome assembly for *P. viticola* may not be adequately encapsulating the diversity these samples of *P. viticola* contain.

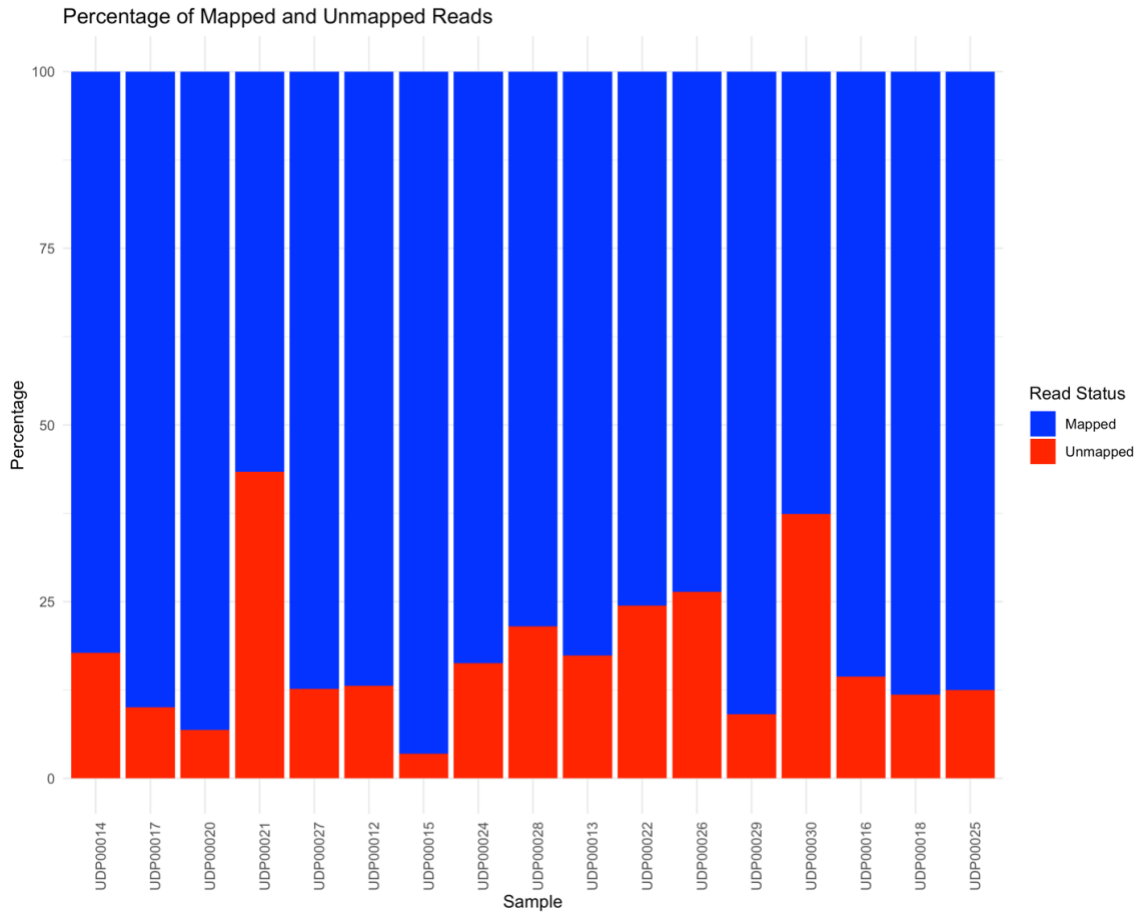

Figure 13: Percentage of mapped reads (blue) compared to unmapped reads (red) of the 17 field samples obtained through samtools flagstat. The first five reads are *P. viticola* that grew on grapevines with no RPV (UDP00014, UDP00017, UDP00020, UDP00021, UDP00027), the next four contained RPV3.1 (UDP00012, UDP00015, UDP00024, and UDP00028), the next five contained RPV10.3 (UDP00013, UDP00022, UDP00026, UDP00029, UDP00030), and the final three contained both RPV3.1 and RPV10.3 (UDP00016, UDP00018, UDP00025).

It should be noted that two out of the three highest unmapped scores contained admixtures of *PvR*, which could partially explain the high percentage. The Dussert et al. (2021), genome is *PvA*, so *PvR* instances may not align well.

Genes and Predicted Proteins in High Selective Sweeps Areas

Using the SweeD (Pavlidis, Pavlos, et al., 2013) sweep data generated by from the mapped reads, the candidate sweep region likelihood scores across the merged samples were visualized. Likelihoods represent the ratio of observing the sweep under different models of selection. In more depth, the ratio of the genetic variation if a selective sweep occurred at a particular location is calculated (Selective Sweep model) as well as the ratio that the genetic variation occurred due to random genetic drift (Neutral model). The "likelihood" or ratio that was calculated under the selective sweep model is then divided by the "likelihood" calculated under the neutral model to get the reported likelihood of that position, indicating potential selective sweeps. Higher likelihoods will indicate stronger signals of selection.

For samples grown on RPV3.1, the log likelihoods were visualized in Q-Q plots (Figure 14). The values along the x and y axis represented the number of quantiles the program divided the data into. A quantile was a percentile. The quantiles represented the likelihood scores arranged in ascending order with the red dashed line indicating a 1:1 comparison of the no RPV susceptible samples' likelihood scores in ascending order. The axes were divided into six quantiles, which was automatically calculated by the Q-Q plot (Almeida, Loy, and Hoffman, 2018) function in R based on sample data size (Ford, 2015). In the case of Figure 14-16, each number represented the '16.7 x n' percentile of likelihood scores, disregarding positions in the genome or scaffold number where n was the quantile number. The x-axis and the red dashed line

contained the log-likelihoods taken from samples grown on vines containing no RPV susceptible samples. The y-axis contains the log-likelihoods of the sample grown on either RPV3.1 vines, RPV10.3 vines, or RPV3.1+RPV10.3 vines. The log was applied to the likelihood scores to attempt to normalize the skewed right distribution.

Between quantile 3 and 3.5 (representing 50-58th percentile of data) as well as quantile 4 and 4.5 (66-75th percentile of data) was where the likelihood scores for resistant samples (RPV3.1) significantly diverged from the expected values of the susceptible samples, suggesting regions of potential importance for resistance. Because the divergence was above the 1:1 comparison, the likelihood scores were higher than expected for these particular ranges. This overall divergence was further supported by the box plot(Figure 17). The log of RPV3.1 resistant samples was significantly different ($P < 0.001$) when compared with the log of susceptible samples within the box plot according to a Wilcoxon rank-sum test. The high number of outliers was most likely explained by the fact that the likelihood scores for all of the samples were skewed right, despite attempts to normalize it through log transformation. Combining these results (Figure 14 and Figure 17) suggested that the difference in likelihood is not by random chance alone. However, the most statistically significant regions (above 99th percentile) did not significantly diverge from expectations, suggesting that no major selective sweep exists on RPV3.1 vines.

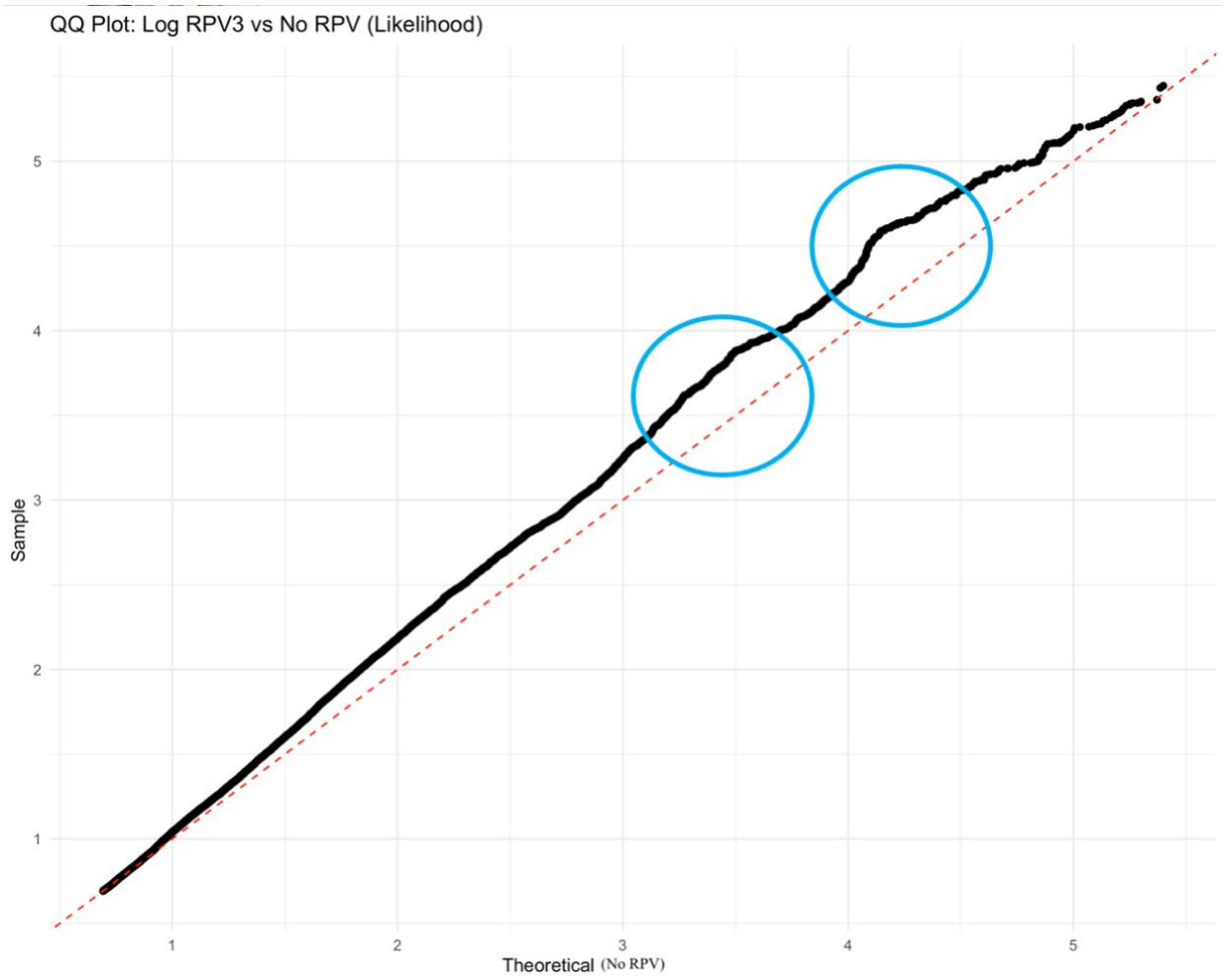

Figure 14: The qq plot for the log of likelihoods of *P. viticola* samples the grew on RPV3.1 vines (y-axis) versus the log of likelihoods of *P. viticola* samples that grew on no RPV (x-axis). Since there were 6 quantiles, each one represents the 16.7 x n percentile of the data where n was quantile number.

For samples grown on RPV10.3, the log likelihoods were visualized in Q-Q plots (Figure 15). Between quartile 4.75 and quartile 5.5 (79-91st percentile) was extreme divergence from the log transformed likelihoods of the susceptible samples. Because the divergence was above 1:1 comparison, the likelihood scores were greater on samples grown on RPV10.3 vines than with samples grown on No RPV. There was a shift below 1:1 comparison around part 2-3 (33-50th percentile) but was determined not to be significant. If divergence *was* below and considered significant, the likelihood scores for that range would be lower than expected.

To further support this, the log transformed likelihoods of RPV10.3 resistance samples were compared with those of no RPV susceptible samples (Figure 17), resulting in extreme significance ($p < 0.001$) and verifying that these regions were not brought about by random chance.

The sweep data for samples grown on RPV3.1+RPV10.3 were visualized in Q-Q plots (Figure 16). Unlike the results described above for RPV3.1 resistant and RPV10.3 resistant samples, RPV3.1+RPV10.3 resistant samples significantly deviate from no rpv susceptible at 1.5 (25^{th} percentile and above) onward. This could suggest a compound resistance mechanism as opposed to the centralized predicted mechanisms of RPV3.1 resistant samples and RPV10.3 resistant samples. Because the divergence is above the 1:1 comparison, this indicated that the log likelihood values are above the expected threshold. The box plot in Figure 17 also shows significant deviation ($p < 0.001$) of the log transformed likelihoods of RPV3.1+RPV10.3 resistant samples from the log transformed likelihoods of no RPV susceptible.

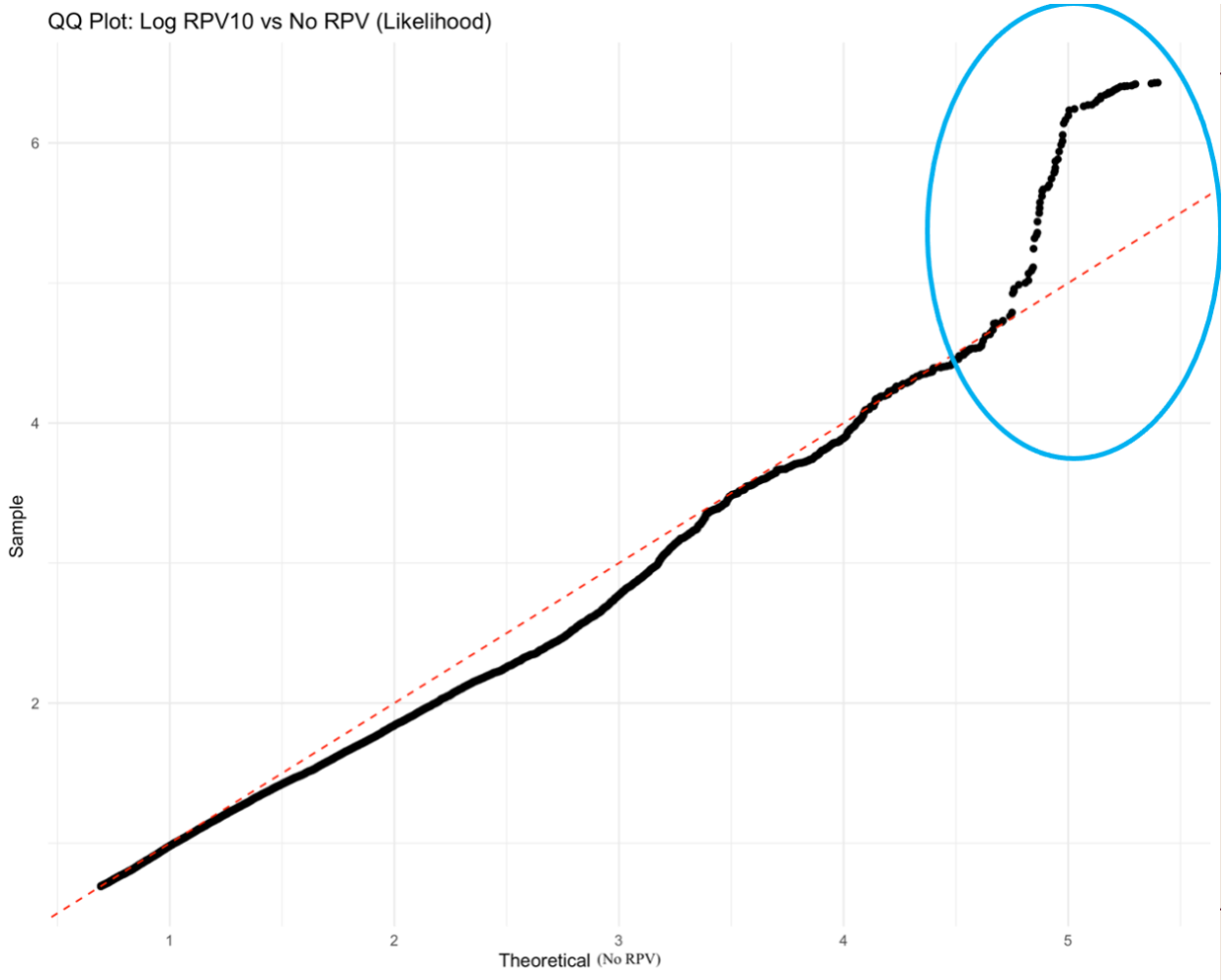

Figure 15: The Q-Q plot for the log of likelihoods of *P. viticola* samples that grew on RPV10.3 vines (y-axis) versus the log of likelihoods of *P. viticola* samples that grew on no RPV (x-axis). Since there were 6 quantiles, each one represents the 16.7 x n percentile of the data where n was quantile number.

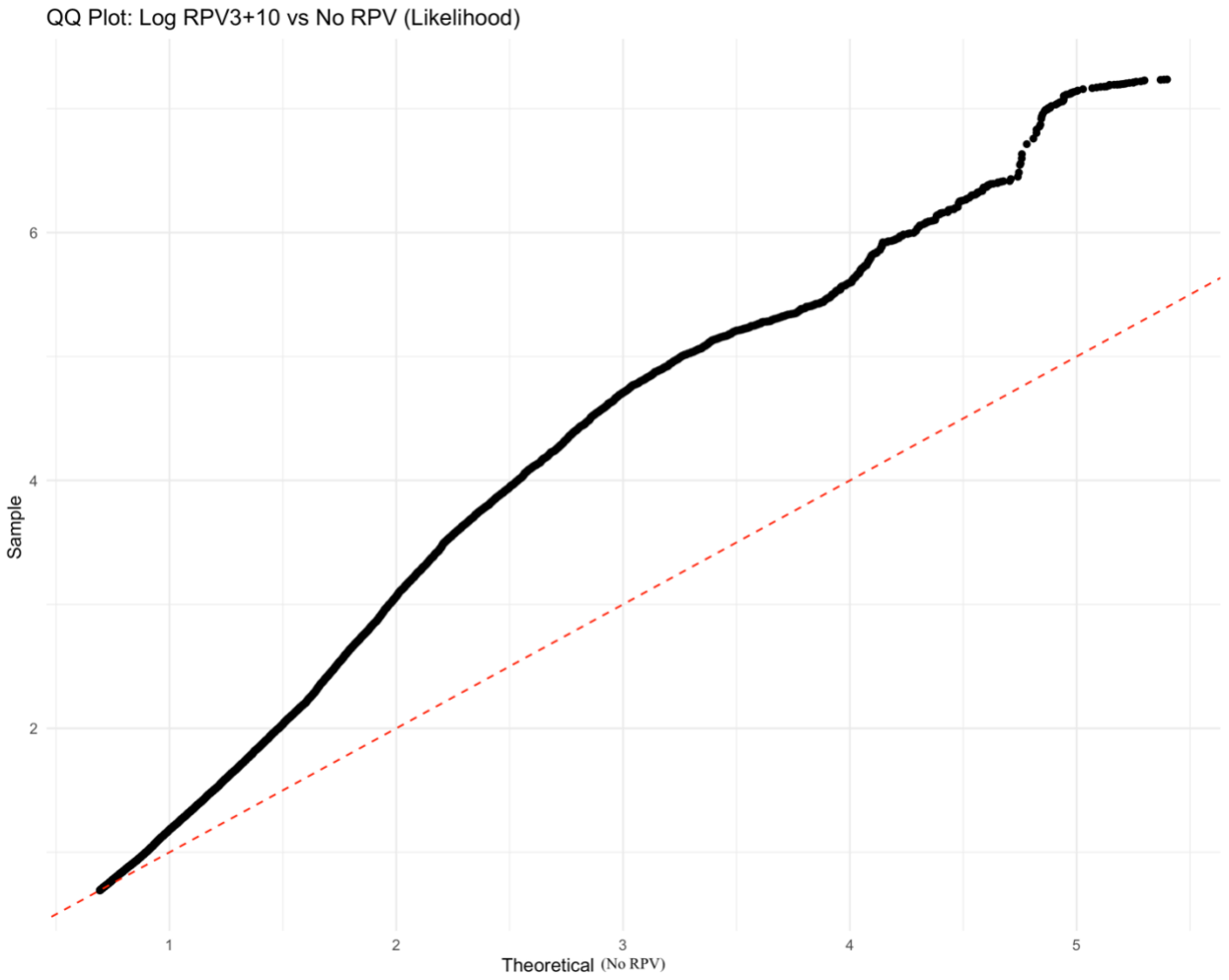

Figure 16: The Q-Q plot for the log of likelihoods of *P. viticola* samples that grew on RPV3.1+RPV10.3 vines (y-axis) versus the log of likelihoods of *P. viticola* samples that grew on no RPV (x-axis). Since there were 6 quantiles, each one represents the 16.7 x n percentile of the data where n was quantile number.

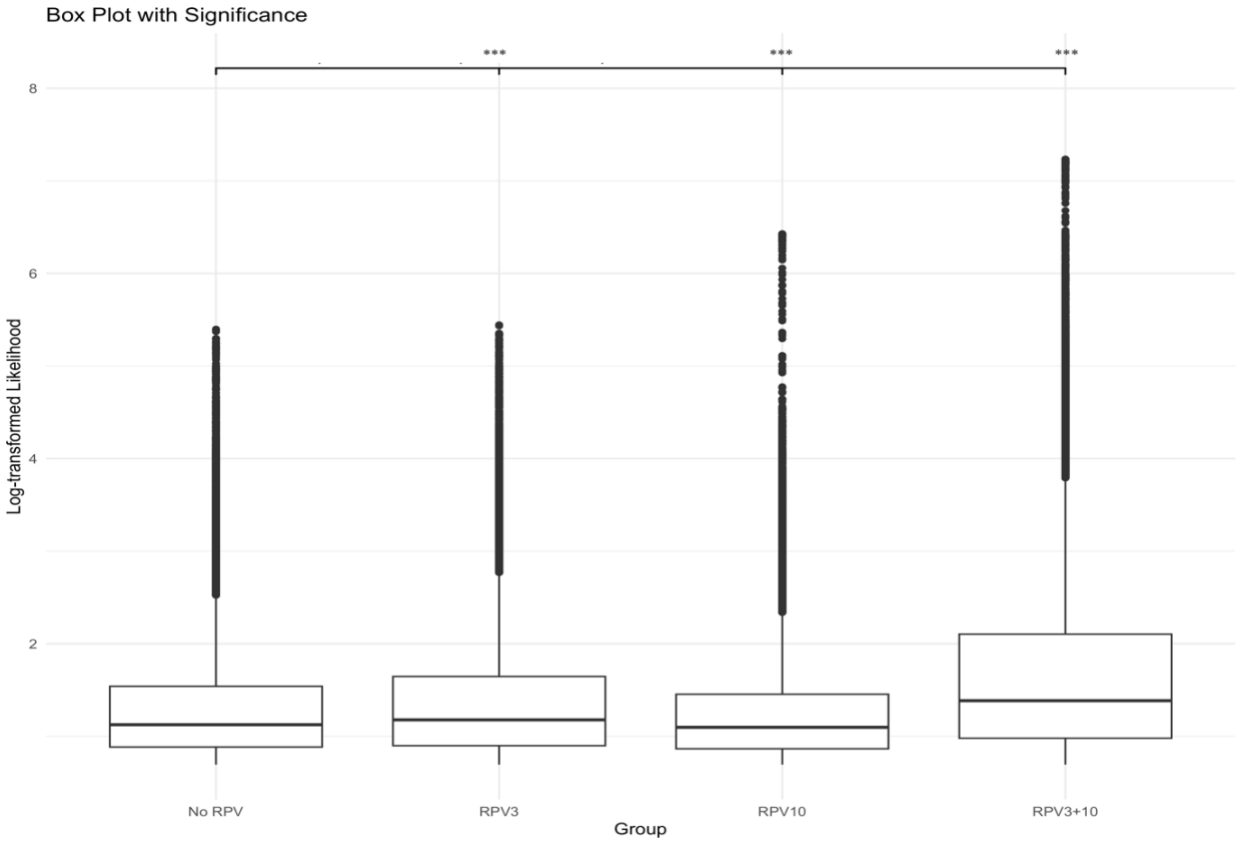

Figure 17: The four different groups tested along with their log transformed likelihoods. The asterisks above each of the groups represented the significance level of a Wilcoxon rank-sum test when compared with the log transformed likelihoods of the no RPV group.

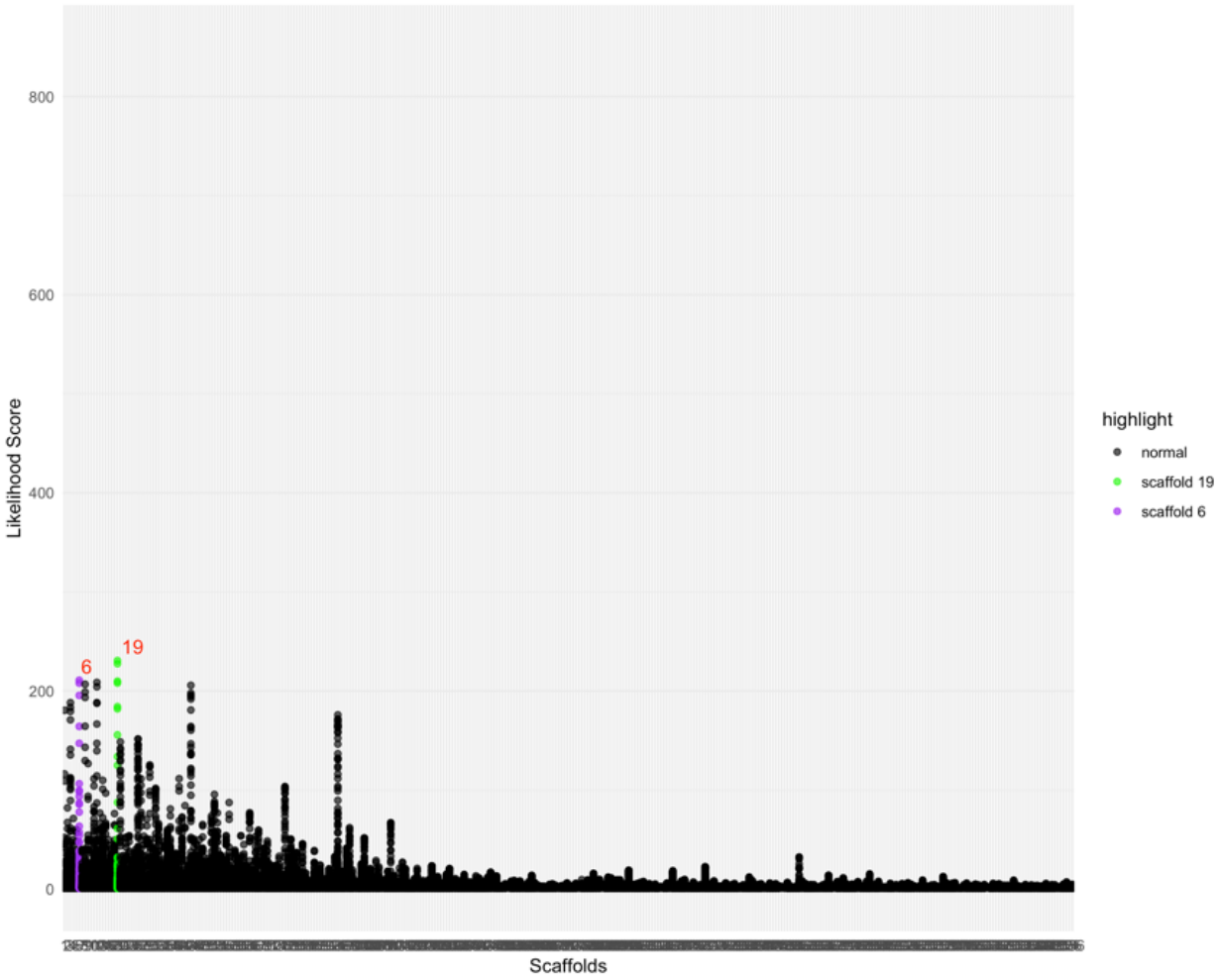

Figure 18: A genome-wide look at how the likelihood scores were distributed across the 358 scaffolds. The top scaffolds were highlighted. This plot was representative of RPV3.1 grown samples.

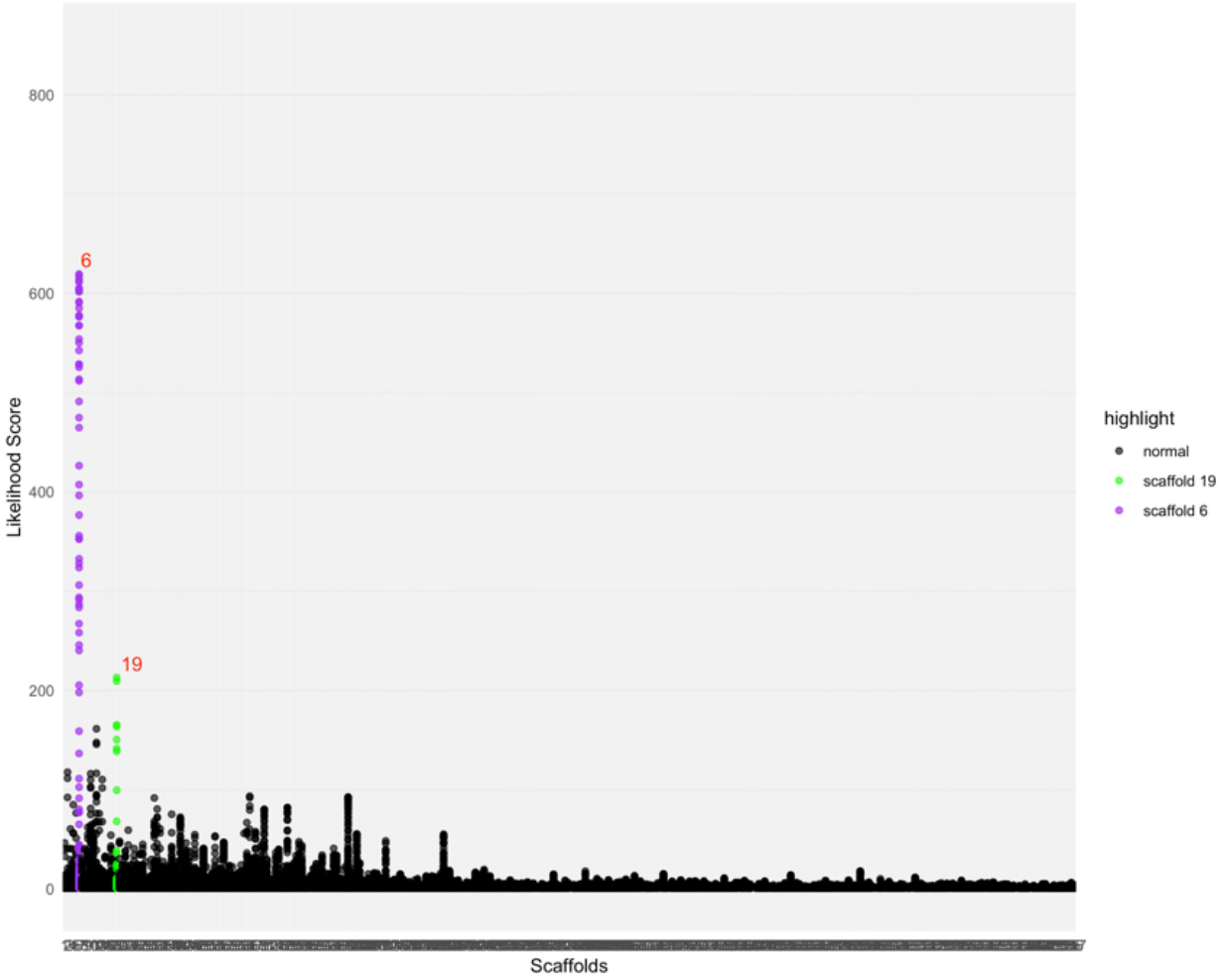

Figure 18 (Continued): A genome-wide look at how the likelihood scores were distributed across the 358 scaffolds. The top scaffolds were highlighted. This plot was representative of RPV10.3 grown samples.

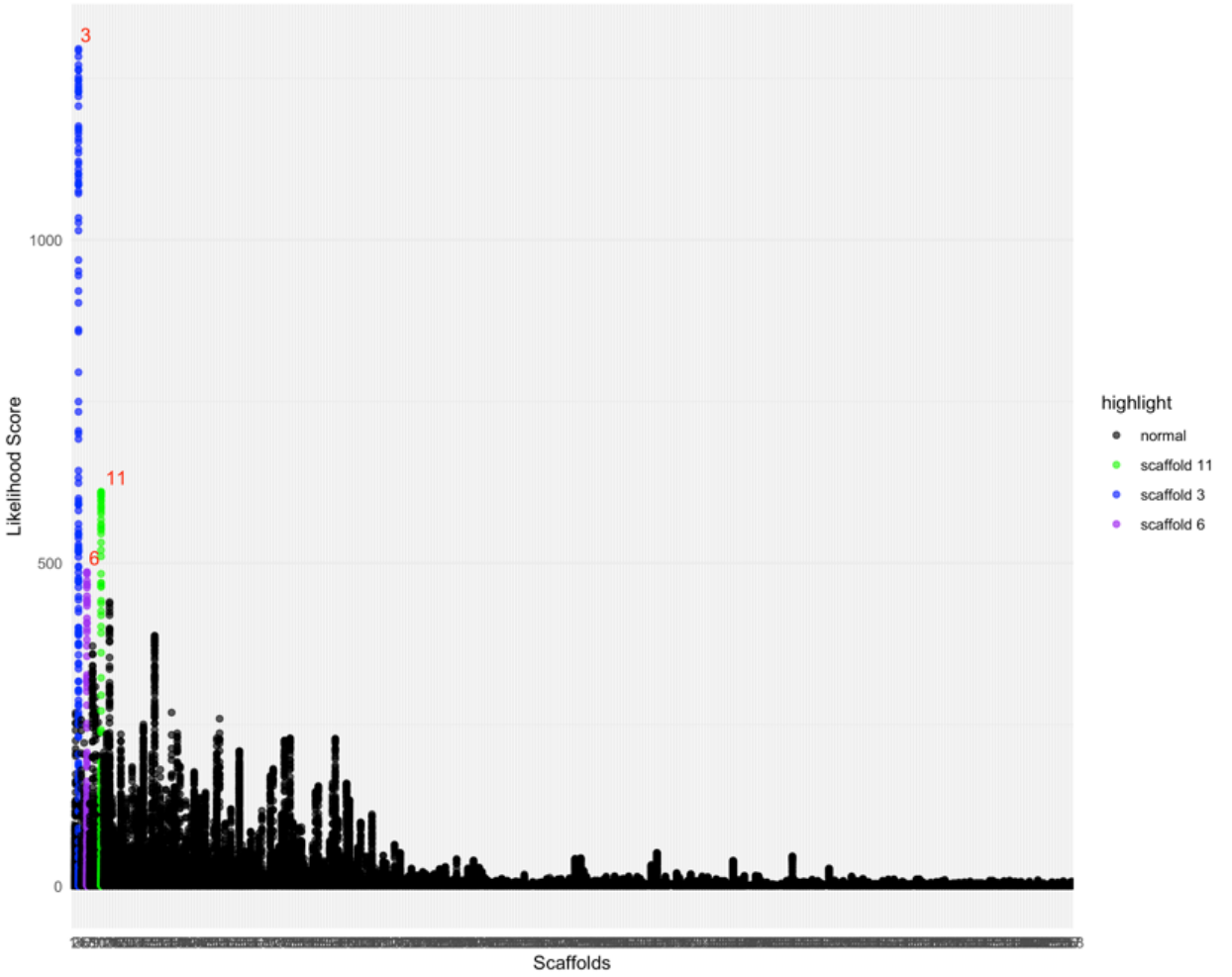

Figure 18 (Continued): A genome-wide look at how the likelihood scores were distributed across the 358 scaffolds. The top scaffolds were highlighted. This plot was representative of RPV3.1+10.3 grown samples.

The likelihood scores from across the entire RPV3.1, RPV10.3, and RPV3.1+10.3 grown samples' genomes were plotted (Figure 18). The top scatterplot showed the results from RPV3.1 grown samples. The highest peaks were for scaffold 19 and 6 with likelihood scores around 200. However, these top two peaks were not as distinct as the peaks observed in RPV10.3 and RPV3.1+10.3, resulting in difficulty determining if these scaffolds were under selection. The middle scatter plot shows the results from RPV10.3 grown samples. Scaffold 6 had the highest likelihood scores for RPV10.3 at around 600, which suggested this scaffold may have a selective sweep as the next highest scaffold had a 400 likelihood score difference. Rpv3.1+10.3 grown samples also contained scaffold 6 as one of the scaffolds having a selective sweep with the differences in likelihood score being 100. Because this scaffold appeared to have a selective sweep in both RPV10.3 and RPV3.1+10.3 and possibly in RPV3.1, it can be hypothesized that this scaffold contained elements that are linked to general resistance in *P. viticola* or at the very least, resistance mechanisms in RPV3.1 and RPV10.3. Within RPV3.1+10.3 samples, scaffolds 3 and 11 appeared to have a selective sweep due to their extremely high likelihood scores of around 1,000 and 650 respectively, indicating that these two scaffolds contained genes that are critical for overcoming the combined resistance rather than individual resistances.

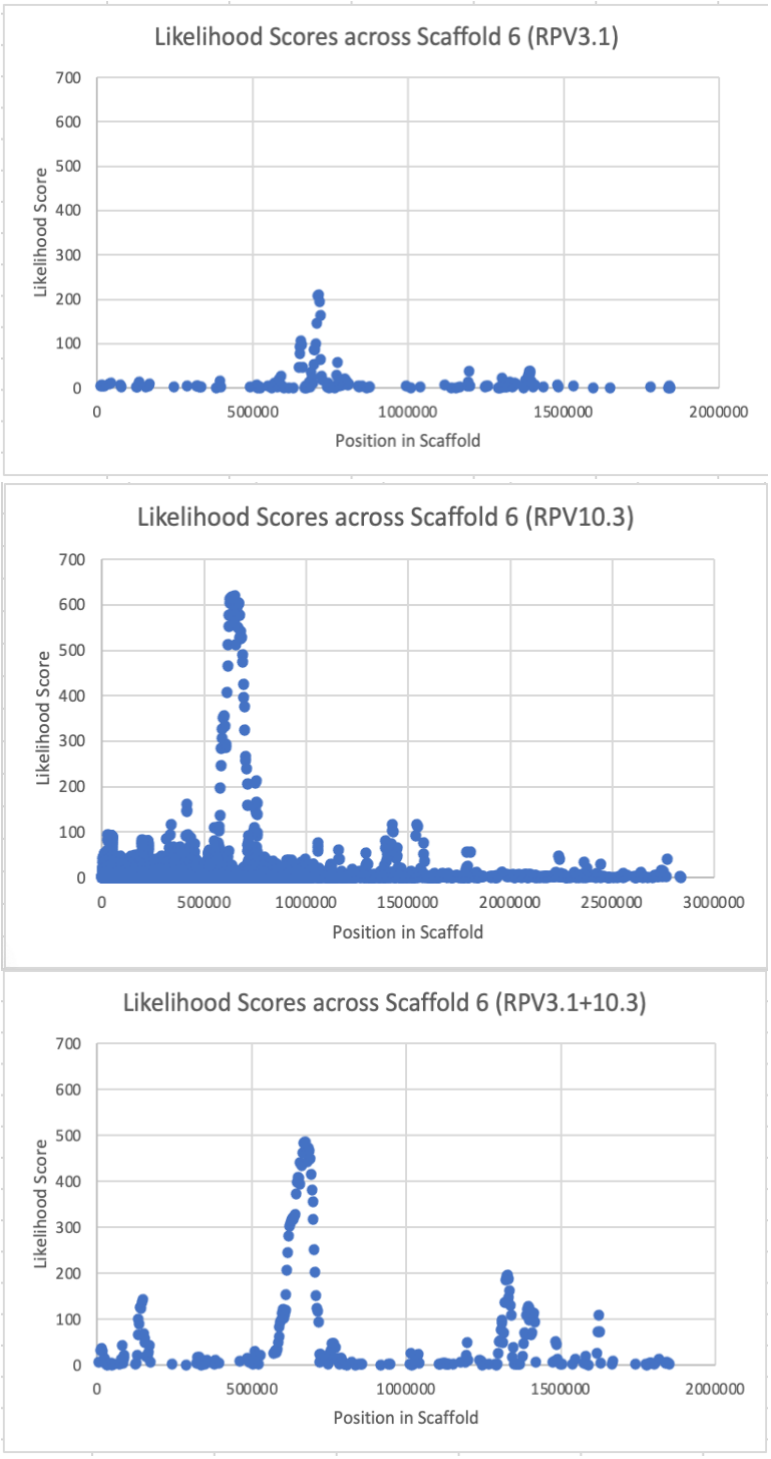

Figure 19: The likelihood scores distributed across scaffold 6 for RPV3.1 (top), RPV10.3(middle), and RPV3.1+RPV10.3 (bottom).

Due to all three host resistances containing relatively high likelihood scores on scaffold 6, their likelihood scores across this scaffold were plotted in scatter plots (Figure 19). The areas of a selective sweep occured in the same region (~510,000-750,000) of the scaffold for RPV10.3 (upper right corner) and RPV3.1+10.3 (bottom center). With RPV3.1 samples, it was difficult to tell if the small peak was under true selection. The linkage disequilibrium was also evaluated on scaffold 6 (Figure 20). The r^2 statistic was used with a score of 0.8 or greater applied to determine the strength of the LD. The area was the area identified as. Within the site of the sweep, a strong LD (>0.75) is observable towards the 750,000 bp position, which was within the sweep region (circled in red in Figure 20). Moving away from the site of the sweep, a decrease in LD was evident. Because RPV3.1 was difficult to determine if a selective sweep was occurring, it was excluded from further LD analysis, gene annotation, and protein prediction. Scaffold 3 also contained a significant sweep region, and the LD was assessed (Figure 21). Unlike with the scaffold 6 LD, this LD pattern was difficult to determine as there was not a clear "high" region with the exception at position ~910,000. The site of the sweep was circled in red. Scaffold 11 was also identified as significant in RPV3.1+10.3. The LD pattern was evident as the LD increased towards the end of the sweep region (Figure 22) with the selective sweep site being identified by the red box. Around position 200,000, LD was above 0.75, indicating strong LD.

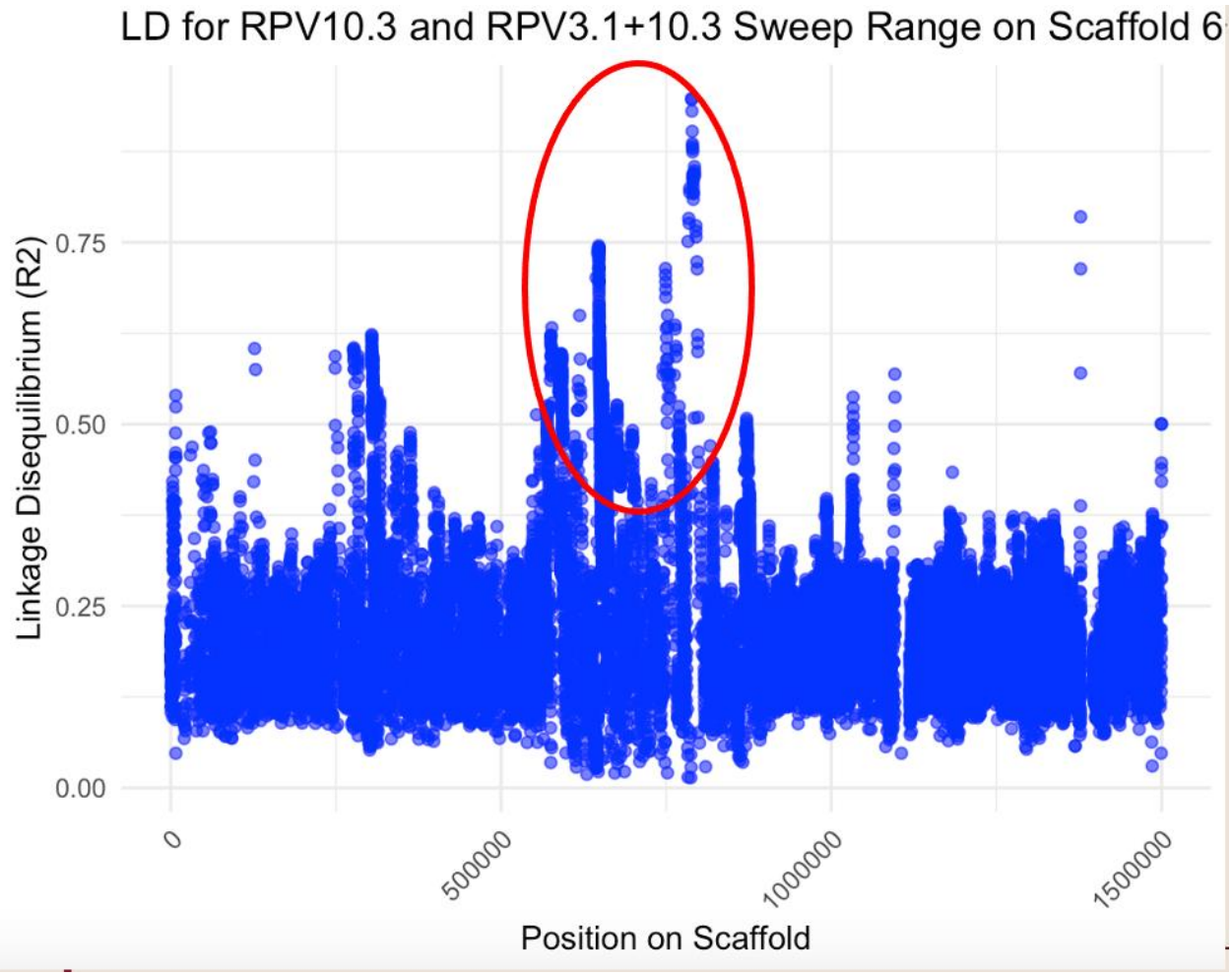

Figure 20: The linkage disequilibrium (LD) calculated using VCFtools for scaffold 6 in both the RPV10.3 and RPV3.1+RPV10.3 grown samples. The red circle corresponded to the sweep region.

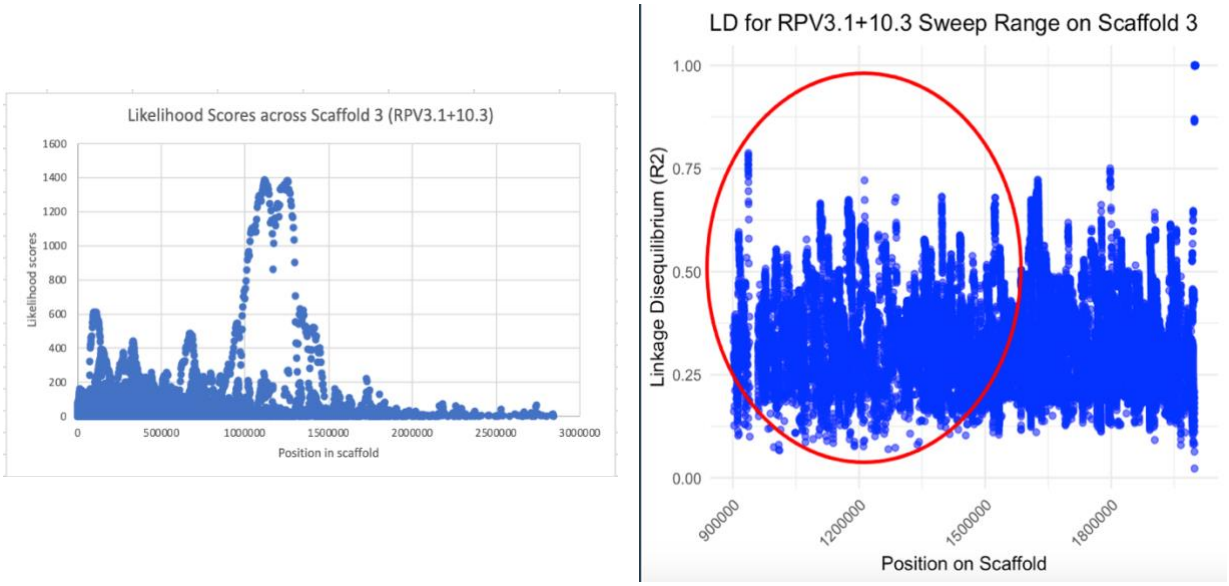

Figure 21: The likelihood scores across scaffold 3 (left), and the LD scores across scaffold 3 (right) for RPV3.1+10.3 grown samples. The red circle corresponded to the sweep region.

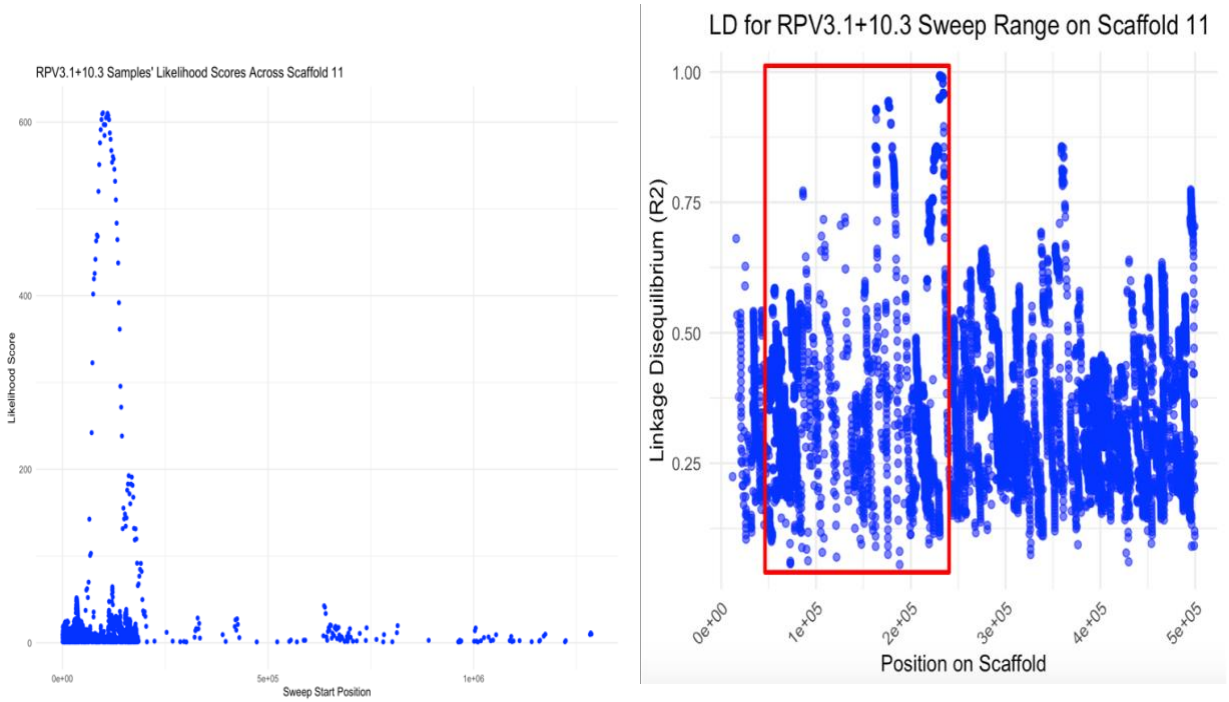

Figure 22: The likelihood scores across scaffold 11 (left), and the LD scores across scaffold 11 (right) for RPV3.1+10.3 grown samples. The red square corresponded to the sweep region.

The genes present in the sweep regions (Figure 18) were assessed using the protein sequences generated from Augustus (Stanke et al., 2006) and were recorded (Tables 6 and 7). Several of these proteins have general roles in pathogenesis but may not explain the specific virulence on RPVs that led to a selective sweep. For example, a glycoside hydrolase was found and predicted to have cellulase activity. Cellulase are responsible for cellulose degradation (Ejaz, Sohail, and Ghanemi, 2021), which is a major component of the cell wall. By degrading this, *P. viticola* may be able to penetrate the plant tissue more effectively and gain access to nutrients. 39S ribosomal protein are needed for protein synthesis including any protein involved in virulence or stress response. Heat shock transcription factor is likewise involved with the stress response (Tran Thi Ngoc et al., 2023). In *P. viticola,* this could mean adapting to the grapevine's stress response to survive in varying conditions.

Several of the proteins have logical roles in the specific host-pathogen interactions, such as secreted RxLR effectors. To establish an infection, oomycetes deploy effector proteins with the largest group being RxLR effectors. In an experiment by Lan et al., it was found that PvRxLR targets leucine-rich repeat receptor-like kinases (LRR-RLKs)-associated inhibitor BKI1 for promoting infection. Since LRR-RLKs control downstream resistance responses, targeting this is effective for disrupting the corresponding signaling (Lan et al., 2019). Through EffectorP, this proposed RxLR factor (PVIT_0003606) was identified as a cytoplasmic effector. This type of effector is delivered into the host cell's cytoplasm. Once inside, the cytoplasmic effector can suppress the plant's defense responses as well as alter the host cell structure and function (Presti et al., 2015). Apoplastic effectors, on the other hand are delivered into the apoplast, and function to neutralize the host's defense mechanism as well as protect the pathogen from the plant's defense response (Rocafort et al., 2020). Finally, highlighted in yellow was a predicted

hypothetical protein (PVIT_0003584), however, after putting the protein sequences through SignalP, this was identified as a potential signal protein with the Signal Peptide (Sec/SPI) being 0.5362. These short signal peptides (usually between 15-50 amino acid residues in length) are usually present at the N-terminus of proteins that are directed towards the secretory pathway. As the name suggests, these proteins are involved with transmitting information to regulate multiple biological processes, specifically carrying information for protein secretion. This protein may have played a role in the overcoming of resistance as it is in a region identified to be under a selective sweep. Overall, due to these proteins being in areas predicted to have selective sweeps and therefore strong, positive selection, it can be inferred that the proteins may be important for *P. viticola's* adaptation and interaction with the host grapevines.

Gene ID	Gene Start	Gene End	Predicted Protein Name	Predicted Protein Function	Species
PVIT_0003564	1440310	1441401	Glycoside hydrolase family	Cellulase activity; Polysaccharide catabolism	Phytophthora sojae
PVIT_0003572	1468861	1469754	39S ribosomal protein mitochondrial like	Structural molecule activity	Plasmopara halstedii
PVIT_0003574	1471063	1471773	Ankyrin repeat-containing domain	Unknown	Phytophthora cinnamomi
PVIT_0003576	1474766	1476172	Heat Shock transcription factor	RNA binding; transcription regulator activity	Plasmopara halstedii
PVIT_0003584	1503969	1504889	Hypothetical Protein	Unknown	-
PVIT_0003585	1505052	1506674	Phospholipid hydro peroxide glutathione	Unknown	Plasmopara halstedii
PVIT_0003589	1527909	1528927	Carbonic	Lyase activity; zinc ion binding	Plasmopara halstedii
PVIT_0003594	1560715	1574597	Dynein heavy chain C-Terminal domain	ATP binding; chain binding; microtubule motor activity	Phytophthora infestans
PVIT_0003596	1589893	1590759	Abyhydrolase domain-containing protein	Lyase activity	Phytophthora nicotianae

Table 6: The gene IDs, their predicted encoded protein, the predicted function, and the species of the top protein match for genes located in RPV10.3 and RPV3.1+10.3's scaffold 6. The protein data was collected from Uniprot (Bateman et al., 2022).

PVIT_0003597	1590962	1592200	Glutathione peroxidase	Unknown	Phytophthora cactorum
PVIT_0003600	1603622	1604830	Oxygen-dependent coproporphyrinogen-III oxidase	Potential oxidoreductase activity	Phytophthora citrophthora
PVIT_0003606	1630255	1631265	RxLR effector protein	Virulence factor; suppress basal immunity	Phytophthora cinnamomi
PVIT_0003607	1638683	1637895	Cysteine-rich protein	ATP binding; metal ion binding; unfolded protein binding	Phytophthora nicotianae
PVIT_0003608	1638638	1640614	Leucine-rich repeat domain	Signaling	Phytophthora cactorum

Table 6 (continued)

Avirulence (Avr) proteins represent effectors with a function initially discovered by their triggering a successful defense response, thereby their presence makes the pathogen avirulent on the host (Du et at., 2018). Two homologues of avirulence proteins (PVIT_0002214 and PVIT_0002215) were found in a high likelihood area for scaffold 3 selective sweeps. These proteins may be evolving rapidly to avoid detection by the resistance mechanisms (i.e. RPV3.1+RPV10.3). Other proteins have general roles in pathogenesis but may not explain the specific virulence on RPVs. The osmotic stress-sensitive mutant 1 is involved with helping the organism's response to changes in osmotic pressure. This suggests that in *P. viticola*, this protein may be important for managing stress during infection as well as changed in environmental conditions, and due to its place in the high likelihood sweep region, this may be crucial for adaptation and survival of *P. viticola*. S-M checkpoint control protein is involved in RNA processing and modification. Since this protein is in a high likelihood sweep region, it may play a role in regulating genes for adaptation to the host's environmental and stress conditions. Finally, tRNA methyltransferase is involved in the modification of tRNA (Björk et al., 2001). This protein may be undergoing selection due to its role in adapting to the host's defense or other

environmental pressures. To confirm these functions, further research will have to be done. Finally, for scaffold 11, there were not any RxLR effector proteins or avirulence proteins predicted, but there was a signal peptide predicted in the area where a hypothetical protein was originally predicted (Table 8).

Gene ID	Gene Start	Gene End	Predicted Protein Name	Predicted Protein Function	Species
PVIT_0002213	2569340	2570026	Urease accessory protein	Nickel cation binding	Phytophthora nicotianae
PVIT_0002214	2570206	2576157	Avirulence (Avr) protein	pathogenicity	Phytophthora megakarya
PVIT_0002215	2576249	2576842	Avirulence (Avr) protein	pathogenicity	Phytophthora megakarya
PVIT_0002216	2577167	2578950	Osmotic stress-sensitive mutant 1	Unclear	Plasmopara halstedii
PVIT_0002218	2579517	2582756	S-M checkpoint control protein CID1 and related nucleotidyltransferases	Transferase activity; catalytic activity, acting on RNA	Plasmopara halstedii
PVIT_0002228	2643859	2644656	tRNA (guanine-N(7)) methyltransferase	Catalytic activity	Phytophthora citrophthora

Table 7: The gene IDs, their predicted encoded protein, the predicted function, and the species the top protein match for genes located in RPV3.1+RPV10.3 scaffold 3. The protein data was collected from Uniprot (Bateman et al., 2022).

Gene ID	Gene Start	Gene End	Predicted Protein	Species
PVIT_0006058	369087	370133	Hypothetical Protein	-

Table 8: Scaffold 11 gene prediction and protein prediction for RPV3.1+10.3

Serial Passaging Outcome

To mimic the field sample disease progress and analysis, serial passaging was carried out. However, despite 46 samples successfully sequenced, the Illumina reads for the serial passaging samples were unable to be adequately mapped to the reference genome. The average mapping percentage was 2.5% mapped, which indicates very poor mapping. Due to the low number of mapped reads, the unmapped reads were put through blast to diagnose the issue. The top results were *Strenotrophomonas rhizophila*, *Microbacterium testaceum*, *Microbacterium* sp. (which is a common lab contaminant), and *P. viticola* mitochondrion. One potential reason why these species appeared in the reads could do with how the spores were transferred from leaf to leaf. Using a spray bottle to mist the leaves and allow the water to drip into a tube which is then used to spray a new leaf may have introduced these species seeing as the cluster of spores was already small on the first generation.

Discussion

Genome Assembly Comparison

In the assembly for this research, 6 kb fragments were sequenced after an attempt was made to increase fragment size during the initial library preparation. After conferring with INRA-PV221 reference, it was found they used fragments above 5 kb (Dussert et al., 2018), which this fit into. There was also an estimated 14x coverage for all four haplotypes, which is relatively low. A deeper coverage is being generated with left over DNA at the PacBio facility. The samples that the single spore was extracted from were in the freezer for at least a month before the extraction, which could cause a degradation of the quality. Using fresh samples from the field and continuing to build up the biomass is hypothesized to improve quality. Assembling a genome of a virulent isolate would help identify structural variation in the regions under selective sweep, in case there are large insertions or deletions involved. Generating RNA-seq

data would also have been beneficial and is a procedure that should be produced in the future. RNA-seq data would help with gap filling, identifying coding regions of the genome, and refine gene models. This assembly can provide novel insight into genes present and how they differ from the published reference's, especially in regard to the avrRpv3.1 locus (Paineau et al., 2023) that is important in the pathogen's resistance mechanism.

Field Sample Analysis

The overarching goal of this research was to provide an understanding into how *P. viticola* adapts to the grapevines' host resistant genes with a prediction that there will be linkage disequilibrium (fixed alleles) around the loci responsible for virulence and random assortment of alleles moving away from the loci. It was also predicted that selective sweeps patterns would differ response to different host resistant genes.

The SweeD analysis of samples grown on RPV3.1, RPV10.3, and RPV3.1+10.3 provided insight into the regions of the pathogen's genome that are undergoing selective sweeps in response to certain RPVs such as scaffold 3 in RPV3.1+10.3 grown samples and scaffold 6 in RPV10.3 grown samples. However, scaffold 6 also appears to show evidence of a selective sweep in RPV3.1+10.3 grown samples, and further investigation determined that the annotated genes showing selective sweeps on this scaffold for this RPV were the exact same annotated genes showing selective sweeps for RPV10.3 on this scaffold, suggesting strong evidence that this region is important for overcoming the RPV10.3 resistance.

The specific regions under selection were identified as a 121 kb region on scaffold 3 and 230 kb region on scaffold 6, further pinpointing where the selective pressure is concentrated. Within these regions, scaffold 6 contains a gene predicted to encode an RxLR effector protein as

well as a gene predicted to encode for a signal protein while scaffold 3 has two genes predicted to encode avirulence (Avr) proteins. RxLR effector proteins have a wide range of modes of action including stabilization, destabilization of target protein targets, inhibition, use of host protein enzyme activity, disruption of host protein complexes, and changing of localization targets (Wang et al., 2023). Based on the findings of this research alone, the exact function of the RxLR effector within scaffold 6 is relatively unclear. Further experimental validation is necessary to confirm this exact functions. Some RxLR proteins are avirulence (Avr) proteins as well. Two Avr proteins were predicted on scaffold 6 in RPV10.3 grown samples, and similar to RxLR effectors, they are part of a pathogen's pathogenicity. Since the scaffold 6 genes in RPV3.1+10.3 are the same as the RPV10.3 genes, this suggests that these Avr proteins could be part of the mechanism for overcoming RPV10.3 resistance. Further wet lab work would have to be done to confirm that these genes encode for this protein to conclusively state that. Because strong pressure on scaffold 3 is only present in RPV3.1+10.3 grown samples as RPV3.1 grown samples do not have a distinctive region of selective sweep signals and RPV10.3 grown samples have scaffold 6, this selective pressure can be hypothesized to be brought about by the stacked host resistance. Furthermore, the presence of other moderate sweep signals as seen in Figure 15 by the slight peaks across the genome suggests the RPV3.1+10.3 has less genetic variation around the first third of the genome. It would be interesting to see if these regions become more pronounced with selective pressure if more generations of sporulation were included. RPV10.3 grown samples can be hypothesized to have more genetic variation as the sweep signals are low. Overall, these findings support the original hypothesis of selective sweeps being concentrated in one loci while regions further away from the selective sweeps shows reduced evidence of sweeps, proposing a complex mechanism for overcoming RPVs in *P.viticola* . Because of this

support, this can provide insight into methods for developing targeted measures for disrupting the resistance mechanisms in *P. viticola*. However, longer term monitoring may be beneficial in identifying the pathogen's long term adaptive mechanism, which may differ from these findings. Furthermore, devising a wet lab procedure for gene knockout in regions proposed to have proteins that are relative to pathogenicity could provide further confirmation of these predictions. It would also be beneficial to have RNA-seq data to analyze gene expression changes as well as also monitoring protein changes.

Conclusion

This research provides new insight into how *P. viticola* adapts to three different grapevine resistant genes. Important findings include the identification of RPV3.1+10.3 and RPV10.3 grown samples' selective sweep likelihood scores as having extreme divergence from the susceptible samples' selective sweep likelihood scores and identifying scaffold 3 and scaffold 6 as regions under significant selective pressure with scaffold 6 seeming to be essential for overcoming RPV10.3 resistance.

Future improvements such as using higher quality DNA, assembling a genome of a virulent isolate, and generating RNA-seq data would improve the assembly to conclude more sound interpretations of genetic differences as well as providing more insight into the field samples analyzed. Functional testing of the candidate genes is needed.

Overall, this research advances the understand of *P. viticola's* adaptation mechanisms and offers a potential foundation for developing targeted strategies to manage the resistance. Future work should thereby focus on confirming the proteins predicted to be encoded and explore the impact of long-term sporulation to determine if the same loci remain under strong selective pressure over many generations.

References

1. Almagro Armenteros, José Juan, et al. “SIGNALP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks.” *Nature News*, Nature Publishing Group, 18 Feb. 2019, www.nature.com/articles/s41587-019-0036-z.
2. Almeida A, Loy A, Hofmann H. “ggplot2 Compatible Quantile-Quantile Plots in R*” The R Journal*, volume 10 number 2, 2018, https://doi.org/10.32614/RJ-2018-051.
3. Altschul, Stephen F, et al. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology*, Academic Press, 6 Feb. 2007, www.sciencedirect.com/science/article/abs/pii/S0022283605803602.
4. Andrews, S. “FastQC: A Quality Control for High Throughput Sequencing Data.” *Babraham Bioinformatics - FastQC a Quality Control Tool for High Throughput Sequence Data*, 2012, www.bioinformatics.babraham.ac.uk/projects/fastqc/.
5. Bateman, Alex, et al. “Uniprot: The Universal Protein Knowledgebase in 2023.” *OUP Academic*, Oxford University Press, 21 Nov. 2022, academic.oup.com/nar/article/51/D1/D523/6835362?login=false.
6. Bellin, Diana, et al. “Resistance to Plasmopara Viticola in Grapevine ‘Bianca’ is controlled by a major dominant gene causing localised necrosis at the infection site.” *Theoretical and Applied Genetics*, vol. 120, no. 1, 11 Oct. 2009, pp. 163–176, https://doi.org/10.1007/s00122-009-1167-2.
7. Bendtsen, J. D., et al. “Feature-based prediction of non-classical and leaderless protein secretion.” *Protein Engineering Design and Selection*, vol. 17, no. 4, 4 May 2004, pp. 349–356, https://doi.org/10.1093/protein/gzh037.

8. Björk, G R, et al. “A Primordial Trna Modification Required for the Evolution of Life?” *The EMBO Journal*, U.S. National Library of Medicine, 15 Jan. 2001, www.ncbi.nlm.nih.gov/pmc/articles/PMC140193/.
9. Brischetto C, Bove F, Fedele G, Rossi V. A Weather-Driven Model for Predicting Infections of
10. Bushnell, Brian. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. United States: N. p., 2014. Web.
11. Grapevines by Sporangia of *Plasmopara viticola*. Front Plant Sci. 2021 Mar 9;12:636607. doi: 10.3389/fpls.2021.636607. PMID: 33767721; PMCID: PMC7985336.
12. Browning, Brian L, et al. “Fast Two-Stage Phasing of Large-Scale Sequence Data.” *American Journal of Human Genetics*, U.S. National Library of Medicine, 7 Oct. 2021, www.ncbi.nlm.nih.gov/pmc/articles/PMC8551421/.
13. Camargo, M. P., et al. “Cryptic species and population genetic structure of *plasmopara viticola* in São Paulo State, Brazil.” *Plant Pathology*, vol. 68, no. 4, 19 Feb. 2019, pp. 719–726, https://doi.org/10.1111/ppa.12993.
14. Campbell, Sarah E, et al. “Efficacy of Fungicide Treatments for Plasmopara Viticola Control and Occurrence of Strobilurin Field Resistance in Vineyards in Georgia, USA.” *Crop Protection*, Elsevier, 2 Sept. 2020, www.sciencedirect.com/science/article/abs/pii/S0261219420303045.
15. Cheng, Haoyu, et al. “Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm.” *Nature News*, Nature Publishing Group, 1 Feb. 2021, www.nature.com/articles/s41592-020-01056-5.

16. Danecek, Petr, et al. “Twelve Years of SAMtools and BCFtools.” *OUP Academic*, Oxford University Press, 16 Feb. 2021, academic.oup.com/gigascience/article/10/2/giab008/6137722.
17. Dobin, Alexander, et al. “Star: Ultrafast Universal RNA-Seq Aligner.” *OUP Academic*, Oxford University Press, 25 Oct. 2012, academic.oup.com/bioinformatics/article/29/1/15/272537.
18. Du, Y, et al. “RXLR Effector Diversity in *Phytophthora Infestans* Isolates Determines Recognition by Potato Resistance Proteins; the Case Study AVR1 and R1.” *Studies in Mycology*, U.S. National Library of Medicine, Mar. 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC6002335/.
19. Dussert, Yann, et al. “A High-Quality Grapevine Downy Mildew Genome Assembly Reveals Rapidly Evolving and Lineage-Specific Putative Host Adaptation Genes.” *OUP Academic*, Oxford University Press, 7 Mar. 2019, academic.oup.com/gbe/article/11/3/954/5371071.
20. Dussert, Yann, et al. “Genome assembly and annotation of Plasmopara viticola, the grapevine downy mildew pathogen”, *Rèpublique Française*, Recherche Data Gouv, 16 Apr. 2018, https://doi.org/10.15454/4NYHD6.
21. Ejaz, Uroosa, et al. “Cellulases: From Bioactivity to a Variety of Industrial Applications.” *Biomimetics (Basel, Switzerland)*, U.S. National Library of Medicine, 5 July 2021, www.ncbi.nlm.nih.gov/pmc/articles/PMC8293267/.
22. Fontaine, Michael C, et al. “Europe as a Bridgehead in the Worldwide Invasion History of Grapevine Downy Mildew, Plasmopara Viticola.” *Current Biology*, Cell Press, 25 Mar. 2021, www.sciencedirect.com/science/article/abs/pii/S0960982221003481.

23. Fontaine, Michael C., Fréderic Austerlitz, et al. “Genetic signature of a range expansion and leap-frog event after the recent invasion of Europe by the grapevine downy mildew pathogen *plasmopara viticola*.” *Molecular Ecology*, vol. 22, no. 10, 18 Mar. 2013, pp. 2771–2786, https://doi.org/10.1111/mec.12293.

24. Ford, Clay. “Understanding QQ Plots.” *UVA Library*, 15 Aug. 2015, library.virginia.edu/data/articles/understanding-q-q-plots.

25. Garrison, Erik, and Gabor Marth. “Haplotype-Based Variant Detection from Short-Read Sequencing.” *arXiv*, 20 July 2012, arxiv.org/abs/1207.3907.

26. Gobbin, D., et al. “Evidence for sporangial dispersal leading to a single infection event and a sudden high incidence of grapevine downy mildew.” *Plant Pathology*, vol. 56, no. 5, 25 July 2007, pp. 843–847, https://doi.org/10.1111/j.1365-3059.2007.01651.x.

27. Jones, David S., and Patricia S. McManus. “Distinctive symptoms and signs of downy mildew on cold-climate wine grape cultivars.” *Plant Health Progress*, vol. 18, no. 3, 1 Jan. 2017, pp. 192–195, https://doi.org/10.1094/php-01-17-0009-dg.

28. Kennelly, Megan & Gadoury, David & Wilcox, Wayne & Magarey, Peter & Seem, Robert. (2006). Seasonal Development of Ontogenic Resistance to Downy Mildew in Grape Berries and Rachises. Phytopathology. 95. 1445-52. 10.1094/PHYTO-95-1445.

29. Kim, B. R., et al. “First report of downy mildew caused by *plasmopara viticola* on *vitis coignetiae* in Korea.” *Plant Disease*, vol. 103, no. 7, July 2019, p. 1793, https://doi.org/10.1094/pdis-02-19-0317-pdn.

30. Kokot, Marek, et al. “KMC 3: Counting and Manipulating K-Mer Statistics.” *OUP Academic*, Oxford University Press, 4 May 2017, academic.oup.com/bioinformatics/article/33/17/2759/3796399.

31. Koledenkova, Kseniia, et al. “Plasmopara Viticola the Causal Agent of Downy Mildew of Grapevine: From Its Taxonomy to Disease Management.” *Frontiers*, Frontiers, 19 Apr. 2022, www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.889472/full.
32. Krueger, Felix. “Trim Galore.” *Babraham Bioinformatics*, Babraham Institute, 19 Nov. 2019, www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
33. Lan, Xia, et al. “Plasmopara Viticola Effector PVRXLR131 Suppresses Plant Immunity by Targeting Plant Receptor-like Kinase Inhibitor BKI1.” *Molecular Plant Pathology*, U.S. National Library of Medicine, June 2019, www.ncbi.nlm.nih.gov/pmc/articles/PMC6637860/.
34. Li, Heng, and Richard Durbin. “Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform.” *OUP Academic*, Oxford University Press, 18 May 2009, academic.oup.com/bioinformatics/article/25/14/1754/225615.
35. Marçais, Guillaume, and Carl Kingsford. *JELLYFISH - Fast, Parallel k-Mer Counting for DNA*, 26 July 2011, www.cs.cmu.edu/~ckingsf/software/jellyfish/.
36. Nielsen, Rasmus, et al. “Genomic Scans for Selective Sweeps Using SNP Data.” *Genome Research*, Cold Spring Harbor Lab, 5 June 2005, genome.cshlp.org/content/15/11/1566.
37. Nityagovsky, Nikolay N., et al. “Distribution of Plasmopara Viticola Causing Downy Mildew in Russian Far East Grapevines.” *MDPI*, Multidisciplinary Digital Publishing Institute, 27 Mar. 2024, www.mdpi.com/2311-7524/10/4/326.
38. Paineau, Manon, et al. “An Effector Deletion Leads to the Breakdown of Partial Grapevine Resistance to Downy Mildew.” *bioRxiv*, Cold Spring Harbor Laboratory, 1 Jan. 2023, www.biorxiv.org/content/10.1101/2023.08.17.553663v1.full#F1.

39. Pavlidis, Pavlos, et al. “Sweed: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes.” *OUP Academic*, Oxford University Press, 18 June 2013, academic.oup.com/mbe/article/30/9/2224/999783.

40. Possamai, Tyrone, and Sabine Wiedemann-Merdinoglu. “Phenotyping for QTL Identification: A Case Study of Resistance to Plasmopara Viticola and Erysiphe Necator in Grapevine.” *Frontiers*, Frontiers, 27 June 2022, www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.930954/full.

41. Presti, Libera Lo, et al. “Fungal Effectors and Plant Susceptibility.” *Annual Review of Plant Biology*, Annual Reviews, 29 Apr. 2015, www.annualreviews.org/content/journals/10.1146/annurev-arplant-043014-114623.

42. Rocafort, Mercedes, et al. “Apoplastic Effector Proteins of Plant-Associated Fungi and Oomycetes.” *Current Opinion in Plant Biology*, Elsevier Current Trends, 2 Apr. 2020, www.sciencedirect.com/science/article/abs/pii/S1369526620300224#:~:text=Plant%2Dassociated%20fungi%20and%20oomycetes%20deliver%20effector%20proteins%20into%20the,plant%20cell%2Dsurface%20immune%20receptors.

43. Rouxel, Mélanie, et al. “Phylogenetic and Experimental Evidence for Host-Specialized ...” *Wiley*, Wiley, 22 Sept. 2012, www.researchgate.net/publication/233422452_Phylogenetic_and_experimental_evidence_for_host-specialized_cryptic_species_in_a_biotrophic_oomycete.

44. Sargolzaei, Maryam, et al. “Rpv29, Rpv30 and Rpv31: Three Novel Genomic Loci Associated with Resistance to Plasmopara Viticola in Vitis Vinifera.” *Frontiers in Plant Science*, U.S. National Library of Medicine, 8 Oct. 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7583455/.

45. Simão, Felipe A, et al. “Busco: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.” *OUP Academic*, Oxford University Press, 9 June 2015, academic.oup.com/bioinformatics/article/31/19/3210/211866.

46. Sperschneider, Jana, and Peter N. Dodds. “Effectorp 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes.” *Molecular Plant-Microbe Interactions®*, vol. 35, no. 2, Feb. 2022, pp. 146–156, https://doi.org/10.1094/mpmi-08-21-0201-r.

47. Stanke, Mario, et al. “Augustus: Ab Initio Prediction of Alternative Transcripts.” *OUP Academic*, Oxford University Press, 1 July 2006, academic.oup.com/nar/article/34/suppl_2/W435/2505582.

48. Tarailo-Graovac, Maja, and Nansheng Chen. “Using Repeatmasker to Identify Repetitive Elements in Genomic Sequences - Tarailo-graovac - 2009 - Current Protocols in Bioinformatics - Wiley Online Library.” *Wiley*, Current Protocols, 1 Mar. 2009, currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471250953.bi0410s25.

49. Taylor, Andrew, and David C Cook. “An Economic Assessment of the Impact on the Western Australian Viticulture Industry from the Incursion of Grapevine Downy Mildew | Request PDF.” *ResearchGate*, Journal of Plant Diseases and Protection, Feb. 2018, www.researchgate.net/publication/323227527_An_economic_assessment_of_the_impact_on_the_Western_Australian_viticulture_industry_from_the_incursion_of_grapevine_downy_mildew.

50. Taylor, Andrew S., et al. “Population genetic structure and cryptic species of *plasmopara viticola* in Australia.” *Phytopathology®*, vol. 109, no. 11, 17 Sept. 2019, pp. 1975–1983, https://doi.org/10.1094/phyto-04-19-0146-r.

51. Tran Thi Ngoc, Anh, et al. “Dnaj, a Heat Shock Protein 40 Family Member, Is Essential for the Survival and Virulence of Plant Pathogenic Pseudomonas Cichorii JBC1.” *Research in Microbiology*, Elsevier Masson, 24 June 2023, www.sciencedirect.com/science/article/abs/pii/S0923250823000694#:~:text=Bacterial%20plant%20pathogens%20must%20cope,in%20response%20to%20environmental%20stress.

52. Turrà, David, and Antonio Di Pietro. “Fungal Germ Tube.” *Fungal Germ Tube - an Overview | ScienceDirect Topics*, 2015, www.sciencedirect.com/topics/agricultural-and-biological-sciences/fungal-germ-tube.

53. Vercesi, A., Tornaghi, R., Sant, S., Burruano, S., and Faoro, F. (1999). A cytological and ultrastructural study on the maturation and germination of oospores of *Plasmopara viticola* from overwintering vine leaves. *Mycol. Res.* 103, 193–202. doi: 10.1017/S095375629800700X

54. Vurture, Gregory W, et al. “GenomeScope: Fast Reference-Free Genome Profiling from Short Reads.” *Bioinformatics (Oxford, England)*, U.S. National Library of Medicine, 15 July 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5870704/.

55. Wang, Shumei, et al. “Genius Architect or Clever Thief—How Plasmodiophora Brassicae Reprograms Host Development to Establish a Pathogen-Oriented Physiological Sink.” *APS*, APS, 22 Dec. 2023, apsjournals.apsnet.org/doi/10.1094/MPMI-03-19-0069-CR.

56. Wei, Zuoying, et al. “Phylogeny and Taxonomy on Cryptic Species of Forked Ferns of Asia.” *Frontiers*, Frontiers, 8 Nov. 2021, www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2021.748562/full.

57. Wickham, Hadley. “Ggplot2: Elegant Graphics for Data Analysis.” *SpringerLink*, Springer New York, 3 Oct. 2009, link.springer.com/book/10.1007/978-0-387-98141-3.

58. Wingerter, Chantal, et al. “Grapevine Rpv3-, RPV10- and RPV12-mediated defense responses against Plasmopara viticola and the impact of their deployment on fungicide use in viticulture.” *BMC Plant Biology*, vol. 21, no. 1, 14 Oct. 2021, https://doi.org/10.1186/s12870-021-03228-7.