Theses

5-7-2021

SecretePEPPr: Computational Prediction and Characterization of Effector-like Proteins Secreted from Vitis vinifera

Patrick Rynkiewicz

pxr8793@rit.edu

Recommended Citation

RIT

SecretePEPPr: Computational Prediction and Characterization of Effector-like Proteins Secreted from *Vitis vinifera*

By Patrick Rynkiewicz

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

May 7th, 2021

`

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 Patrick Rynkiewicz, a candidate for the Master of Science degree in Bioinformatics, has submitted his 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.

_____   _____
André O. Hudson, Ph.D.

_____   _____
Lance Cadle-Davidson, Ph.D.

_____   _____

_____

Feng Cui, Ph.D.   475-4115 (voice)
Director of Bioinformatics MS Program   fxcsbi@rit.edu

Acknowledgements

I would like to extend my greatest acknowledgement and appreciation for my committee members, Dr. Michael V. Osier, Dr. Lance Cadle-Davidson, Dr. André O. Hudson, and Dr. Eli J. Borrego, for their countless hours of patience, encouragement, thought-provoking questions, willingness to share subject matter expertise, and continued faith in this project. I owe them my gratitude for the many opportunities I have had throughout this project to gain new skills in bioinformatics and biology.

I would also like to thank my close friends and family for their never-ending support: from proof-reading my initial proposal document to attending my thesis defense. They, like my committee, served as a constant reminder of the collaborative nature of scientific work and the immense value of clear scientific communication.

Additionally, I would like to extend a special thank you to Anna Underhill at the USDA ARS for her efforts in preparing haustorial samples during the validation portion of this project, Raquel Becker for designing the logo for the SecretePEPPr computational tool, and Jeremy Jackson for his assistance in the laboratory.

`

TABLE OF CONTENTS

`

LIST OF FIGURES

`

LIST OF TABLES

`

ABSTRACT

Plant pathology research has long placed a focus on pathogen-derived effector proteins: small, secreted proteins translocated into host cells where they subvert host basal immunity and promote infection. Recent studies suggest that some plant species secrete similar effector-like proteins during mutualistic plant-fungal interactions and affect fungal growth. In this project, a computational tool named SecretePEPPr (Secreted Plant Effector-like Protein Predictor) was written, evaluated, and tested for the purpose of predicting candidate plant effector-like proteins from a set of whole genome annotations. Among other factors, this prediction tool considered classical and non-classical secretion, protein size, and the prevalence and presence of clathrin-mediated endocytic motifs. Analysis on testing data revealed a subcellular localization prediction specificity of 90% on a set of over 500 intracellular plant proteins and sensitivity of 55% on a set of experimentally validated secreted plant proteins. Across four analyzed grape proteomes, several germin-like proteins were identified as potentially haustorially localized through clathrin-mediated endocytic means. Protein length distributions revealed that effector-like candidates containing clathrin-mediated endocytic motifs were mostly in the 150-300 amino acid length range. Follow up *in vivo* validation was conducted in *Erysiphe necator*-infected Chardonnay grape leaves. Through this process, the first *Erysiphe necator* haustorial extraction method was devised using a Percoll Density Gradient followed by fluorescent labeling and fluorescence-activated cell sorting (FACS), resulting in 14 and 18 million purified fungal haustorial cells from 20 grams of heavily infected leaves. Computational streamlining of plant effector-like protein prediction established in this project provides a foundation for the top-down discovery and characterization of this recently discovered class of plant proteins, which have important implications in plant-microbe interactions and may act as targets for breeding and gene editing in plants.

`

INTRODUCTION

I. Grape Powdery Mildew Impact and Management

The ascomycete fungus *Erysiphe necator* (syn. *Uncinula necator)* is an economically impactful obligate biotroph which infects the green tissues of cultivated grapevines and causes grape powdery mildew disease (GPM), significantly reducing grape quality and yield. As of 2017, 7.5 million hectares are occupied globally by vineyards, and the wine market alone is valued at 30 billion USD worldwide (OIV, 2018). It is expected that GPM-resistant varieties of grapevine would yield economic benefits as high as $48 million USD per year across wine, table, and raisin grape industry segments in California (Fuller et al., 2014). While there exist several grape species (*Vitis sp.*) with varying levels of resistance to GPM, most commercially sold grapes are produced by the GPM-susceptible species, *Vitis vinifera*, which originated in Europe and produces grapes with high yields and desirable flavor**.** Evidence supports the origin of grape powdery mildew in eastern North America, where *E. necator* evolved independently of *V. vinifera* until the pathogen was introduced to Europe (Brewer & Milgroom, 2010). The native host of *E. necator* at the species-level is not currently known, although two distinct haplotypes have been observed between *V. rotundifolia* and all other *Vitis* species (Brewer & Milgroom, 2010). With the widespread use of *V. vinifera* to produce commercially desirable grapes, work such as the *Vitis*Gen project (http://www.vitisgen.org/) address the global need for creating and improving pathogen resistant cultivars.

Various strategies have been developed for the management of GPM, including the widespread use of organic and inorganic fungicides (Gadoury et al., 2012). Commonly used fungicides consist of elemental sulfur, benzimidazoles, ergosterol biosynthesis inhibitors such as sterol demethylation-inhibitors (DMIs), quinone-outside inhibitors such as strobilurins, and succinate dehydrogenase inhibitors (SDHIs). Like antibiotics, antifungals used for GPM management generally target the formation or aggregation of membrane components in fungi. Many contemporary fungicides applied to crop diseases like GPM have single-site targets, and resistance can be established with as little as a single base pair change or a small number of copy number variations (Délye et al., 1997). Mechanisms of acquired

resistance to single-site fungicides in plant pathogens commonly includes the use of efflux pumps, and less commonly overexpression of the target gene and degradation of the fungicide (Lucas et al., 2015). Various studies have explored DMI resistance and shown its polygenic characteristics, such as mutation of target enzyme 14α-demethylase CYP51 and overexpression of resistance genes or decreased uptake of DMIs as a result of overactive efflux proteins (Délye et al., 1997). In 2015, GPM management alone was found to account for 89 percent of pesticide treatments by grape growers, with the total expenditure on pesticide products in California alone reaching $78 million (Sambucci et al., 2019). Ideally, *V. vinifera* resistance to GPM would be implemented at the plant genome level to circumvent the present drawbacks of mildew management with fungicides. Contemporary genome editing technologies such as CRISPR-Cas9 have been put forward as powerful methods for modifying susceptibility genes and increasing plant disease resistance in tandem with classical resistance genes (Langner et al., 2018). Traditional breeding approaches optimizing for heightened resistance rely on hybridization of target cultivars with resistant cultivars to incorporate resistance genes (Yin & Qiu, 2019). The identification of specific targets for breeding or genome editing such as R genes or chitinases could greatly enhance latent plant defense capabilities and significantly reduce the annual economic burden of fungicide purchase application (Kumar et al., 2018).

II. Mechanics of *Vitis vinifera* Defense from GPM

GPM infection is a multi-stage process. Upon germination, *E. necator* forms an appressorium with a penetration hypha capable of breaching the plant cell wall, invading the host epidermal cell and allowing for the exchange of metabolites and nutrients directly between host and pathogen cytoplasm, as well as the secretion of fungal effector proteins capable of subverting generalized plant immunity. A variety of factors have been shown to decrease the incidence or severity of *E. necator* infection, including the age of the host (Ficke et al., 2002) and previous abiotic stress, such as low temperatures (Weldon et al., 2019). The host's first line of defense is known as PTI, or Pathogen-Associated Molecular Pattern (PAMP) Triggered Immunity. This innate and generalized immune response relies on the detection of

`

molecular structures inherent to pathogens by surface-localized pattern recognition receptors (PRRs). Chitin, an integral component in the cell wall of fungi, acts as a PAMP triggering the plant's first line of defense upon interaction with epidermal cells. Upon detecting a PAMP, the host initiates an immune response composed of stress signaling, cell wall strengthening, and the production of antimicrobial molecules including reactive oxygen species (ROS).

Consequently, pathogens must evolve mechanisms to overcome or prevent PTI largely by employing small, highly specific peptides called effector proteins. In the case of *E. necator*, the pathogen uses a specialized invaginating structure, called a haustorium, to secrete effectors and inhibit host defenses. Unlike other fungal invaginating structures such as penetration hyphae, haustoria are four-layered invaginating structures which incorporate membrane and protein elements from the plant in the outermost layer. Subsequently the plant deploys a second, more target-specific, layer of host immunity known as effector-triggered immunity, or ETI, in which resistance (*R*) genes detect specific effector proteins, named avirulence (*Avr*) genes. Upon detection, the host cell initiates localized programmed cell death to "starve" the obligate biotrophic pathogen of the host's nutrients and limit the extent of infection. Several genetic loci have been attributed to effector-triggered immunity, with one found in *V. vinifera,* known as *REN1* (Coleman et al., 2009), and a dozen others in more resistant *Vitis* species such as *REN2* and *REN4*, delineated mainly by the strength, speed, and race-specificity of resistance (Cadle-Davidson 2019). The *REN1* gene restricts hyphal growth of specific *E. necator* strains, or races, and the *REN1* locus spans several nucleotide-binding site leucine-rich-repeat regions (NBS-LRRs) which are expected to be central to cytoplasmic detection of pathogen effectors and subsequent immune response (Hoffmann et al., 2008; McHale et al., 2006).

III. Role of Secreted Proteins in the GPM-*Vitis* Pathosystem

The apoplastic space between the host plasma membrane and host cell wall is a hotbed for early communication between plant and pathogen (Figure 1). Among other signaling molecules, proteins play a major role in host-pathogen interaction at this interface. Previous studies in rice have identified the

`

involvement of glycosyl hydrolase family proteins, proteases, esterases, peptidases, and chitinases in apoplastic interactions with the *M. oryzae* pathogen (Kim et al., 2013). Evidence of similar interactions has been found in *Arabidopsis* and *Brassica* species (Floerl et al., 2012). Host secreted proteins appear to be prevalent in these interactions. Other proteins implicated in the initial apoplastic interaction include peroxidases, thionins, xylanase inhibitors, other pathogenesis related (PR) proteins, and DUF26s (Belchí-Navarro et al., 2013; Gupta et al., 2015; Sels et al., 2008; Van loon & Van strien, 1999). In *Vitis*, endo ß-1,3-glucanase, four peroxidases, and an endo 1,4-ß-mannosidase have been experimentally isolated from the extracellular proteome (Belchí-Navarro et al., 2013).
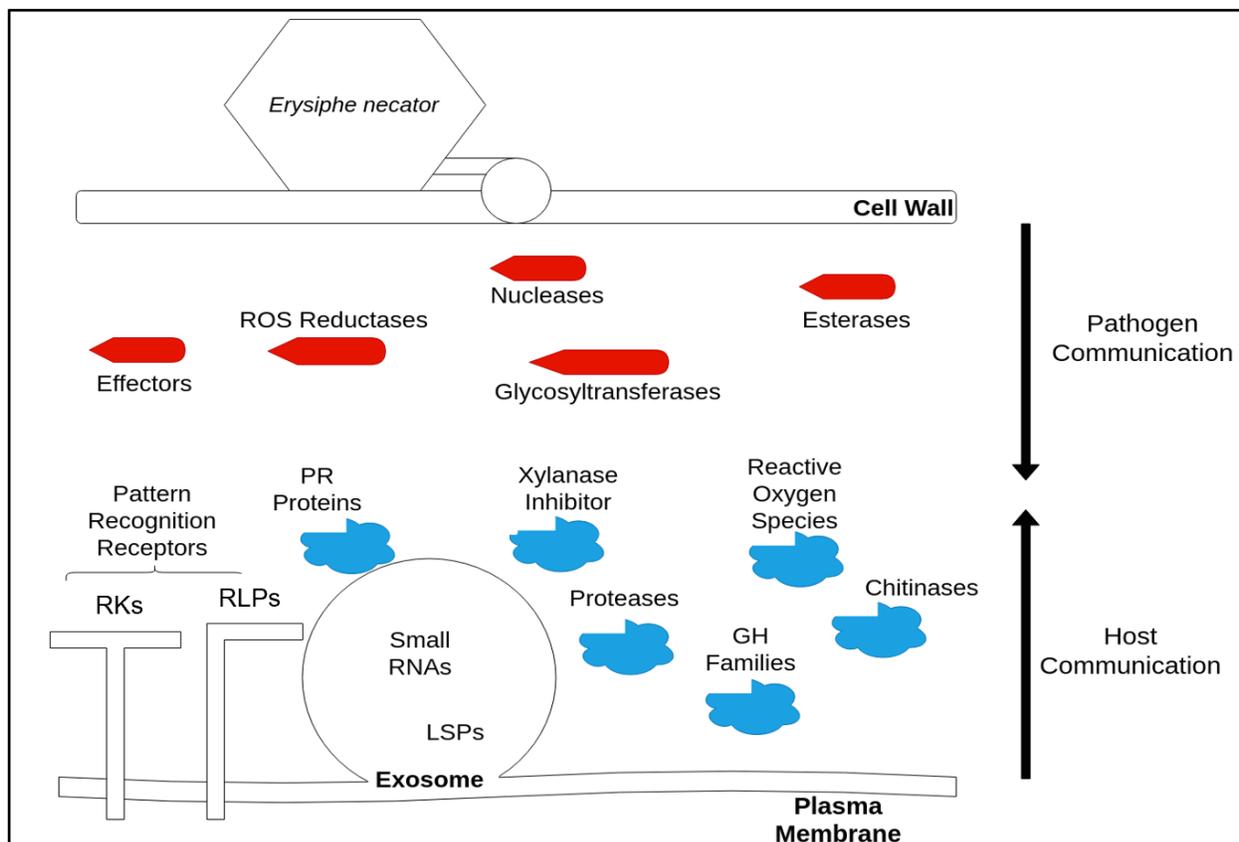
Figure 1. Secreted protein communication between *E. necator* and *V. vinifera* in the apoplastic space.
ROS=Reactive Oxygen Species, GH=Glycosyl Hydrolase, RK=Receptor Kinase, RLP= Receptor-like protein, PR=Pathogenesis Related protein. Not shown: Haustorium at the interface between the plant plasma membrane and fungal cytoplasm.

`

Secreted proteins have primarily been identified and studied in the context of the classical Endoplasmic Reticulum-Golgi pathway. N-terminal signal peptide sequences that are 15-30 amino acid long have been characterized through *in vitro* methods and are used as evidence of classical secretion, indicative of post-translational modification in the Golgi apparatus (Alexandersson et al., 2013). In the context of pathogen-host interactions, some effectors appear to stymie the secretion of classically secreted proteins through specific protein-protein interactions in the host cytoplasm. In the context of plant defense, certain secreted proteins have been found to increase in quantity during the first two hours of infection (Cheng et al., 2009).

A pathogenesis-relevant discovery in plant secretome studies is the presence of non-classically secreted proteins circumventing the ER-Golgi pathway and possessing no signal peptide or post-translational modification. Leaderless secreted proteins (LSPs) constitute, on average, 50% of the total plant secretome (Agrawal et al., 2010). Additionally, plants undergoing biotic or abiotic stresses tend to contain more LSPs than non-stressed plants, indicating a correlation between environmental signals and the secretion of LSPs into the apoplast. LSPs such as glycolytic enzymes and methionine biosynthesis enzymes normally residing in the intracellular space have been identified in the apoplast. Additionally, LSPs have been demonstrated to secrete quickly during the first 2-3 hours of stress, indicating that non-classical secretion may play a large role in early plant defense (Cheng et al., 2009). LSPs, normally being intracellular proteins, appear to have alternative functions once they are secreted into the extracellular space. In the case of thioredoxin, a protein normally responsible for redox balance in plants becomes an extracellular signaling molecule and participates in the inflammatory process as a cytokine (Nickel & Seedorf, 2008). Non-classically secreted proteins exit the cytoplasm via exosome/multivesicular body-dependent secretion, exovesicle or plasma membrane blebbing, lysosome-dependent secretion, and plasma membrane-specific transporter-dependent secretion (Agrawal et al., 2010). The relevance of secreted proteins of pathogen and host origin, as well as the consideration for both classically and non-classically secreted proteins in studying host-pathogen interactions through the lens of proteomics is evident and remains poorly understood.

`

While various investigations into the interactions between *V. vinifera* and *E. necator* have focused on fungal effectors and their impact on host cellular functions after crossing the plasma membrane, effector-like proteins of host origin have been discovered in mutualistic interactions between *Populus trichocarpa* and the fungus *Laccaria bicolor* (Plett et al., 2017), as well as in the pathogenic interaction between *Triticum aestivum* and a fungal pathogen, *Zymoseptoria tritici* (Zhou et al., 2020). These effector-like proteins (ELPs) have been shown to directly influence fungal growth and morphology, demonstrating their role in fungal colonization. These novel insights provide an opportunity for understanding the complex interactions between *Vitis* and *E. necator* outside of the cytoplasm from the perspective of a new class of secreted proteins. Little work has been done on characterizing these effector-like plant proteins. Furthermore, effector-like plant proteins have not yet been explored in the context of a pathogenic relationship with a haustoria-forming fungus, such as *E. necator*.

IV. The Haustorium - A Primary Interface for Plant-Pathogen Communication

The intracellular structure responsible for the exchange of effectors and nutrients is the haustorium, which acts as a channel for pathogen-host interaction during infection. While haustorial biology has not been thoroughly studied in *E. necator,* much of its function can be inferred from better characterized haustorial fungi (powdery mildews, rusts) and oomycetes. Through the haustorium, the pathogen secretes effector proteins capable of suppressing the host's primary immune defenses. In return, the pathogen retrieves nutrients such as hexoses, amino acids, and vitamins (Qiu et al., 2015). The haustorium is a multi-membranous, globose cellular structure connected to a haustorial mother cell by a neckband, sharing cytoplasmic fluid but possessing its own nucleus (Szabo & Bushnell, 2001). Between the haustorial fungal cell and the surrounding plant cytoplasm are the fungus-derived haustorial plasma membrane and haustorial wall, followed by a carbohydrate-enriched extrahaustorial matrix, and finally surrounded by the extrahaustorial membrane which is a derivative of the host plasma membrane (Szabo & Bushnell, 2001). There is evidence to support fungal hijacking of plant endocytic trafficking towards the extrahaustorial membrane during pathogen colonization, as well as a depleted host-derived protein

`

repertoire in the extrahaustorial membrane in comparison to the plant plasma membrane (Bozkurt et al., 2015) Among the remaining protein repertoire is RabG3c, a late endosome marker protein indicative of endosomal traffic rerouting to the fungal interface (Bozkurt et al., 2015). Furthermore, proteins from several host origins including the plasma membrane, vacuolar membrane, endocytic vesicles, plasmodesmata, and the endoplasmic reticulum have been found within the extrahaustorial membrane, suggesting that several stress-related transport pathways may be redirected in the formation of haustoria (Bozkurt & Kamoun, 2020).

The exact mechanism or mechanisms of effector translocation from within fungal cells across these membranes and into the plant cytoplasm has been a topic of inquiry for some time (Lo Presti & Kahmann, 2017; Mendgen et al., 2011; Rafiqi et al., 2012). Three potential pathways of fungal effector translation include the use of receptor-mediated endocytosis, fusion of extracellular vesicles with the extrahaustorial membrane, and/or active transport utilizing a translocon (Bozkurt & Kamoun, 2020; Lo Presti & Kahmann, 2017). However, these pathways are derived from observed similarities between invaginating fungal pathogens and *Plasmodium* species, and direct evidence of effector translocation by any of these means has yet to be demonstrated *in vivo* with the exception of exosome vesicles, which were observed within the haustoria and extrahaustorial matrix of the powdery mildew fungus *Golovinomyces orontii* (Bozkurt & Kamoun, 2020; Micali et al., 2011).

V. Fungal Effector Proteins

Effector proteins in fungi are small (<315 amino acid in length), often signal sequence-containing peptides capable of translocating from within the hyphae or haustorium through the interface between pathogen and host, and ultimately localizing within the apoplast, cytoplasm, chloroplast, or other plant organelles. Fungal effector proteins can be classified as virulence factors if they induce metabolic changes favorable to the pathogen, or avirulence factors if they are recognized as foreign and induce ETI (Chaudhari et al., 2014). The xylanase BcXyl1 in *Botrytis cinerea,* a necrotrophic fungal pathogen that affects grapevine and several other plant species, was demonstrated to induce plant cell death even in the

`

absence of key residues responsible for its xylanase (e.g. cell wall deteriorating) activity, suggesting that some effectors may have dual functionality and are not restricted to one active site (Yang et al., 2018). While mechanisms of effector secretion from haustoria have yet to be pinpointed, it has been suggested that some effectors may localize within the plant cytoplasm through pathogen-independent mechanisms (Rafiqi et al., 2010).

One of the most well understood functions of pathogenic effectors involves the subversion of generalized plant immunity, or PTI, by masking common PAMPs or neutralizing host proteolytic activity responsible for their recognition. In the fungal pathogen *Cladosporium fulvum,* the effector Avr2 inhibits papain-like proteases Rcr3 and PiP1, which belong to a class of cysteine proteases known to induce salicylic acid-related immunity in plants (H. Liu et al., 2018). Effector proteins can also redirect subcellular localization of target host proteins, as is the case of oomycete effector AVRblb2 which specifically targets and inhibits the secretion of papain-like protease C14 in *N. benthamiana*, preventing it from reaching the apoplast (T. O. Bozkurt et al., 2011). *E. necator* has previously been screened for candidate effector proteins and contains EKA-like proteins and ribonuclease-like proteins detected in early infection time points suggesting a potential early involvement in infection (Jones et al., 2014). EKA-like proteins are homologous to those in the *Avr k 1* and *Avr a 10* gene family, which have been shown to elicit a hypersensitive response expected in ETI-triggered gene for gene interactions (Amselem et al., 2015).

Fungal effectors are always at risk of being detected by new resistance proteins in plants and eliciting a hypersensitive response, and as a result are subject to rapid evolution (Lo Presti et al., 2015). This dynamic is evidenced by a generally greater prevalence of positive selection on genes coding for secreted proteins rather than those with transmembrane domains, or cytoplasmic proteins. For easier identification, motif searching of candidate secreted effector proteins present in fungi, oomycetes, and bacteria has been employed to find conserved domains that can be used as fingerprints in studying effector evolutionary history and markers for determining potential secreted effector proteins (Stergiopoulos & de Wit, 2009). This approach has been particularly successful in oomycetes, in which

`

candidate effector proteins are often identified and characterized in the class of RXLR motifs possessing several virulence functions (Morgan & Kamoun, 2007). Albeit the fact that these motifs are conserved among effector proteins in their respective species, they generally are not found to interact directly with plant proteins. Fungal effectors are not as well characterized by common motifs, but some powdery mildew effector proteins have been identified as having an N-terminal Y/F/WxC motif (Godfrey et al., 2010). Overall, this highlights the inherent complexity of classifying fungal effector proteins and studying their evolutionary history.

While the effector-like protein repertoire of crop plants such as *Vitis* remains to be determined, certain parasitic species of plants are known to produce haustoria and modulate host metabolism within their life cycle. Broomrapes of the family *Orobanchaceae* form haustoria within host roots to steal water and nutrients, likely employing effector proteins in the process and triggering host immunity in a similar way as fungi (Saucet & Shirasu, 2016). Like fungal and bacterial pathogens, root parasitic weeds can trigger a hypersensitive response commonly associated with ETI. In the case of *Striga gesnerioides*, the 195 amino acid long protein SHR4z has been demonstrated to bind and inhibit ubiquitin E3 ligase POB1 which is responsible for promoting a hypersensitive response during infection (Su et al., 2020). This supports the notion that plants are capable of producing and secreting effector-like proteins when interacting with other species. To date, only one study has described the phenomenon of effector-like plant peptide secretion and localization within the cytoplasm of ascomycete fungi (Plett et al., 2017). These proteins were less than 250 amino acids in length, differentially expressed during fungal interaction, and possessed an N-terminal secretion signal. It is important to note that these effector-like proteins were studied in a symbiotic fungus. However, fungal effectors are known to be used by both symbiotic and pathogenic fungi to bypass the host defense system and appear to localize frequently to the host apoplast and host cytoplasm (Zeilinger et al., 2016).

`

VI. Bioinformatics Methods for Pathogenesis-Relevant Protein Functions

Protein localization prediction, particularly secretion prediction, has been successfully employed in the identification of fungal effector and plant effector-like proteins, and can be considered a major initial screening method for their discovery. Bioinformatics tools have made great strides in signal peptide and subcellular localization prediction. Sophisticated HMM and neural network-based tools have been released displaying a high confidence (>90% accuracy) in classically secreted protein prediction as well as progress towards accurate distinction between truly secreted proteins and those with transmembrane domains (Petersen et al., 2011; Savojardo et al., 2018; Sperschneider et al., 2017; Tsirigos et al., 2015). In bioinformatics prediction of signal peptides, the predicted secretome is defined by the collection of proteins possessing an N-terminal signal peptide but no transmembrane domains, GPI-anchor, or subcellular localization signal (Agrawal et al., 2010).

To discover *Vitis* peptides that could be considered effector-like, a pipeline workflow (Secreted Plant Effector-like Protein Predictor, or SecretePEPPr) was developed using a sequence-based genome annotation mining approach. Consistent with a previous study in *Populus trichocarpa* (Plett et al., 2017), subcellular localization prediction was emphasized and used as a principle upon which to exclude intracellularly localized proteins. Additionally, a size cutoff of 315 amino acids was impose. This was done in accordance with an analysis of a wide range of fungal effectors conducted by Sperschneider et al. (Sperschneider et al., 2016, 2018), which determined that experimentally verified fungal effectors from a wide range of species were no longer than 315 amino acids in length. Presence of an N-terminal secretion peptide, size exclusion, and the presence of transmembrane helices was also predicted in other plant secreted protein studies (Plett et al., 2017; Zhou et al., 2020).

In this project, a comprehensive computational method for identifying candidate secreted effector-like proteins of *Vitis* origin was developed and evaluated, and the resulting pool of candidate proteins from several *Vitis vinifera* cultivars was functionally characterized. Additionally, actionable progress towards validating candidate proteins is accomplished by developing a haustorial extraction

`

methodology specific to the *Vitis vinifera* / *Erysiphe necator* pathosystem, paving the way towards whole-scale proteomic analysis of *Erysiphe necator* haustoria.

`

MATERIALS AND METHODS

I. Protein Feature Prediction Tools and Pipeline Workflow

Effector-like proteins were predicted based on their subcellular localization, size, and the presence of an endocytic motif. A literature search was conducted to determine the most appropriate subcellular localization prediction tools to use in the SecretePEPPr (Secreted Plant Effector-like Protein Predictor) pipeline, based largely on the Sperschneider et al. 2017 evaluation of subcellular localization prediction tools (Sperschneider et al., 2017). Subcellular localization tools included in the computational pipeline, named SecretePEPPr (Secreted Plant Effector-like Protein Predictor) performed highly in this evaluation and had a command line interface, as a web interface would have a significantly lower throughput and pose an unfeasible challenge to automate reliably. Intracellular protein localization prediction was added with the inclusion of LOCALIZER, predicting chloroplast, nuclear, and mitochondrial proteins, and NLStradamus, predicting nuclear-localized proteins. SignalP5 and TargetP were used to predict classically secreted, N-terminal signal peptide-containing proteins which are secreted via the ER-Golgi pathway. The inclusion of ApoplastP allowed for the prediction of non-classically secreted proteins lacking an N-terminal signal peptide. TargetP was also added due to its ability to predict classically predicted proteins, chloroplast localized proteins, and mitochondrial proteins. KohGPI was employed to predict GPI anchors, and TMHMM was used for transmembrane domain prediction. The resulting SecretePEPPr pipeline architecture is shown below in Figure 2.

`
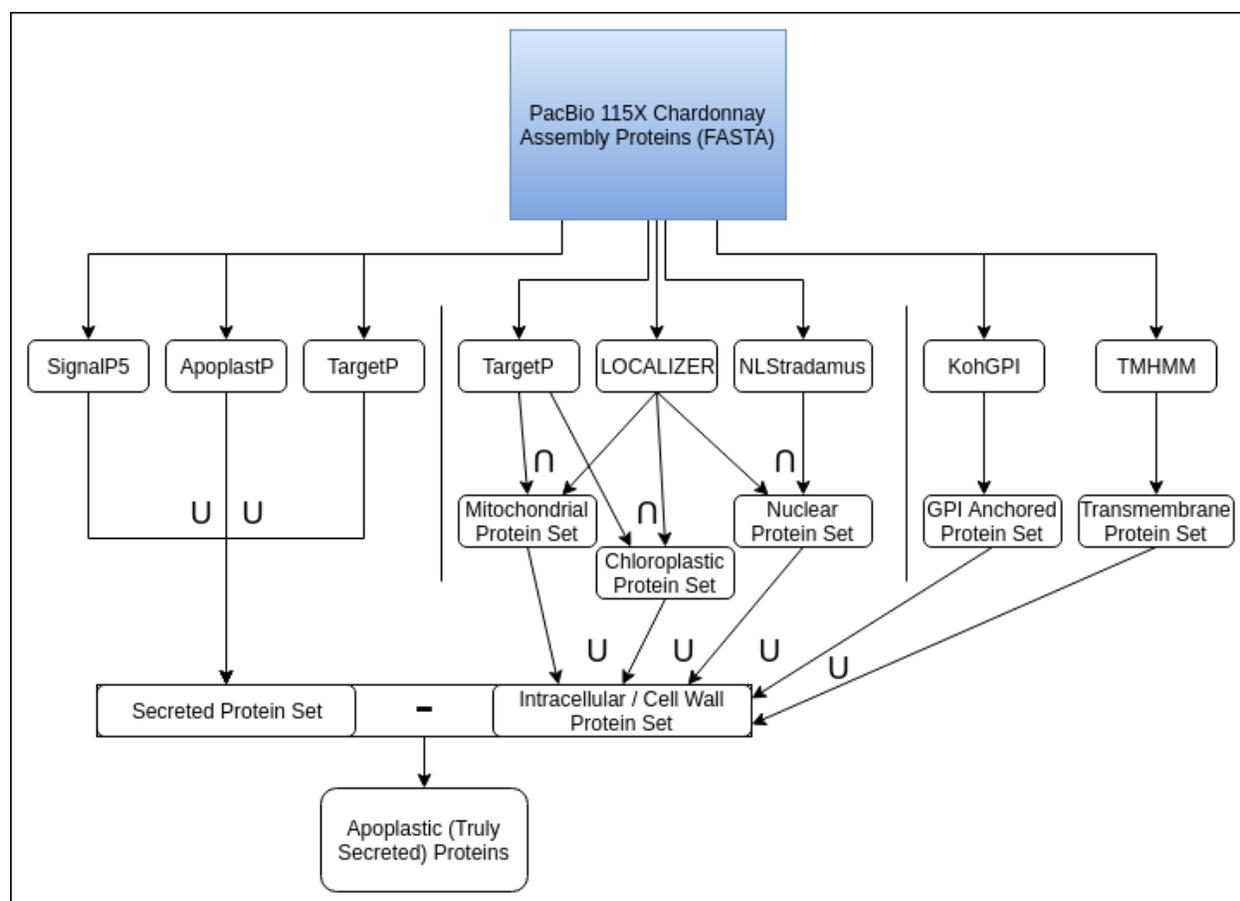
Figure 2. Secreted Protein Prediction Pipeline (SecretePEPPr) Workflow. ∪ represents portions of the pipeline where protein lists were combined from predictors in an additive fashion and deduplicated. ∩ represents portions of the pipeline where the intersect in protein lists between predictors was used. Intracellular / Cell Wall localizing proteins were removed from the Secreted Protein Set as an additional screening measure.

II. Pipeline Architecture -

The SecretePEPPr pipeline was written in Python 3 implementing the configparser, subprocess, os, argparse, Bio::SeqIO, and re (regular expressions) additional libraries. The pipeline consists of the scripts parse_secretepeppr.py, run_TargetP.sh, SecretePEPPr_motif_search.py and secretepeppr.py, the latter of which is the only script needed to be run by the user. SecretePEPPr was intended to be run on the Bash command line due to the use of the python os and subprocess modules for controlling workflow. Configuration options including paths to each predictor, input parameters such as maximum number of threads and FASTA file path, output parameters such as the option to save temporary files, and cutoffs for

`

each respective predictor are all configurable in the CONFIG.ini file with delimited sections and data input fields. Each SecretePEPPr run should have its own directory with only the custom CONFIG.ini file within that directory, for exact reproducibility of each run. Upon calling the secretepeppr.py script within the chosen directory, the pipeline will create a run/ and tmp/ directory for managing intermediate files alongside final output files. Each of the seven predictors is run by this script, which results in seven final output files located in the run/predictor_outputs/ directory in the format '*file.predictorname*'. The tmp/ directory contains split FASTA files for input to KohGPI and TargetP, as well any other temporary files associated with predictors. Once all seven final predictor outputs are present in the run/predictor_outputs directory, the parse_secretepeppr.py script parses each individual predictor output file and returns the final list of predicted proteins to the user. Additionally, the program outputs a short summary of the run and a FASTA file of the predicted proteins, located in the run/directory. It is also possible to run the *parse_secretepeppr.py* script on previous runs of any of the seven included predictors, so long as the output files follow the '*file.predictorname*' format, and there are no duplicates within a directory. Finally, endocytic motifs of the user's choice as indicated in the CONFIG.ini file are identified with the *SecretePEPPr_motif_search.py* script and output to the user in the run/subdirectory.

III. Endocytic Motif Identification

As an additional optional analysis step in the SecretePEPPr pipeline, the option was added to further filter proteins based on the top five percent of most prevalent endocytic motifs. This was included to provide two protein sets for downstream validation: one larger and more broadly defined plant secretome for discovery-based identification methods, and another more narrowly defined motif-containing secreted protein set for targeted identification and validation. The set of proteins without endocytic motif prediction is intended for broad-scale proteomic studies where pairwise comparison of predicted and experimentally observed proteomes do not require as strict levels of specificity, and where the hypothesis is not centered specifically around clathrin-mediated endocytosis as the protein translocation mechanism. The much smaller set of proteins filtered with additional endocytic motif

`

screening is intended for targeted studies such as screening transgenic *Vitis vinifera* samples, and labeling candidates with a fluor to determine localization. As a further confirmatory analysis in this study, protein lists were further filtered by those present in the top five percent of most prevalent endocytic motifs across all four *Vitis vinifera* cultivars.

IV. SecretePEPPr Predictor Parameterization and Secretion Prediction Assessment

For the confirmation of pipeline subcellular localization prediction accuracy and parameterization of subcellular localization prediction tools, several sets of plant protein annotations were retrieved (Table 1).

Table 1. Categorized plant protein data used for SecretePEPPr pipeline testing and validation.

Category	Dataset	Abbreviation	Proteins	Reference
Parameterization Data	Delaunois et al. 2013	APSp	60	(Delaunois et al., 2013)
Validation Data	cropPAL, intracellular	CPL-	114	(Hooper et al., 2016)
Validation Data	cropPAL, extracellular	CPL+	481	(Hooper et al., 2016)
Validation Data	Arabidopsis proteins, nucleolar	ATN-	124	(Berardini et al., 2015)

To mitigate any inaccuracy stemming from the use of subcellular localization tools trained on several taxonomic kingdoms, and to further increase sensitivity for *V. vinifera* proteins, the APSp set of *in vitro* isolated apoplastic Chardonnay proteins was the initial test set used to determine cutoffs for each predictor in the pipeline. While LOCALIZER and ApoplastP were developed and trained specifically on plant data, other predictors such as SignalP and KohGPI were trained on proteins from various taxa. The APSp proteins also served to reflect translated and secreted proteins regardless of mechanism or expression level, as these proteins were identified directly through mass spectrometry methods by (Delaunois et al., 2013). With the APSp set of apoplastic proteins in Chardonnay, individual predictors were fine-tuned to predict for the maximum sensitivity, i.e. prediction of these proteins as secreted. APSp

`

proteins originally mapped to accessions of a Chardonnay genome that was no longer supported in the NCBI database were re-mapped to the I10V1 Chardonnay genome annotations using BLASTP.

Following parameterization, a secretion prediction assessment was performed. Validation test sets were designed to encompass a range of known secreted and non-secreted proteins, mostly of rice, corn, and *Arabidopsis thaliana* origin. The CRP+ validation test set of secreted proteins was collected under the condition of being validated via GFP assay and present in all possible secretory pathways (Hooper et al., 2016). Since cropPAL only provided the Ensembl Protein IDs of entries, the Ensembl REST API was used to retrieve sequences in FASTA format**.** The CRP- validation test set of non-secreted proteins was collected under the same conditions as CRP+, and additionally nucleolar proteins from The Arabidopsis Information Resource were used to represent non-secreted proteins in a non-crop model plant (Berardini et al., 2015).

V. Effector-like Protein Prediction in Four *V. vinifera* Genomes

Genome annotations from the reference PN40024 grape genome assembly originally sequenced by Sanger sequencing, as well as three additional *V. vinifera* genome annotations derived from PacBio-long read sequenced Cabernet Sauvignon, Carmenere, and Chardonnay were used in a multi-cultivar prediction of effector-like proteins in *V. vinifera* (Table 2). BUSCO scores, representing the completeness of a genome annotation by assessing its contents for the presence of universal single-copy ortholog genes, were the metrics used to assess the quality of the REFg, CABg, CRMg, and CHRg genomes. The SecretePEPPr pipeline was used to analyze each genome annotation dataset using the default parameters and 16 threads on a Ryzen 7 2700 CPU with 32 gigabytes of RAM.

`

Table 2. Annotation methods for all four *Vitis vinifera* proteomes used in this study. Proteomes that were screened for effector-like proteins are largely annotated using gene modeling software and homology tools with the addition of EST and RNAseq data.

Genome Abbreviation	Annotation Method(s)	Number of Proteins	Year Published	Citation
PN40024 Reference (12X) REFg	- Automated feature annotation (GAZE) - cDNA evidence - Alignment-based annotation with proteins/genes from other species	29,927	2011	(Jaillon et al., 2007)
Cantu Cabernet Sauvignon CABg	- Transposable and noncoding element automated annotation - Repeat sequence annotation - Automated gene annotation (MAKER) - Transcriptomic evidence (Iso-Seq) - BLASTP against RefSeq Plant	96,331	2019	(Roach et al., 2018)
Cantu Carmenere CRMg	- Repeat sequence annotation - EST evidence - PN40024 CDS homology - Transcriptomic evidence (Tannat/Corvina, CABg Iso-Seq) - BLASTP against RefSeq Plant	73,109	2019	(Minio et al., 2019)
Chardonnay I10V1 (115X) CHRg	- Repeat sequence annotation (custom repeat library) - Automated feature annotation (MAKER) - Transcriptomic evidence (I10V1 short reads)	112,320	2018	(Minio et al., 2019)

`

V. Candidate Effector-like Protein Characterization

To identify candidate effector-like proteins at the cultivar level, the four sets of genome annotations in FASTA format consisting of Chardonnay, Carmenere, Cabernet Sauvignon, and PN40024 grape varietal proteins were run through the SecretePEPPr protein and analyzed. Three total sets of proteins were compared through COG category assignment with eggNOG-mapper: all proteins in the initial genome annotations, candidate effector-like proteins predicted by SecretePEPPr with size exclusion but without endocytic motif screening, and candidate effector-like proteins with the most prevalent endocytic motifs.

To support probable expression of candidate proteins during infection, two sets of RNA expression data were queried for homologues to the final four sets of host-derived effector-like protein candidates. The first set of expression data was initially processed in a separate thesis by Miriam Barnett at RIT, and consisted of paired-end 2x100bp reads of *V. vinifera cv. Chardonnay* infected with *E. necator* isolate LNYM sequenced on an Illumina GA HiSeq sequencer at 6 days post infection (Barnett, 2015). The second set of expression data is taken from (Weng et al., 2014) and consisted of *E. necator* (Schw.) Burr-infected *V. pseudoreticulata* transcripts predicted by the authors to code for secreted proteins. In this study, RNA samples were sequenced at 24h timepoints between 1- and 7-days post infection through the synthesis of cDNA libraries using random hexamers, and run on an Illumina HiSeq 2000. Homologues between candidate effector-like proteins in the SecretePEPPr + Motif protein group in Chardonnay and the first set of expression data were conducted via BLASTP. Proteins predicted from other cultivar data sets were excluded from this portion of the study as an exact match could not be made between the candidate ELPs and sequencing data. Homologues between candidate ELPs in the SecretePEPPr + Motif protein group and the second set of expression data were conducted via BLASTP as well, and all cultivars were considered as this expression study contained sequencing reads from *V. pseudoreticulata*, not *V. vinifera*.

`

VI. Haustorial Isolation Through A Novel Biphasic Extraction Method

To validate secretion and translocation of Chardonnay ELPs into *E. necator* haustoria, a biphasic haustorial extraction was performed following a modified protocol originally designed for the isolation of wheat stripe rust haustoria (Garnica & Rathjen, 2014). Ten days before extraction, healthy Chardonnay leaves were detached, surface sterilized in 10% bleach, rinsed, air-dried, and infected with *E. necator* isolate LNYM in agar plates, as previously described (Feechan et al., 2015). On the day of extraction, the inoculated leaves were removed from agar plates and surface sterilized with tap water, ice cold 2% bleach solution, 70% ethanol, and sterilized Nanopure water. Leaves were dried and deposited in a Waring blender, then two thirds of 250mL homogenization buffer with freshly added 2-mercaptoethanol was deposited in the blender and the mixture was homogenized for 25 seconds. Homogenized material was passed through a 100μm filter into sterile 50mL Falcon centrifuge tubes, leaving 15-20mL of space at the top, and solid particulates from the filter were returned to the blender. The remaining ~100mL of homogenization buffer was added to the Waring blender and again homogenized for 15 seconds, then distributed to the Falcon tubes through a 100μm filter along with the initial filtrate. Each tube of filtrate was passed through a 20um mesh for secondary filtration and distributed to fresh 50mL centrifuge tubes, then transferred to six chilled Oak Ridge round-bottom centrifuge tubes with 45mL capacity.

The Oak Ridge tubes were weight-balanced by pouring carefully between tubes while on a scale, then centrifuged at 1100xg for 15 minutes on a floor-centrifuge set at 4 degrees C. The supernatant was discarded, and the pellets were resuspended in 2mL of cold 1X isolation buffer gently using a 1 ml serological pipette to gently dispense buffer onto the pellet twice. The pellet and buffer mixture was then gently swirled to resuspend. The resuspended pellets were combined in a single conical centrifuge tube, and the volume of the mixture was brought to 20mL with 1X isolation buffer. At this point, each of four tubes with chilled Percoll-30 solution (total 24mL Percoll, 30mL Nanopure water, and 6mL 10x isolation buffer) had 5mL of the resuspended haustorial pellets gently pipetted to the top of the solution. These Oak Ridge tubes were then centrifuged at 25,000xg for 30 minutes at 4 degrees C.

`

To determine the gradient with the highest density of haustoria, one Oak Ridge tube was removed from the prep, and 1mL fractions of the tube were collected from the top. These fractions were mixed 1:1 with lactophenol cotton blue stain, and two microliters of this mixture was viewed under a tabletop microscope to determine the gradient with the highest abundance of haustoria. Following this optimization step, the remaining fractions corresponding to the highest abundance of haustoria were pooled and diluted 1:8 in 1x isolation buffer, then distributed to eight chilled 45 mL Oak Ridge tubes. The tubes were again centrifuged at 1100xg for 15 minutes on a floor centrifuge at 4 degrees C, then the supernatant was discarded, and the pellets were resuspended to a final volume of 4mL by pouring and gently mixing the same 4mL of 1X isolation buffer between each Oak Ridge tube. This volume was transferred to a 15mL conical centrifuge tube.

Then, 200 uL of 1 mg/mL WGA-AlexaFluor488 was added to the final haustorial fraction, and the 15mL conical centrifuge tube was covered in tin foil. The tube was mixed gently on a rotary mixer for 20 minutes at room temperature. The haustoria were pelleted on a benchtop centrifuge with a swing-out rotor at 2800 rpm for 5 minutes at room temperature. The supernatant was removed, and the pellet was washed twice with 1X isolation buffer, then finally again resuspended in 4mL of 1x isolation buffer. The sample was kept on ice until used in subsequent steps.

Cells were immediately transported to a cell sorting facility and sorted on a BD FACSAria II with a 100um nozzle, and HEPES buffer as sheath fluid. Samples were diluted in a 1:2 ratio with 1X PBS to reduce the cell population for accurate cell sorting. A 685/35 nM bandpass filter was used to detect chloroplast autofluorescence, and a 525/50 nM bandpass filter was used to detect fluorescence of AlexaFluor488-tagged haustoria. These differed slightly from confocal microscopy bandpass filters due to cell sorter restrictions.

`

VII. Confocal Microscopy of the Haustorial Fraction

To confirm the association of WGA-AlexaFluor488 with haustoria, 2uL of the final haustorial fraction was pipetted onto a clean microscope slide and overlaid with a cover glass. The cover glass was sealed with clear nail polish, and the slide was inverted when placed into the confocal microscope. Two 488nm laser channels were used: one at a 530/30 nM bandpass filter to excite the AlexaFluor488, and another at a 695/40 nM bandpass filter to excite chlorophyll autofluorescence.

VIII. Haustorial Protein Database Creation

A transcriptome built with the Trinity transcript assembler tool was retrieved for *E. necator* isolate LNYM and contained partial-length mRNA sequences (Barnett, 2015). In order to convert the partial sequences to full-length transcripts, the reads were aligned to full length transcript sequences derived from a different variant of *E. necator* known as C-strain (Jones et al., 2014). BLASTN was used to map the sequences to their full-length transcripts. The top scoring significant alignment (E-value < 0.005) for each read was saved as a representative full-length transcript, and the set of full-length transcripts was once again aligned using BLASTX against the C-strain protein database using the BLOSUM80 matrix to select for highly similar sequences with relatively low evolutionary divergence, given that the subject and query were of the same species. This produced the final database of LNYM haustorial protein sequences for use in mass spectrometry identification of peptides.

`

RESULTS

I. Literature-reported Subcellular Localization Prediction Accuracy

An accuracy assessment of intracellular protein predictors conducted by (Sperschneider et al., 2017) as well as individual plasma membrane-localized and classically secreted proteins determined the utility of including the LOCALIZER, TargetP, NLStradamus, TMHMM, and SignalP algorithms in the SecretePEPPr pipeline (Figure 3).

Figure 3. Literature-reported prediction accuracy of subcellular localization prediction tools.

`

II. Grape Genome Annotation Assessment

The CHRg genome, released by the Australian Wine Institute, was of particular interest because the APSp protein set was isolated from Chardonnay grapes, and *in vitro* validation was planned to be conducted on Chardonnay grape leaves (Roach et al., 2018). The Chardonnay genome was sequenced by the Australian Wine Institute using the PacBio SMRT sequencing platform and resulted in a 490 Mb primary assembly with an N50 of 935.8 Kb and 115X sequencing depth. Notably, the Chardonnay I10V1 genome annotation dataset contained more proteins than the other three genome annotations analyzed with SecretePEPPr (Table 2) but had more fragmented or missing genes based on BUSCO analysis (Figure 4).
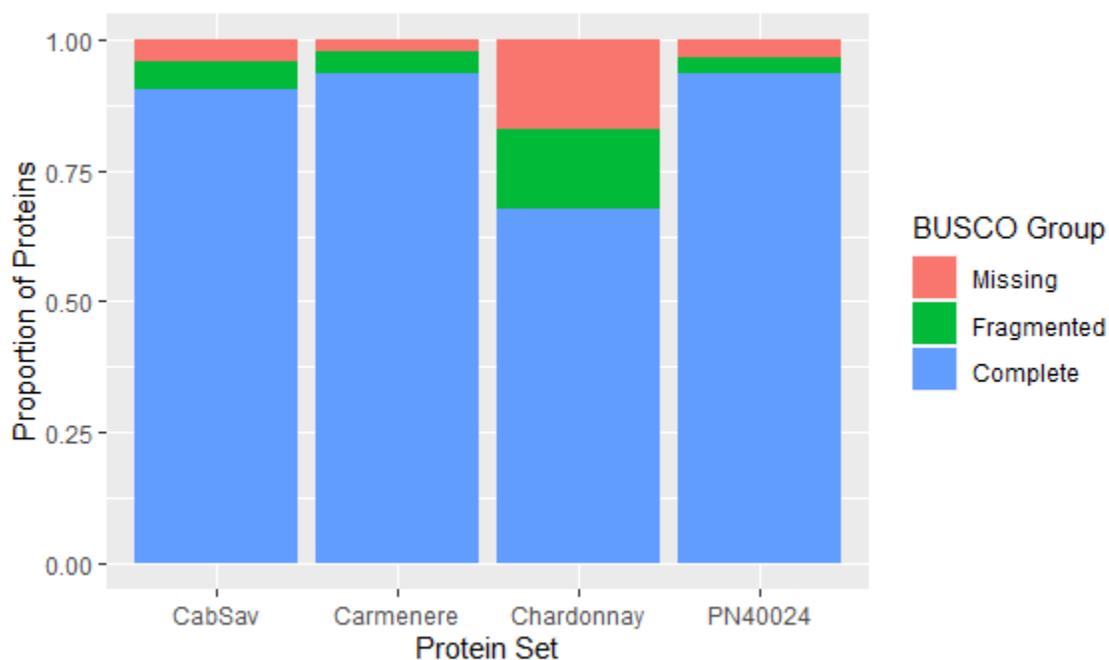
Figure 4. All genomic annotation data sets across four cultivars contain most BUSCOs. Chardonnay proteins contain the most fragmented BUSCO groups and proportionally more missing BUSCOs. Complete BUSCOs are present in the proteome at their full length, Fragmented BUSCOs are present in the proteome and truncated, and missing BUSCOs are not detectable in the proteome through gene model homology search.

`

III. Identification of Apoplastic Protein Homologues in I10V1 Genome Annotations

The set of apoplastic proteins identified by (Delaunois et al., 2013) were found to be mapped to an early version of the *Vitis* genome , and a homology search with BLASTP was conducted to map each apoplastic protein entry to a protein in the CHRg protein set. The scripts *create_Delaunois_fasta.py* and *Delaunois.R* were used to match and extract homologues to APSp proteins in the CHRg proteome. The distribution of function in the 66 apoplastic protein homologues in the Chardonnay genome is shown in Figure 5.

`
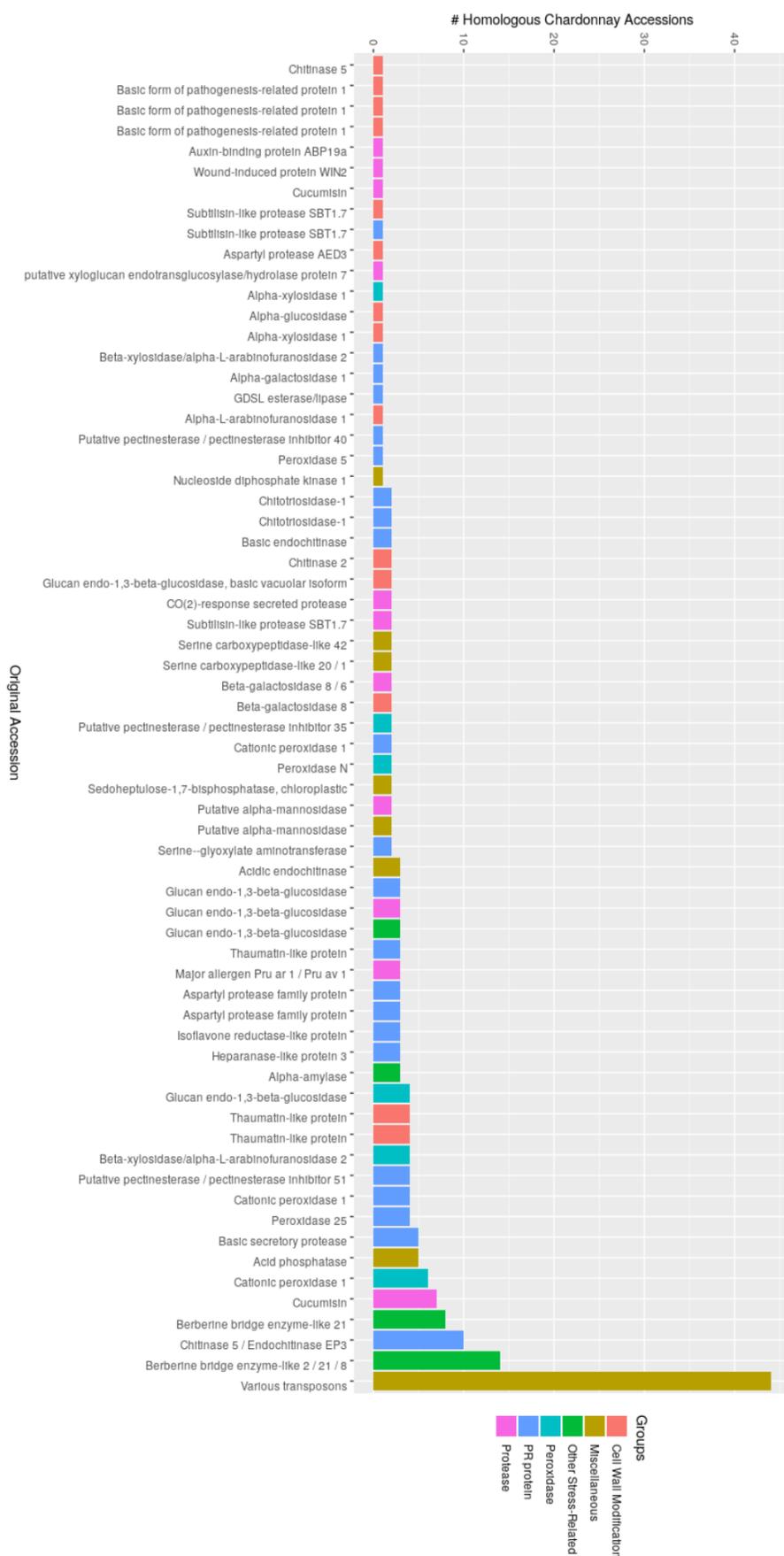
Figure 5. Distribution of homologues between original APSp and CHRg protein sets. Each of the 60 APSp proteins and their functional annotation are shown on the vertical axis. Bars correspond to the number of homologues in the CHRg genome. The single best homologue in CHRg for each original accession was used for downstream analysis.

Transposon homologues were overrepresented in the data set and were subsequently excluded from the protein set used to parameterize the SecretePEPPr pipeline.

Of the 66 proteins in the initial test set, 60 had an exceptionally high scoring match to a protein in the 115X Chardonnay genome. As could be expected, many proteins with the same functions had the same high-scoring matches in the Chardonnay genome assembly and could therefore introduce bias. Therefore, the initial test set was comprised of only the top match for each unique protein entry between the CHRg and APSp data sets. In the initial pipeline design, all proteins were run through all predictors to generate an exhaustive list of localization predictions for each individual protein. However, the run time of the pipeline was severely negatively impacted by this workflow, and an initial run of the 112,320 proteins in the CHRg assembly annotations took 60 hours to complete. Of the predictors in Figure 2, only LOCALIZER and SignalP natively supported parallelization, and two others (KohGPI and TargetP) required splitting the input FASTA file, as it was too large. To improve runtime performance, the pipeline was restructured to first run the positive, secretion predictors (TargetP, SignalP5, and ApoplastP) and then extract from the entire set of proteins only those which were predicted to be secreted. This FASTA file was then passed to the rest of the predictors and parsed the same way as before. Runtime was reduced to 11 hours for the set of 112,320 Chardonnay proteins, and 83% (50) of APSp positive test proteins were correctly predicted by SecretePEPPr as secreted. The exclusion of 17% of the APSp data set could be attributed to the presence of intracellular signaling motifs or transmembrane domains, which would be screened out as part of the prediction process.

IV. Protein Length Distribution Analysis

A comparative analysis was conducted to assess the distribution of protein lengths between proteomes before and after being run through the SecretePEPPr pipeline.

`
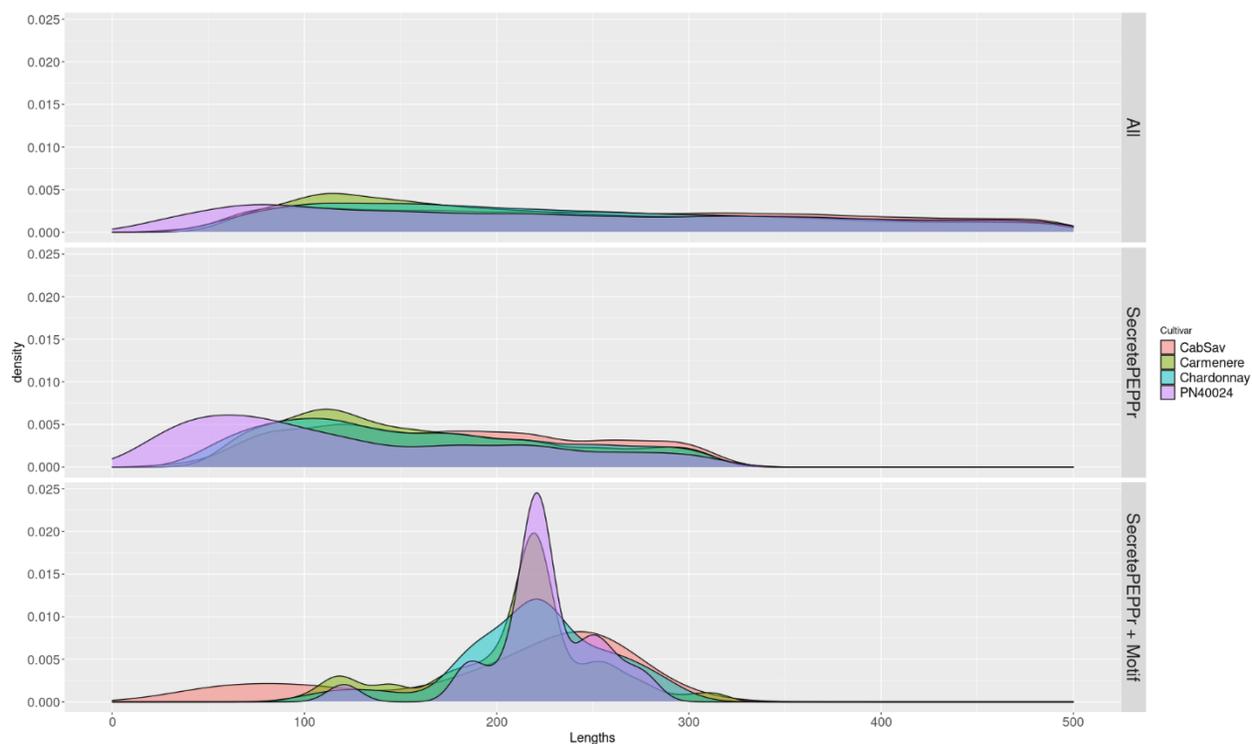
Figure 6. Protein length distributions for original and predicted proteomes. Protein lengths are cut off at 500 amino acids for visibility.

Protein length distributions across all four cultivars were very similar, especially when comparing entire proteomes (Figure 6). Proteins output by SecretePEPPr were overall slightly enriched at all lengths below 315 amino acids and did not extend past that length, as this was the size cutoff. Following selection with SecretePEPPr and following additional filtering through endocytic motif finding, there was a considerable uptick in proteins having lengths between 215 and 250 amino acids for PN40024, Chardonnay, and Cabernet Sauvignon proteins.

`

X. Individual Prediction Tool Evaluation

An initial test run was performed using WoLF PSORT, SignalP 5.0, and TMHMM without custom cutoffs. Proteins predicted to be classically secreted by both WoLF PSORT and SignalP 5.0 but not present in TMHMM predictions were considered the final predicted secreted protein set. Through this approach, only 15 proteins from the APSp protein set were recovered in the final list of predicted proteins. Later, the predictor ApoplastP was also added to the pipeline due to its ability to predict non-classically secreted proteins. To determine which predictors would be used in the final SecretePEPPr pipeline, an analysis was performed in which the benefit of adding a new predictor was compared to the number of secreted proteins already predicted by another tool at variable parametric cutoffs. Table 3 depicts this comparison between SignalP and ApoplastP in relation to the list of 60 test proteins. It was subsequently discovered that WoLF PSORT was not contributing any unique proteins to the result and was subsequently removed from the pipeline. The inclusion of ApoplastP in the pipeline allowed for 6 additional proteins in the positive test set to be predicted, at the default cutoff.

Table 3. Cutoff Threshold Determination for Protein Secretion Algorithms. Classically secreted proteins were predicted by SignalP, and LSPs were predicted by ApoplastP. Results shown are with respect to the APSp protein set.

Cutoff	Predicted Secreted (Total)	Predicted Secreted (SignalP)	Predicted Secreted (ApoplastP)	Predicted Secreted (Both)	Predicted Transmembrane (TMHMM)
Default	48	42	33	27	16
% Total	80%	70%	55%	45%	26%
0.75	37	37	10	10	14
% Total	61%	61%	16%	16%	23%
0.85	37	37	7	7	6
% Total	61%	61%	11%	11%	10%
0.95	33	33	0	0	0
% Total	55%	55%	0%	0%	0%

VI. Pipeline Performance Evaluation

In the APSp set, 83% of proteins were predicted by the pipeline to be secreted and not membrane-bound. The ten missed proteins were composed of mainly enzymes such as two pectinesterases, as well as

`

three proteases and a diphosphate kinase. Interestingly, all nine chitinases in the test set were predicted as apoplastic as well as all eight PR proteins, including three thaumatins and two berberine bridge enzymes. Chitinases and PR proteins are directly relevant to fungal infection, and these proteins were present in the test set even though the authors extracted proteins from an uninfected host. Both aspartyl proteases and serine proteases were present in the APSp set (n = 7), and only one of each was incorrectly predicted by the pipeline.

Prediction of secretion in the CRP+ dataset was less successful, with only 38% of secreted proteins predicted by the pipeline to be secreted. Since the pipeline specifically selects for proteins that are not membrane-bound, these proteins can be excluded as members of the positive test set. eggNOG annotation does not indicate whether a protein is likely to take part in GPI-anchoring, so this is not a quality of the test proteins that could be empirically determined. The pipeline also stringently excludes any proteins with intracellular localization motifs, which can be present in intercellular signaling peptides and others. 25% (18 out of 70) of proteins in the CRP+ data set that were marked incorrectly predicted by SecretePEPPr were found to be transmembrane following re-annotation. As a result, the true positive prediction rate for CRP+ proteins was expected to be significantly artificially deflated due to the intentional filtering of transmembrane and GPI-anchored proteins. Like the APSp protein set results, all chitinases were predicted by SecretePEPPr to be truly secreted.

In CRP-, 459 out of 481 (95%) of proteins were correctly predicted as non-secreted by SecretePEPPr. Similarly, 119 out of 124 (96%) of proteins from the ATN- set were correctly discarded by the pipeline.
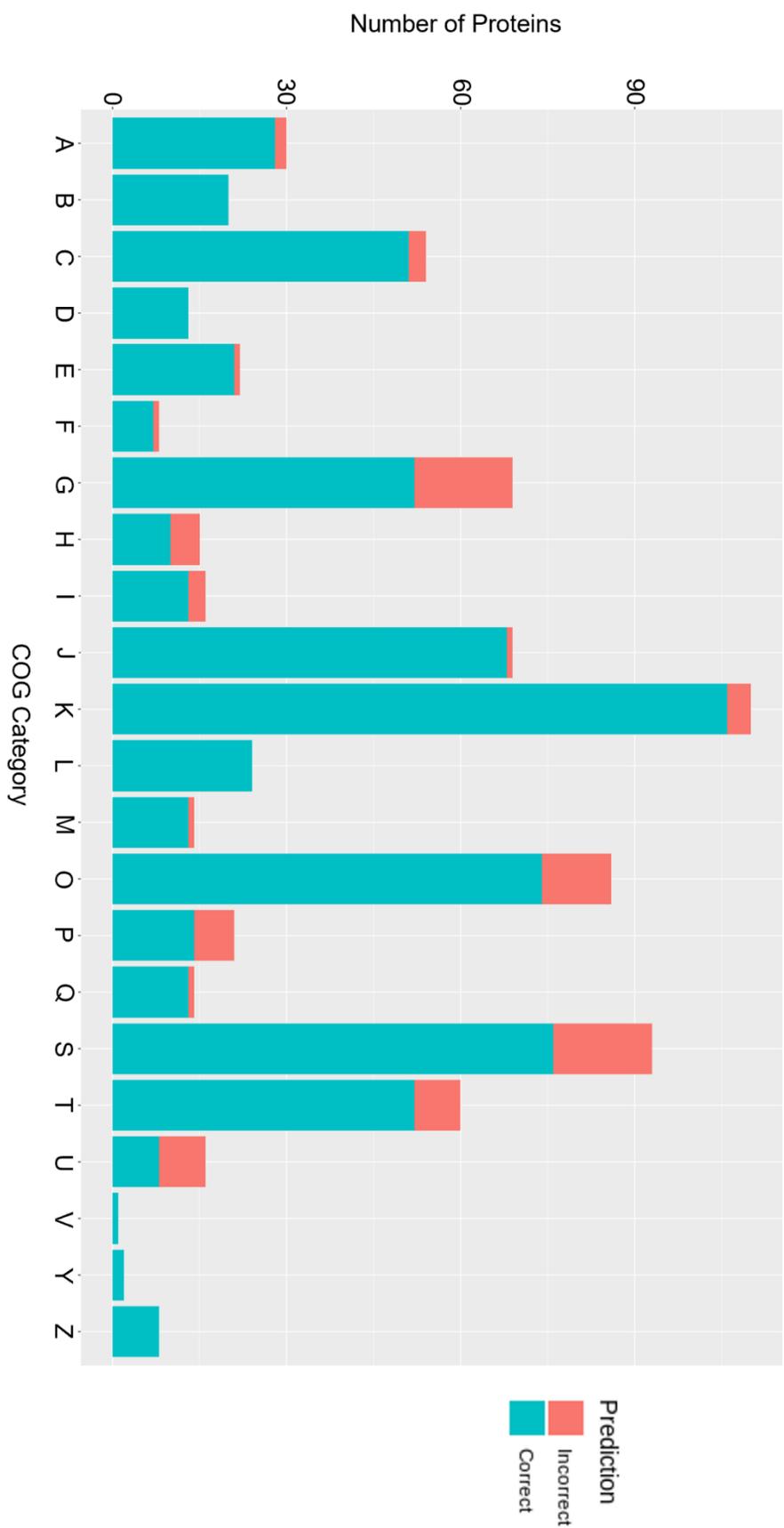
`

Figure 7. SecretePEPPr prediction accuracy of test protein sets by COG category. Blue indicates correct subcellular localization prediction (e.g. intracellular or extracellular), red indicates incorrect subcellular localization prediction. Proteins shown are from the CPL+, CPL- and ATN- test data sets.

To determine if putative function affected accuracy of SecretePEPPr results, the CRP+, CRP-, and ATN- datasets were by SecretePEPPr and functional COG categories were annotated by eggNOG (Figure 7). Highly represented COG categories containing 60 or more proteins included carbohydrate metabolism and transport (G), post-translational modification and chaperone functions (O), unknown function (S), signal transduction (T), transcription (J) and translation (K). Several COG categories had 10 or fewer proteins, including defense mechanisms (V), nuclear structure (Y), cytoskeleton (Z) and nucleotide transport/metabolism (F). SecretePEPPr displayed perfect classification accuracy across categories consisting of defense mechanisms (V), nuclear structure (Y), cytoskeleton (Z), replication, recombination, and repair (L), cell cycle control (D) and chromatin structure (B) (Figure 7). Categories with less than 80% correct predictions by SecretePEPPr include carbohydrate transport (G), coenzyme transport (H), post-translational mechanisms (O), unknown function (S), and intracellular trafficking (U).

As indicated by the area under the curve in Figure 8 the SecretePEPPr pipeline can correctly classify proteins as intracellular or truly secreted 76.3% of the time based on the test set of 766 positive and negative test proteins. Given the architecture of intracellular localization prediction in SecretePEPPr and known ambiguity surrounding transmembrane protein prediction, it is not surprising that the pipeline has a notably higher specificity than sensitivity. This was addressed with the inclusion of checks for agreement between intracellular protein prediction algorithms such as TargetP and LOCALIZER. Additionally, the inclusion of proteins related to the cell membrane biogenesis (COG category M) and ion transport (COG category P) may be artificially deflating the predictive confidence of SecretePEPPr, as membrane-bound proteins are explicitly discarded by the pipeline and not easily distinguished from secreted proteins by many automated annotation methods.

`
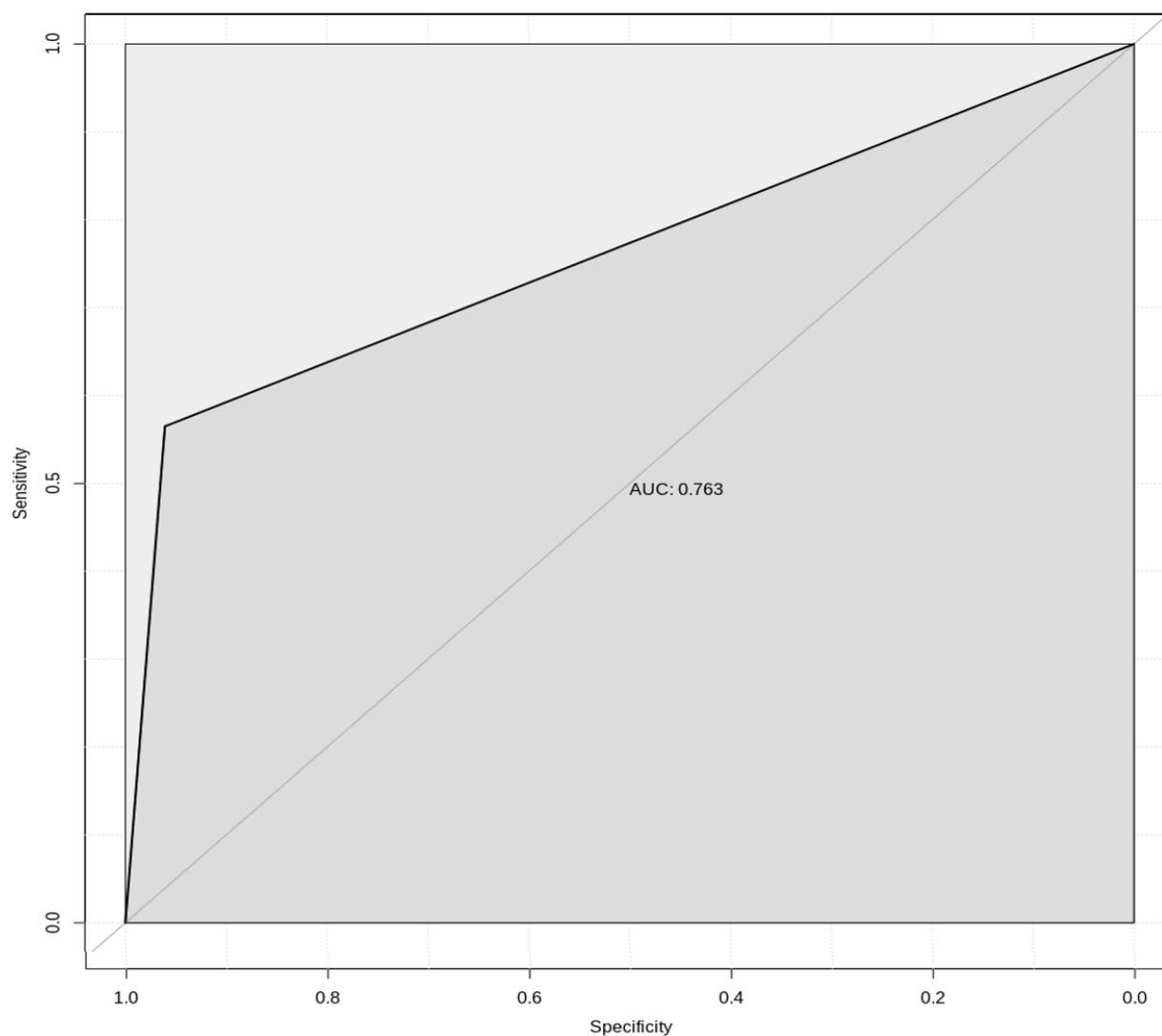
Figure 8. SecretePEPPr Performance Across all Test Sets. The faint gray line bisecting the square indicates a theoretically indistinguishable True Positive and False Positive rate.

VII. Known Effector-like Protein Characterization

Table 4 shows the results of effector-like validation methods conducted by (Plett et al., 2017) for

the three *P. trichocarpa* proteins, as well as secretion prediction by SecretePEPPr and the number of

grape genomes analyzed in this study which had at least one homologue to the given protein., as well as

their secretion prediction by SecretePEPPr and the number of *Vitis* secretomes containing homologues to

the proteins.

`

Table 4. *P. trichocarpa* proteins capable of localizing within fungal hyphae.

Pt Protein	Length	Secretion Prediction Confidence (Plett et al, 2017)	Secretion Confirmed in Yeast	Secretion Predicted by SecretePEPPr	*Vitis* genomes with Secretion Predicted by SecretePEPPr
Potri.009G063200	69	High	N	N	0
Potri.007G006800	84	High	Y	Y	4
Potri.010G251000	69	Medium	Y	Y	1

Gene Ontology, Pfam domain, and eggNOG annotations were inconclusive for the three proteins, except

for Potri.007G006800.1, which was identified as a phytosulfokine (PSK), a short disulfated peptapeptide

that is universally present in higher plants and binds to a transmembrane receptor protein with

extracellular leucine-rich repeat (LRR) regions (Matsubayashi et al., 2002; Sauter, 2015). Subsequently, a

multiple sequence alignment was performed on the three *P. trichocarpa* proteins using the MUSCLE

alignment method on the EMBL-EBI protein alignment web service to elucidate their sequence similarity

(Figure 9) (Madeira et al., 2019).

```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)


Potri.009G063200.1      MAFAAMKLFAAAVVMAMLASVAVSAQDLGELAPAPAPGKDKGAASFS------------
Potri.007G006800.1      --MANVKVTTLFLIVSLLLC------STLTYAARPEPGFPNGSLAKNQQKVVDAEHAEVM
Potri.010G251000.1      --MGAMQIFD---VRDKLIK------KQGAGSACPCCGGPVMAMDYD------------
                          :. :::      :   *          .   :. * *    :   .

Potri.009G063200.1      ------LGMSGALI-----CSSLFLSMLSLLRH------
Potri.007G006800.1      EESCEGVGEEECLM---RRTLAAHTDYIYTQKHKP----
Potri.010G251000.1      ------SHLYFCFIPISHRNKRKFSCVVCSRRLVPVPVQ
                              .::            .    :   .
```

Figure 9. Regions of Similarity Among Three Validated Effector-like Plant Proteins. * Indicates identical residues across all sequences, : indicates a high degree of conservation among substitutions, and . indicates a moderate degree of conservation among substitutions.

`

The alignment shows sparse regions of local similarity among the three short protein sequences. An alignment between three proteins is not sufficient to determine a shared motif or conserved region, but it does highlight the value of performing a BLAST search against *Vitis* proteomes with all three proteins, as they are sufficiently divergent to identify unique grape homologues.

Model Organism Support for Effector-like Protein Secretion

Homologues to the three *P. trichocarpa* effector-like proteins from the model organism *Arabidopsis thaliana* were found in Phytozome 12 (*Arabidopsis thaliana* Araport11) and evaluated for literature-based evidence of secretion (Table 4). The SUBA4 subcellular localization database was used to identify *Arabidopsis* proteins which had their subcellular localization confirmed by MS/MS or FP tagging (Hooper et al., 2017). Potri.007G006800.1 had a 66.7% similarity to AT3G49780, a phytosulfokine 4 precursor with a subcellular localization in the Golgi apparatus confirmed by MS/MS and a total length of 79 amino acids.

`

VIII. *Vitis vinifera* **Annotated Proteome Screening for Candidates**

Each of the four *Vitis vinifera* proteomes (REFg, CHRg, CABg, CRMg) derived from genome annotations were run through the SecretePEPPr pipeline for secretion prediction. BLASTP of the three *Populus trichocarpa* proteins was performed on the FASTA files output by SecretePEPPr (Table 4).

Potri.007G006800.1 was the only query with homologues found across all four predicted secretomes, and Potri.010G251000.1 had a single homologue in the reference grape genome annotations. 21 *Vitis vinifera* proteins with strong homology (E $<= 2x10^{-15}$) to Potri.007G006800.1 from across all four predicted secretomes, and one grape protein with strong homology to Potri.010G251000.1 from the predicted secretome of the reference *Vitis vinifera* PN40024 genome annotation (E $= 5x10^{-22}$) were identified as strong candidates for *in vitro* validation. No homologues of Potri.009G063200.1 were found in any of the four predicted secretomes, but its three closest homologues in the reference grape genome annotations were included in the final candidate set (E $<= 6x10^{-8}$). This resulted in a total of 25 *Vitis vinifera* homologues to two previously identified effector-like proteins in *Populus trichocarpa*.

`

IX. Endocytic Motif Screening

To provide additional support for protein localization within fungal cells, four known endocytic motif patterns were screened against the 25 candidates and the three originally identified effector-like proteins. The peptide motifs YXXΦ (where Φ is a bulky hydrophobic residue), [D/E]XXXL[L/I], and FXNPXY have been associated with clathrin-mediated endocytosis, and the NPFXD motif has been implicated as an endocytosis signal in yeast (Kozik et al., 2010; Tan et al., 1996). Through a regular expression analysis of the sequence data, one match to the YXXΦ motif was found in Potri.010G251000.1 with the sequence YFCF, and no other matches were found. This suggests the possibility of an endocytosis-independent localization pathway or the presence of previously undiscovered endocytic motifs.

The two most prevalent motifs found in all four *Vitis vinifera* cultivars at the top five percent of identified motifs were YFNL and YVGF, neither of which were found in the *Populus trichocarpa* effector-like proteins. YXXΦ was the most commonly detected potential endocytic motif, followed by [D/E]XXXL[L/I], FXNPXY, and finally NPFXD. The number of proteins in the entire proteome, SecretePEPPr-screened proteome, and most common endocytic motif-containing SecretePEPPr-screened proteome is shown in Figure 10.

`

X. Candidate *Vitis* Effector-like Protein Repertoire
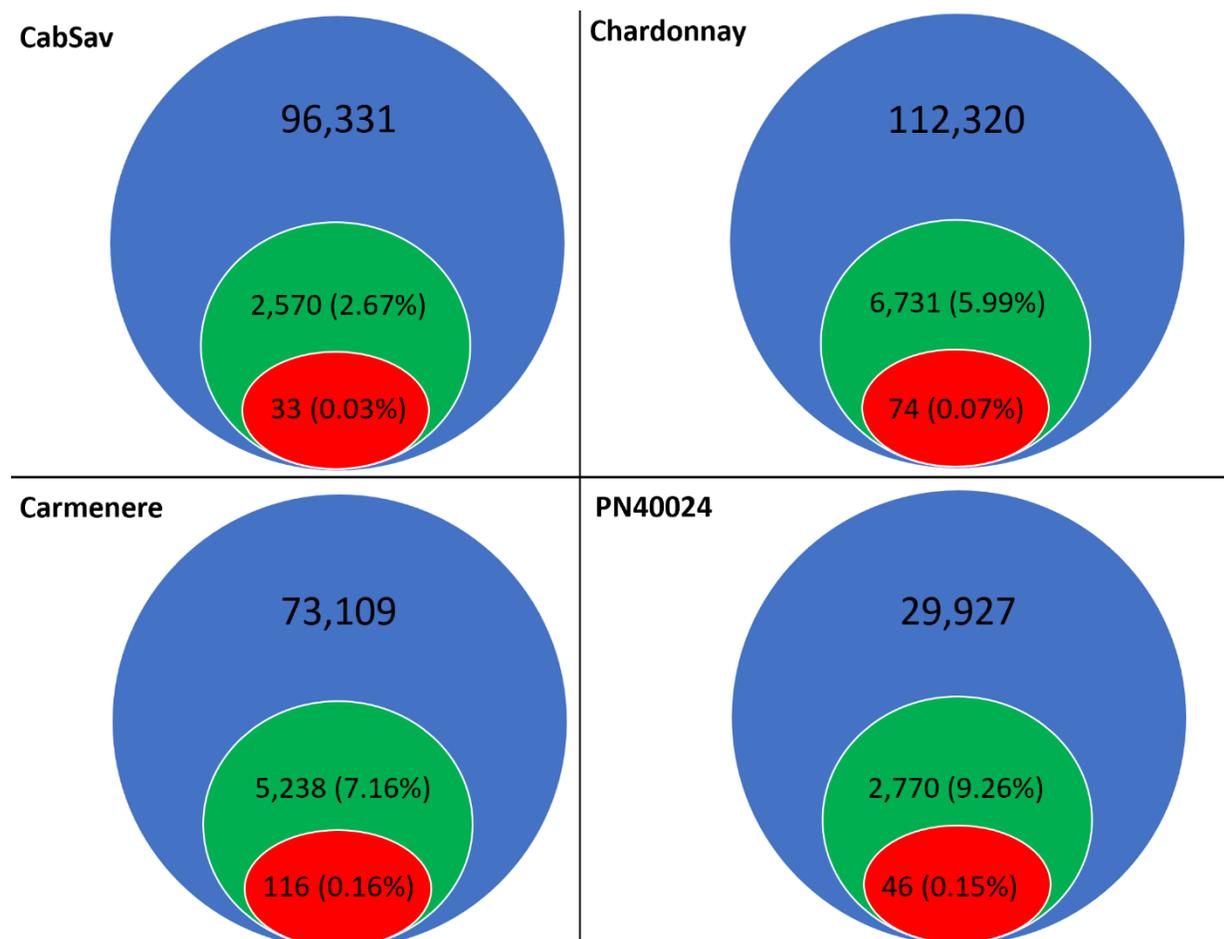
Figure 10. *Vitis* protein sets by predicted category. Blue indicates the entire proteome for that cultivar, green is a smaller subset predicted by SecretePEPPr without endocytic motif filtering, and red indicates a targeted list of proteins indicated to contain the most common endocytic motifs.

Targeted effector-like protein sets, intended for assays like GFP-tagging and targeted proteomics, ranged from 116 to 33 proteins across the four cultivars (Figure 10). Without endocytic motif identification, between 2.7% and 9.3% of the total proteome across cultivars was indicated as potentially effector-like based on secretion and size exclusion. COG categories of proteins in each cultivar were assessed (Figure 11). Category S (unknown function) was consistently abundant in protein sets, particularly in the SecretePEPPr + Motif protein sets where most proteins were of COG category S. Category T (signal transduction) was also represented in three of the four cultivars in the SecretePEPPr + Motif protein sets.

`
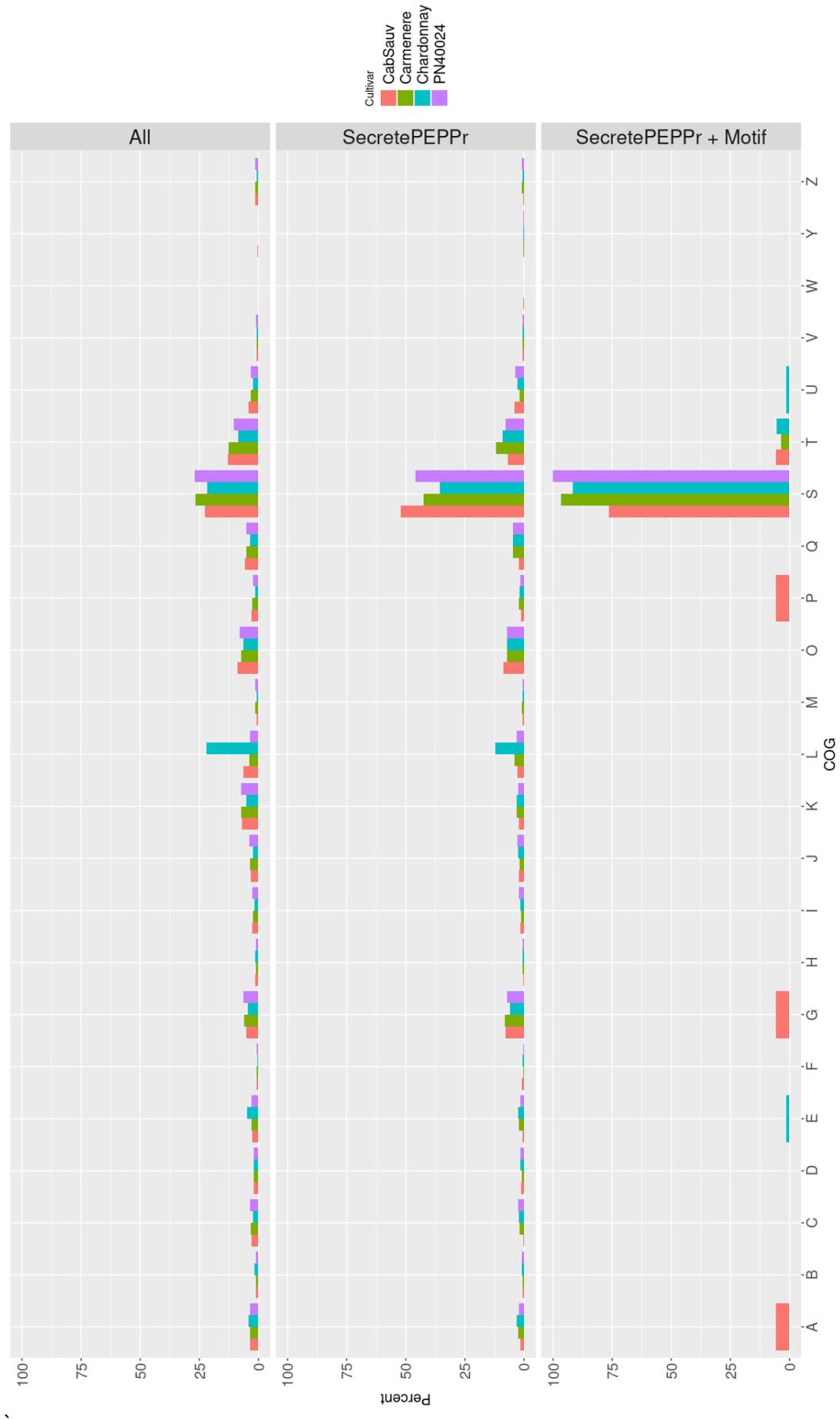
Figure 11. COG categories of proteins across four cultivars. Data includes whole proteome (All) broad secretome (SecretePEPPr) and targeted secretome (SecretePEPPr + Motif).

For functional categorization of candidate *Vitis* effector-like proteins, the SecretePEPPr + Motif protein sets (red groups in Figure 10) from all four cultivars were re-annotated with eggNOG mapper using *Viridiplantae* as the taxonomic scope to limit results to plant orthologues. The most represented protein orthologues across all four cultivars were germin-like proteins, followed by various expansin proteins, and photoassimilate-responsive proteins (PAR1). Chardonnay, PN40024, and Carmenere candidate proteins included a lipid-transfer protein, and Carmenere and PN40024 protein sets also contained three total cyclase family protein homologues. Other noteworthy homologues to proteins in individual cultivars included two epidermal patterning factor proteins in Chardonnay and Carmenere, two zinc finger proteins in Cabernet Sauvignon, a coatomer subunit protein in Chardonnay, a thaumatin-like protein in Cabernet Sauvignon, a pectinesterase protein in Cabernet Sauvignon, and octicosapeptide Phox Bem1P-containing proteins in Carmenere. These results are summarized and ordered by abundance in Table 5 below.

XI. Transcriptomic Support for Candidate ELPs

A comparison of protein homologues across all SecretePEPPr + Motif candidate ELP sets and the two expression studies used for suggestive evidence can be found in Table 5. Both analysis comparing candidates to expression data reflected an abundance of effector-like germin-like protein homologues, a thaumatin-like protein, a coatomer subunit, and lectin-domain containing receptor kinases at 1-7 days post infection (Table 5).

`

Table 5. Functional Annotation of ELPs. ELPs predicted from two sets of expression data are included. PN40024, Cabernet Sauvignon, and Carmenere predictions for Barnett, 2015 were not included because an exact cultivar match was available (Chardonnay).

Cultivar	SecretePEPPr + Motif (Predicted ELPs)	Barnett, 2015 Predicted ELPs	Weng et al. 2014 Predicted ELPs
PN40024	Germin-like proteins Expansins PAR1 proteins Cyclase family proteins Lipid transfer proteins	N/A	Germin-like proteins
Cabernet Sauvignon	Germin-like proteins Expansins Zinc finger proteins LRR receptor-like serine threonine-protein kinase Pectinesterase Thaumatin-like protein PAR1 protein	N/A	Germin-like proteins Thaumatin-like protein
Carmenere	Germin-like proteins Expansins PAR1 proteins Octicosapeptide Phox Bem1p Lipid transfer protein L-type lectin-domain containing receptor kinase Epidermal patterning factor protein Serine threonine-protein kinase	N/A	Germin-like proteins

`

	Cyclase family Lectin-domain containing receptor kinase Rare lipoprotein A (RlpA)-like double-psi beta-barrel containing protein Plant-type cell wall organization		
Chardonnay	Germin-like proteins Expansins PAR1 protein Lectin-domain containing receptor kinase Lipid transfer protein Plant-type cell wall organization Coatomer subunit Epidermal patterning factor-like protein Cationic amino acid transporter	Germin-like protein Lectin-domain containing receptor kinase Coatomer subunit	Germin-like proteins

XII. Confocal Microscopy of Haustoria Isolated from a Density Gradient Extraction

Confocal microscopy was performed on the final haustorial fraction following Percoll density gradient extraction. Two simultaneous channels were active on the confocal: one consisting of a 530/30 nM bandpass filter to measure AlexaFluor-488 fluorescence and another at 695/40 nM bandpass filter to

detect chloroplast autofluorescence. Subsequent figures depict AlexaFluor488-tagged particles in green,

and autofluorescence of chloroplasts in red.
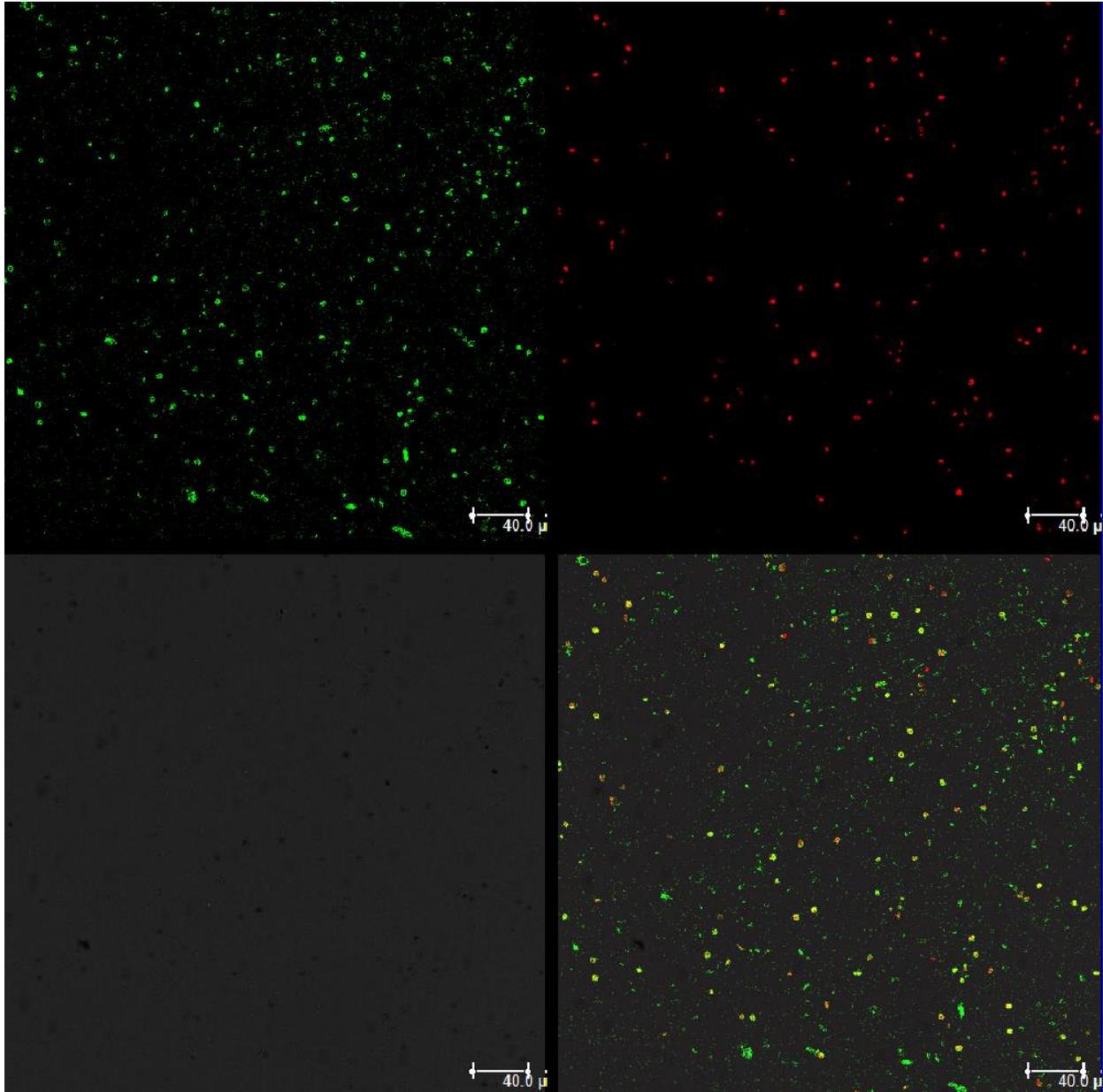
Figure 12. Overview of all confocal panes displaying haustorial sample fluorescence. Top Left: green fluorescence from only AlexaFluor488. Top Right: red autofluorescence of chloroplasts. Bottom Left: the slide with no specified scanned wavelengths (greyscale). Bottom Right: an overlaid image of all three other panes.

The initial confocal image was taken with the same scaling as the original images taken by

(Garnica & Rathjen, 2014) (Figure 12). At first glance, there was considerably more chloroplast

`

autofluorescence in Figure 2C of Garnica & Rathjen (2014) than in the PM haustorial sample in the current study (top right panel of Figure 12). This is surprising because the haustorial pellet resulting from the extraction procedure was green in appearance even after labeling with the WGA-AlexaFluor and washing. While this observation could be due to free chlorophyll in the sample, the majority of chlorophyll should have been excluded due to its comparatively lighter density than intact chloroplast. Moreover, there is a considerable size difference between particles all throughout the AlexaFluor488-stained sample imaged in this study, also shown in the top left pane of Figure 12. This likely reflects haustorial shearing and fragmentation in the process of isolation. The bottom left pane of Figure 12 displays several objects that fluoresced both red and green (showing as orange or yellow when present in tandem). Wheat germ agglutinin binds N-acetyl-D-glucosamine and sialic acid, which may be present in the membrane of chloroplasts. For FACS, this necessitated filtering first by red fluorescence at a 695/40 nM bandpass to remove the chloroplasts, and then at the 530/30 nM bandpass to isolate the haustoria.

`
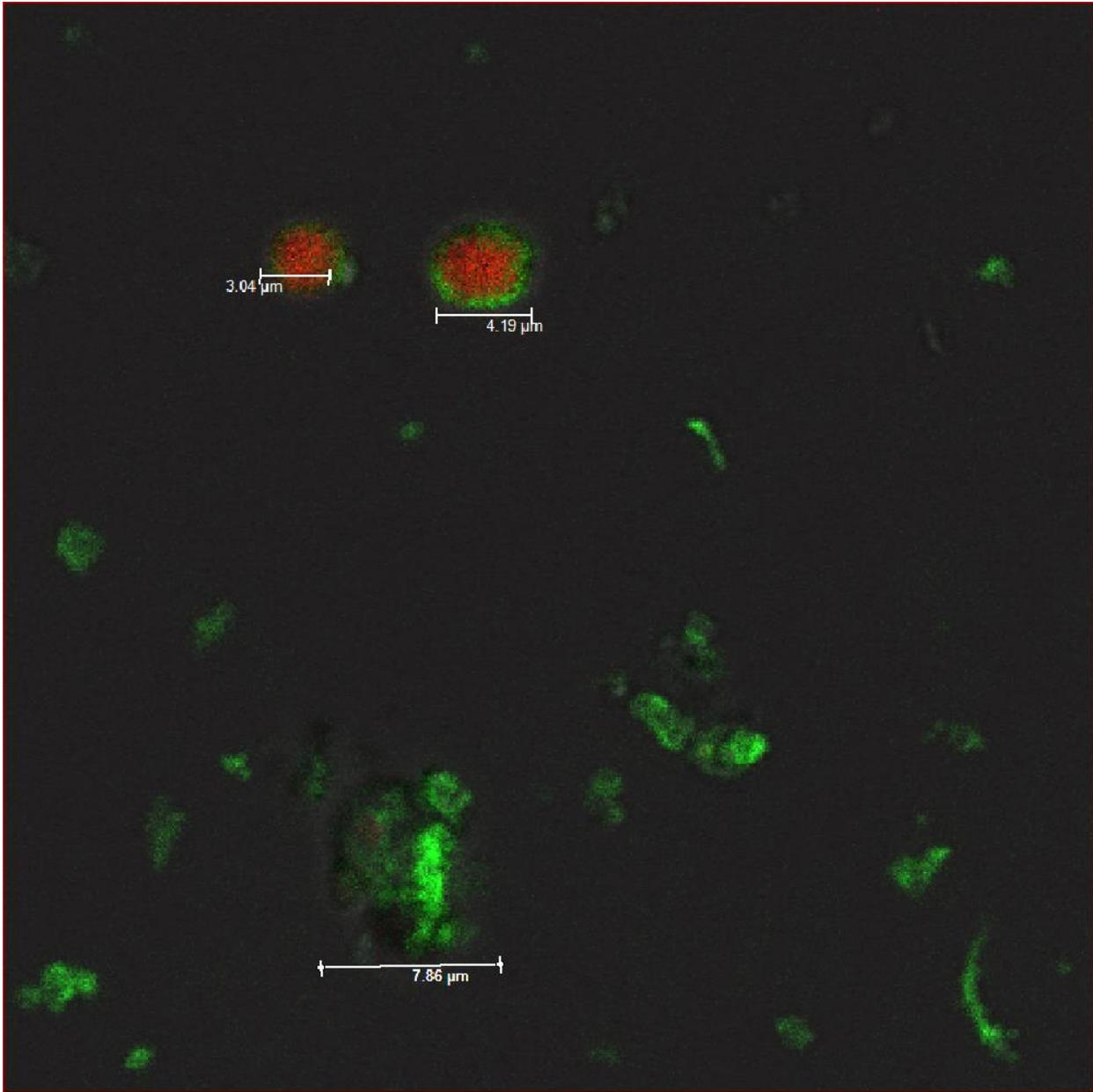
Figure 13. **Juxtaposition of fungal matter and chloroplasts.** Overlaid pane of green (AlexaFluor488) and red (chloroplast autofluorescence) zoomed in on structures suspected to be chloroplasts (top) and haustoria (bottom) with size measurements.

`
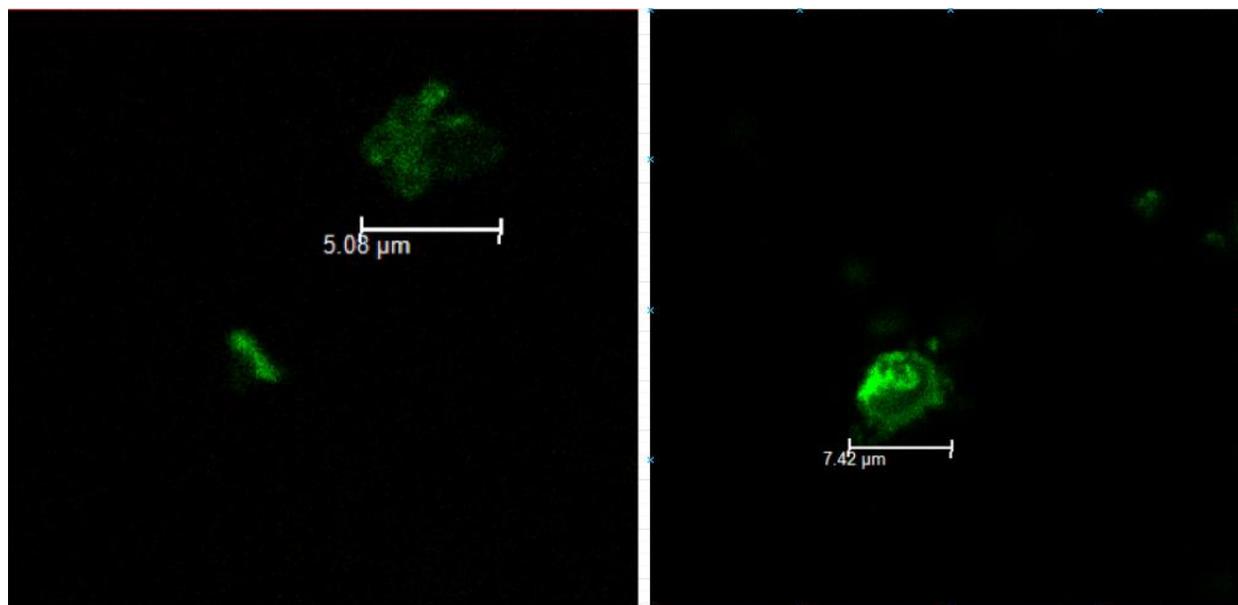
Figure 14. Additional haustorial particles with annotated sizes. Globose particles identified as haustoria are shown with their length in microns. Smaller green particles are likely ruptured haustorial particles damaged during centrifugation.

Per Figures 13 and 14, haustorial particles were around 5-7 um in size and chloroplasts were around 3-5 um in size. For discovery-based mass spectrometry and protein identification by peptide mass fragments, this was not expected to significantly affect results given a high enough yield of haustoria from the extraction protocol. Figure 14 displays two more particles suspected to be haustoria. Note that particles suspected to be haustoria are noticeably smaller than those shown from wheat rust in (Garnica & Rathjen, 2014).

`
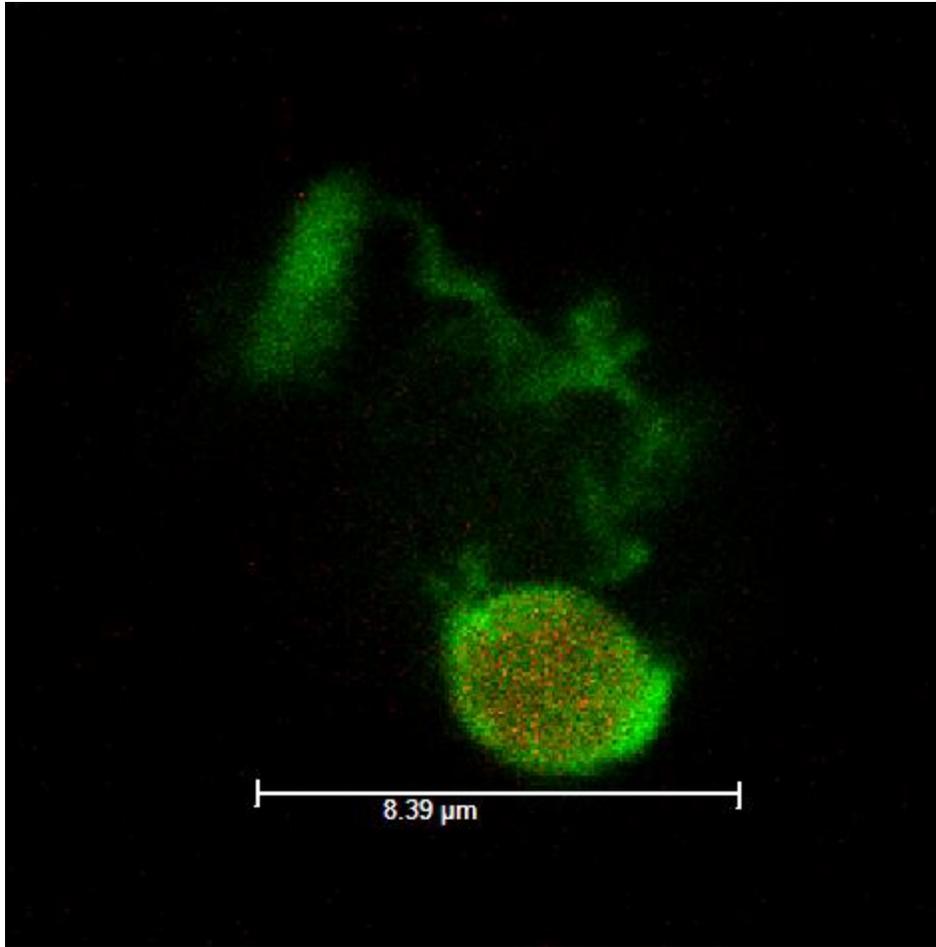
Figure 15. A chloroplast appears to be associated closely in space with a haustorium. WGA/AlexaFluor conjugate (green) is shown binding to the chloroplast at the bottom right and fluorescing, along with chloroplast autofluorescence (red).

In several instances, one or more chloroplasts appeared to associate closely with structures suspected to

be haustoria, as exemplified by Figure 15. 3D focusing of the confocal microscope helped identify

complex structures like the one shown in Figure 15.

`
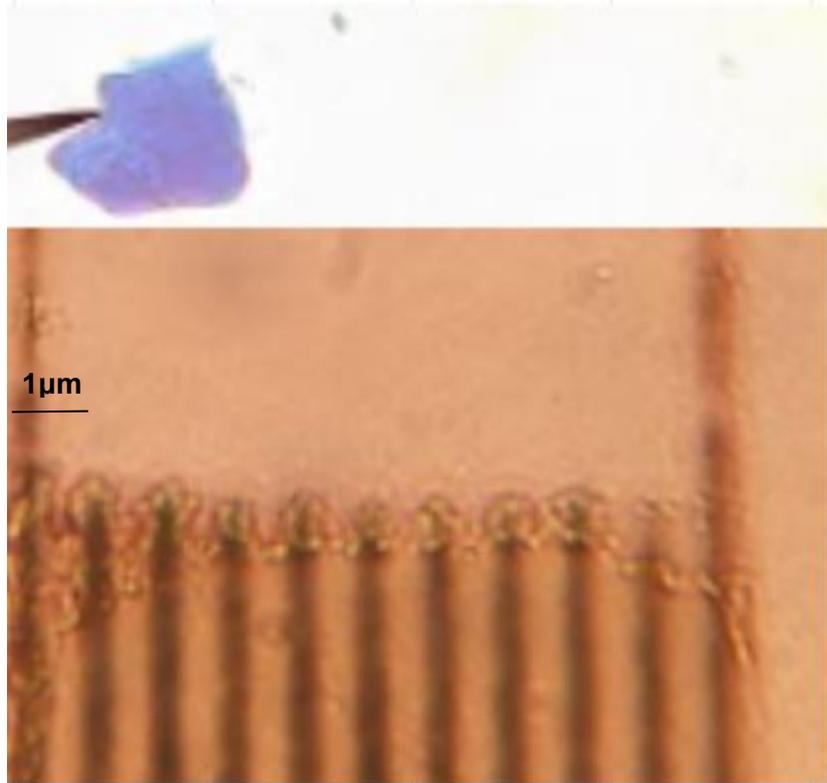
Figure 16. Lactophenol blue staining and measurement by microscope micrometer. The bottom portion represents 10 micrometers, or one micron per segment, indicating the stained structure is slightly more than 3 micrometers wide.

A concern that arose from confocal microscopy given the apparent association of AlexaFluor488 and chloroplasts was whether any of the structures present in the sample were of fungal origin. Thus, the following day the haustorial fraction was stained with lactophenol blue, which binds to chitin, and viewed under a compound microscope (Figure 16). Structures similar to those seen previously under the microscope, such as the blue particle at the top of Figure 16, were seen on the slide. Additionally, these structures appear to be similar in size (~4-8 microns) to those seen on the confocal. No discernible chloroplast-like particles were seen under the benchtop microscope after lactophenol cotton blue staining.

`

XIII. FACS of the Haustorial Gradient
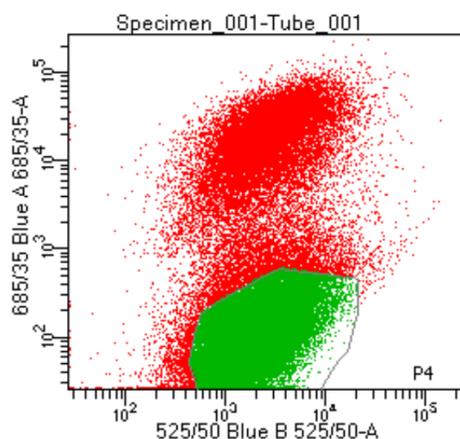
Figure 17. Capture group table and cell population graph for BD FACSAria haustorial cell sorting. Red dots indicate chloroplasts and ambiguous cell populations, green dots indicate the capture group for haustorial cells. The Y axis reflects the approximate emission range of plant chloroplasts, and the X axis reflects the emission range of AlexaFluor488 which represented the target haustorial population.

Cell sorting of fluor-tagged haustoria on a BD FACSAria II resulted in two visually distinct cell populations (Figure 17). Two distinct cell populations were present, representing the chloroplast cell population at the top and the haustorial cell population at the bottom. Pure haustorial cells should have little to no excitation in the 685/35 bandpass range, and as such the capture group for cell sorting was designated as slightly below the intersection between the two cell populations to ensure a higher proportion of pure haustorial cells in the collected sample.

`
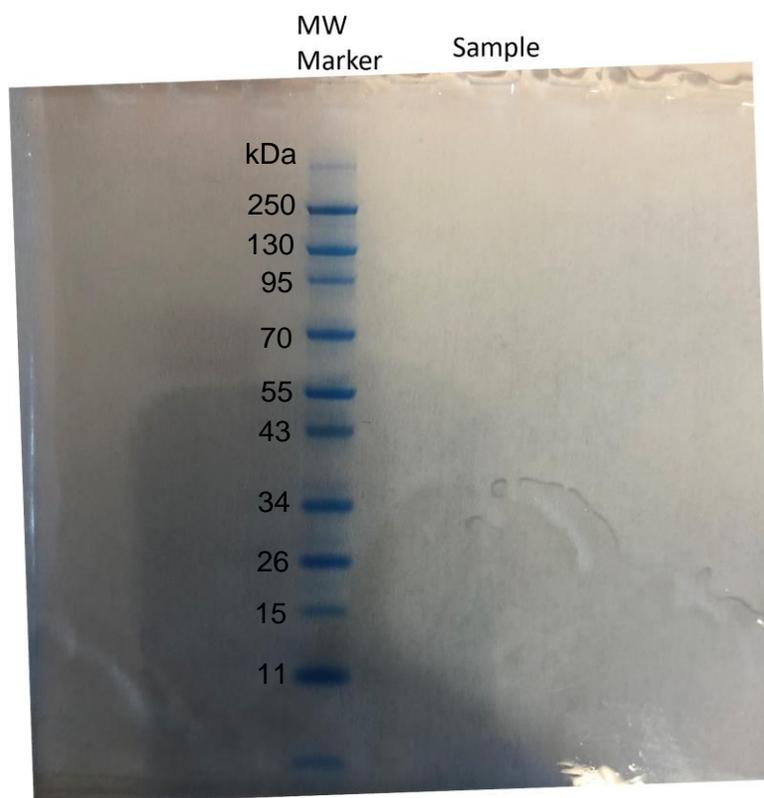
Figure 18. SDS page gel of the purified haustorial fraction. A negative result is displayed alongside a protein molecular weight marker.

Upon running an SDS page gel of the purified haustorial fraction from Figure 17, a negative result was obtained (Figure 18). At this point in the process, a considerable amount of time was spent on optimizing the haustorial extraction protocol up to and including FACS, and it was decided that prohibitively time-consuming additional steps would need to be taken to confirm candidate ELPs *in vivo*. As such, an in-depth overview of the haustorial extraction method in its current form is provided, and it is suggested to be used in further studies focusing on *E. necator* haustoria with some adjustments suggested in the Discussion section.

`

DISCUSSION

In contrast to well characterized pathogen effector proteins, plant effector-like proteins are just now being recognized for their importance in host-pathogen interactions. In this study, a standardized tool for candidate plant effector protein prediction was developed using whole-genome annotation data as the input, and four *V. vinifera* cultivars are screened for effector-like protein content. Consideration of a possible effector-like protein translocation method, clathrin-mediated endocytosis, was included through a motif searching algorithm, identifying four candidate effector-like protein classes. Finally, a haustorial isolation protocol for *E. necator* haustoria was developed, resulting in a return of 10-15 million haustorial cells from an initial sample of 20 grape leaves.

I.   Candidate Effector-like Protein Prediction

The pipeline built in this project was designed to screen grape proteins capable of being secreted via classical (Golgi-ER) and non-classical means, with consideration for additional features suggestive of the capability to translocate to fungal haustoria. This included a filter for proteins localizing to the plant nucleus, mitochondria, chloroplast, and cell membranes (i.e. GPI-anchors and transmembrane domains). In this project, a novel computational methodology is devised, capable of whole-proteome screening for a subset of plant proteins with heretofore largely unexplored function: the ability to translocate to the fungal cytoplasm and affect metabolic processes in a similar fashion to fungal effector proteins. Some current genome-guided studies on plant effector-like proteins, and grape proteins secreted in response to biotic stress more broadly, place a considerable focus on differential expression as an initial filtering tool (Plett et al., 2017; Weng et al., 2014). The genome annotation-guided approach in the SecretePEPPr pipeline allows investigators to conduct FP-tagging and proteomics studies on effector-like protein candidates without the additional up-front work of expression studies, retaining several aspects of previously successful methodologies including subcellular localization prediction, and additional considerations for translocation mechanisms of proteins between the plant and fungal cytoplasm. Additionally, genome-wide screening circumvents variability inherent to analyzing transcriptomes or proteomes at various

`

timepoints throughout infection. Non-classically secreted proteins, which are considered in SecretePEPPr annotation, are generally upregulated within the first ten hours of infection (Cheng et al., 2009). Genome-wide screening methods, particularly those that utilize genome annotations like SecretePEPPr, should be paired with adequate quality control measures such as the BUSCO assessment for annotation quality that was conducted in this study (Figure 4).

Since genome annotations inherently contain significantly higher gene counts than can be expected of RNA expression data, particularly differentially expressed genes associated with infection, certain filtering methods used in the SecretePEPPr pipeline needed to be more stringent to account for the otherwise significantly higher false positive rate. This was accomplished at the subcellular localization prediction level by excluding all proteins found to localize within plant chloroplasts, mitochondria, and nuclei. Since fungal cells do not contain chloroplasts, this screening method does not impact the prediction of proteins potentially localizing within fungal haustoria. However, the exclusion of mitochondrial and nuclear-localizing proteins precludes potential effector-like grape proteins localizing within those subcellular compartments from being included in the final set of candidate effector-like host proteins.

The decision for including these additional intracellular protein filtering rules in the SecretePEPPr pipeline are based on the inclusion of non-classically secreted proteins. Most non-classically secreted proteins have established intracellular functions and, if involved in pathogenesis, likely display dual functionality. Among non-classically secreted proteins found in the extracellular space during infection are enzymes involved in primary metabolism such as G3P-dehydrogenase, phosphoglucomutase, phosphoglycerate kinase, enolase, triose-phosphate isomerase and methionine synthase (Kaffarnik et al., 2009). While these and other proteins secreted by plants for use in intercellular communication may be directly involved in pathogenesis, the consideration of leaderless secreted proteins adds considerable potential for false positive results. Therefore, effector-like proteins of cytoplasmic function are predicted in the SecretePEPPr pipeline, and secreted proteins with mitochondrial or nuclear localization are excluded. With this filter, the set of proteins predicted by subcellular localization prediction alone

`

numbered in the thousands and the additional filter assisted in establishing a smaller set of endocytic motif-containing candidates.

Initial parameterization of predictors was accomplished using an initial set of 60 confirmed apoplastic grape proteins. This step was included to reduce potential bias against non-classically secreted proteins and provide empirically validated parameters to predictors that were not otherwise already fine-tuned to plant proteins specifically. Pipeline accuracy assessment was conducted at the subcellular localization stage, as no database or comprehensive list of effector-like protein exists currently. Experimentally validated secreted proteins from the cropPAL database, as well as experimentally validated intracellular proteins from the cropPAL database and *Arabidopsis* nucleolar proteins were used as positive and negative test sets, respectively. Graphing SecretePEPPr secretion prediction accuracy by COG category revealed that this composite data set was composed of many proteins of unknown function, and several categories with lower prediction accuracy consisted of proteins related to membrane-associated functions such as intracellular trafficking and secretion (COG category U), signal transduction (COG category T), carbohydrate metabolism and transport (COG category G), chaperone function and protein turnover (COG category O), and inorganic ion transport and metabolism (COG category P). This lower prediction accuracy is expected to be at least in part attributed to a lack of annotation for potential GPI-anchored secreted proteins in the cropPAL database and inherent difficulty in differentiating between transmembrane and signal peptide-containing proteins by TMHMM, the transmembrane helix predictor included in SecretePEPPr. Low prediction accuracy could also be exacerbated by the inclusion of non-classically secreted protein prediction in the pipeline, some of which have known intracellular functions. This is corroborated by the presence of serine-threonine protein kinases and L-type lectin domain-containing protein kinases in the final set of candidate effector-like proteins (SecretePEPPr + Motif). While these proteins may translocate to membrane components of haustoria during their formation or throughout infection, their inclusion in the results of the pipeline are unintended.

Another unique inclusion in the SecretePEPPr pipeline is identification, quantification, and filtering of proteins by the presence of endocytic motifs, including motifs involved in clathrin-mediated

`

endocytosis, which serves to add further support and explanatory evidence for translocation of the plant proteins to the haustorium. Early haustorial formation is characterized in part by a massive uptick in plant vesicular activity at the appressorial penetration site (Leborgne-Castel et al., 2010). At this stage of infection, vesicular activity related to the formation of papilla aims to prevent haustorial formation. Later in the infection, vesicles appear to aid the formation of haustoria by shuttling nutrients, and likely fungal effectors, between the cytoplasm of the pathogen and host. Clathrin-mediated endocytosis is a known signaling mechanism in plant defense, making it a valuable additional filtering criterion for determining candidate plant effector-like proteins. By default, sequences containing subsequences within the top five percent of represented potential clathrin-mediated endocytic motifs are saved in the final high-confidence set of candidate effector-like proteins. Since endocytic motifs all contain ambiguous residues, the distribution of present endocytic motifs may vary between different plant species, or even cultivars of the same species. In this study, common and highly prevalent endocytic motifs were cross-referenced between each cultivar, with YFNL and YVGF appearing in the top five percent of all grape proteins predicted as small and secreted by SecretePEPPr. Functional annotation of proteins with these endocytic motifs will be discussed in the next section.

II.   Functional Characterization of Candidates

Candidate ELPs with endocytic motifs common to the cultivars used in this study revealed possible novel functions for previously characterized and previously uncharacterized *Vitis* proteins. Throughout all predicted candidate ELP sets, with and without *in vitro* evidence of expression during infection, germin-like proteins were most abundant. Germin proteins are an incredibly functionally diverse family of proteins, catalyzing 60 or more distinct enzymatic reactions (Himmelbach et al., 2010). Germin-like proteins are defined by an average homology of 50% with wheat germin, and are found in a variety of land plants (Patnaik & Khurana, 2001). The role of germin-like proteins in plant defense against fungal pathogens is well documented, with specific transgenic studies establishing certain germin-like proteins functioning as oxalate oxidases and in cell wall strengthening capacities, reducing wheat susceptibility to

`

powdery mildews such as *Blumeria graminis* (Christensen et al., 2004). Similarly, ontogenic resistance to *E. necator* was shown to coincide with RNA expression of a germin-like protein (VvGLP3) induced by inoculation of ontogenically resistant but not susceptible berries. Expression may coincide not with thicker but perhaps stronger cuticle or cell wall that prevents penetration and haustorium formation (Ficke et al., 2002, 2004). Other examples reduced plant susceptibility without displaying cell wall enforcement functions. In barley leaves attacked by powdery mildew, PAMPS strongly induced the expression of germin-like proteins in the *HvGER4* subfamily (Himmelbach et al., 2010). Germin-like proteins across several plant species have been demonstrated to raise endogenous reactive oxygen species independently, without association with constitutive plant defense pathways or canonical reactive oxygen species synthesis pathways (Beracochea et al., 2015; Ilyas et al., 2016). FP-tagging studies in rice have observed cell surface localization of certain germin-like proteins in uninfected plants. Germin-like proteins have also been shown to trigger the expression of other pathogen defense related proteins, specifically PR proteins including those involved in the salicylic acid signaling pathway (Knecht et al., 2010).

Paradoxically, certain germin-like proteins can also exhibit superoxide dismutase activity commonly associated with the regulation of reactive oxygen species production during infection (Alscher et al., 2002; Barman & Banerjee, 2015). This suggests two potential separate and antithetical roles for germin-like proteins during infection with pathogens like *E. necator*: production of reactive oxygen species within the fungal cytoplasm, or detoxification of reactive oxygen species within the fungal cytoplasm. The implications of these two functions are discussed in the following section, and further studies are necessary to elucidate which germin-like proteins act as true ELPs during infection of *V. vinifera* by *E. necator*. Regardless, enriched presence of germin-like proteins in predicted candidate ELPs with evidence at the expression level in two separate studies suggest a mechanism for germin translocation to the haustorial interface: through vesicular clathrin-mediated endocytosis. Germin proteins with oxalate oxidase functions localizing within fungal haustoria and mediating cytotoxic oxidative stress adds a new element to the growing body of knowledge on the importance of germin proteins in plant defense and provides a mechanism for immediate investigation.

`

Other candidate ELPs with evidence at the expression level during pathogenesis include a thaumatin-like protein in *V. vinifera* cv. Cabernet Sauvignon as well as lectin-domain containing receptor kinases and a coatomer subunit in *V. vinifera* cv. Chardonnay. Thaumatin-like proteins are another large family of plant defense-related proteins, playing different roles in a variety of responses to pathogen attack. In *V. vinifera*, certain thaumatin-like proteins have been associated with stomatal closure in response to fungal PAMP detection among other roles in the salicylic and jasmonic signaling pathways central to plant defense (Yan et al., 2017). Others show direct antifungal activity, such as *VvTLP*-1 and *VvTL*-2 which have been demonstrated to directly inhibit fungal growth including in *E. necator* (Jayasankar et al., 2003; Monteiro et al., 2003). Additionally, thaumatin-like proteins can display glucanase activity and xylanase inhibitor acticity (J.-J. Liu et al., 2010). Lectin-domain containing receptor kinases, like serine-threonine protein kinases also seen in candidate ELPs, are generally cell surface transmembrane domain-containing proteins that act as the products of important resistance genes involved in the detection of PAMPs and activation of PTI (Afzal et al., 2008; Sun et al., 2020). Their inclusion in pipeline results is likely a false positive along with a few other transmembrane proteins, but their involvement with cell membranes and the presence of some transmembrane proteins in the plant-derived extrahaustorial membrane of proteins suggests a potential role for receptor-like kinases in the extrahaustorial matrix following haustorial formation.

Finally, a coatomer subunit protein candidate ELP was found within proteins with evidence of expression during pathogenesis. Coatomer subunits make up a cytosolic complex of non-clathrin coated Golgi vesicles related to membrane budding (Waters et al., 1991). In plants, coat protein-coated vesicles localize to microvesicles surrounding the Golgi apparatus. Their potential role in shuttling of effector-like proteins is presently unknown, as clathrin-mediated endocytic motifs were prioritized in this study. However, they do exemplify how non-clathrin mediated vesicle transport of candidate ELPs is also likely and remains to be explored in future studies. Experimental evidence at the transcriptional level resulted in a narrow set of proteins, and this is expected in part to be due to the timing of sample collection and sequencing. Both expression studies screened for transcripts of candidate ELPs were conducted at 1 to 7

`

dpi, significantly after many non-classically secreted proteins would be expected to be secreted from the plant. Additionally, secreted proteins derived from expression data in (Weng et al., 2014) was pre-screened for N-terminal signal peptides with SignalP, which would have excluded most potential non-classically secreted proteins as well. The following described proteins were present in the final predicted SecretePEPPr results but not directly evidenced in the two expression studies cross-referenced above.

An unexpected but intriguing result found in all SecretePEPPr + Motif protein sets across the four *Vitis vinifera* cultivars included in this study was the presence of expansin-related proteins. Targeted to the plant cell wall, expansin-related proteins are normally associated with cell wall loosening during growth (Cosgrove, 2015). Expansin serves to open the network of cell wall polysaccharides, allowing turgor-driven cell enlargement. Cell wall loosening mediated by expansins is generally acid-induced. Surprisingly, transgenic expression of expansin genes in tobacco and *Arabidopsis* leads to changes in growth and more effective resistance to certain biotic stresses, possibly by improving oxidative stress tolerance (Han et al., 2015). Fungal expansins have been associated with cell wall reorganization in *Aspergillus nidulans* (Bouzarelou et al., 2008). Horizontal gene transfer has been described between some pathogenic fungi and their host plants, providing adaptive advantages in infection (Nikolaidis et al., 2014). This phenomenon has not been described in powdery mildews, however, and it must be put into consideration that expansin-like protein inclusion in the final set of candidate ELPs may be aberrant, with expansins instead facilitating plant cell wall loosening in early stages of haustorial formation.

Another class of proteins represented across all four *Vitis vinifera* cultivar candidate ELPs are photoassimilate-reponding 1 (PAR1) proteins. These proteins are similar in feature to pathogenesis-related proteins, are implicated in plant defense through their strong induction by salicylate, and have also been observed in bacterial infection (Herbers et al., 1995). Their precise function in relation to pathogenesis is largely unknown.

`

III.　　A Model for Novel Function Among Effector-like Protein Candidates

Grape proteins must overcome several barriers to localize within haustoria, including an outermost layer composed of thickened plant cell membrane, an internal saccharide-rich matrix, the fungal cell wall, and the fungal cell membrane. The opposite phenomenon of effector translocation to the host is not extensively described in powdery mildew fungi, and several translocation mechanisms including clathrin-mediated endocytosis, transport through a translocon complex, and extracellular vesicle uptake and subsequent membrane fusion have been proposed (Bozkurt & Kamoun, 2020).

Previously, the propensity for plant proteins to localize within fungi in a mutualistic or parasitic relationship has been characterized only in the interaction between *Populus trichocarpa* (black cottonwood) and ectomycorrhizal fungus *Laccaria bicolor* as well as between wheat and fungal pathogen *Zymoseptoria tritici*, with both interactions taking place largely in the apoplastic space (Plett et al., 2017; Zhou et al., 2020). To the author's knowledge, this is the first exploration of proteins originating from plants and localizing to fungal haustoria. Given the inherent crosstalk between pathogen (or symbiont) and host at various boundaries such as the host cytoplasm and apoplast, as well the well understood importance of effector-target and effector-Avr interactions during fungal colonization, this work sets the foundation for a previously unexplored mechanism by which plants may influence haustoria-forming pathogenic fungi. As such, we propose the following model for effector-like protein translocation from the host cytoplasm to the fungal cytoplasm (Figure 19).

`
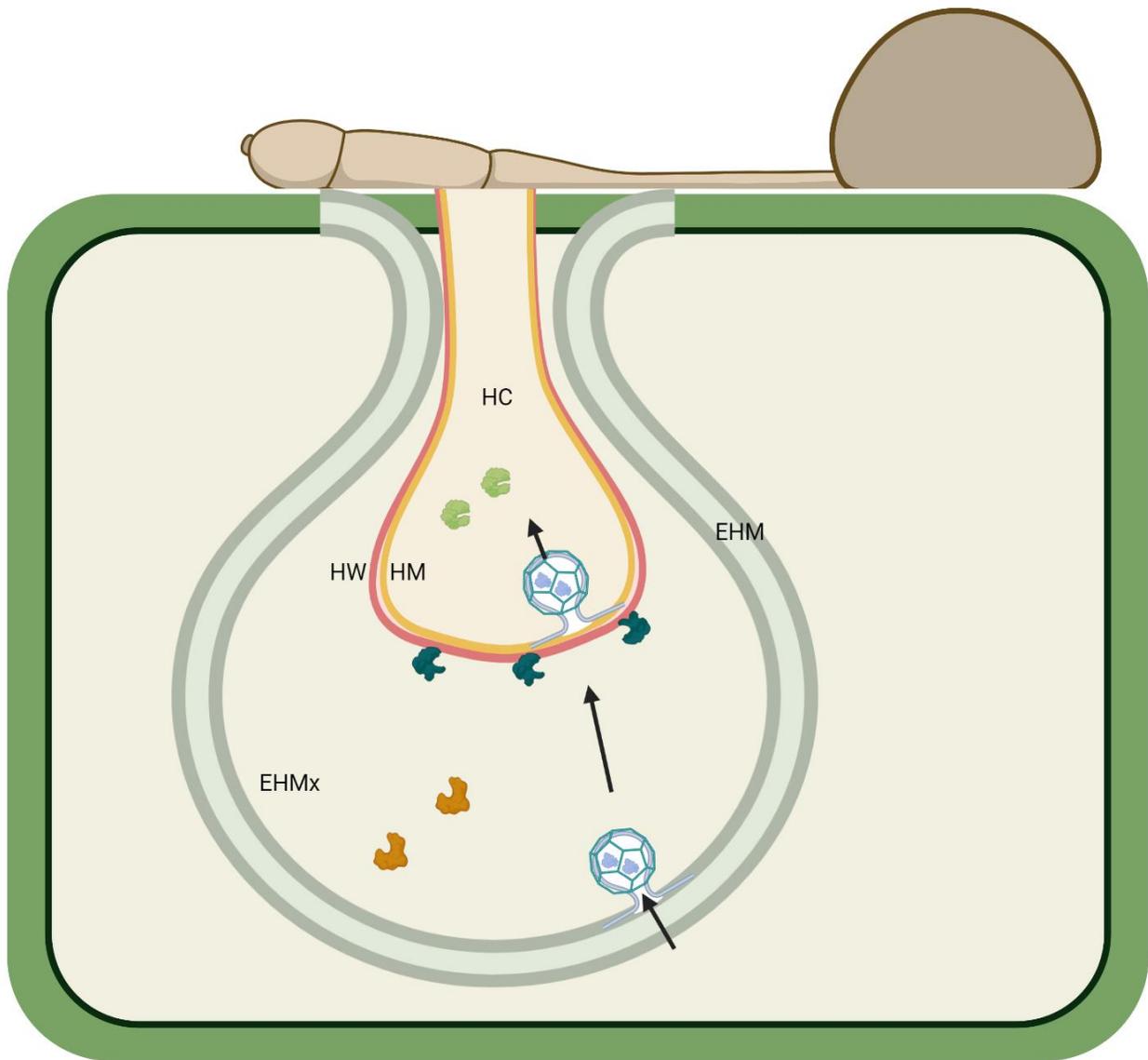
Figure 19. Possible ELP candidate targets within haustoria. Proteins originating in the host cytoplasm are secreted across the plant-derived extrahaustorial membrane (EHM) into the extrahaustorial matrix (EHMx) through vesicles or other secretion methods. Expansin proteins localize to the haustorial wall (HW) and increase permeability. Plant proteins localize to the haustorial membrane (HM) and are shuttled to the haustorial cytoplasm (HC) through clathrin-mediated endocytosis, where they perform functions such as ROS generation and induce metabolic changes in the pathogen.

`

The above model proposes three potential locations that are key for proteins traversing the extrahaustorial membrane: the extrahaustorial matrix, the haustorial cell wall, and the haustorial cytoplasm (Figure 19). Some functions, such as ROS accumulation facilitated by germin-like proteins, could be damaging to the fungus at any of these locations. Others, such as cell wall expansion properties of expansins, are likely to function only within a single site, namely the haustorial cell wall. Since haustorial compartments are complex and precise localizations and functions must be determined using fluorescent tagging or peptide identification-based methods, the above model is limited in its scope to only the proteins identified in this study as potentially participating in clathrin-mediated endocytosis. This is proposed to occur at the plant-derived extrahaustorial membrane, which has been implicated in fungal effector transport in the literature, as well as potentially in the haustorial membrane (Figure 19) (Bozkurt & Kamoun, 2020). Notably, this model is highly simplified and serves to illustrate one possible mechanism for host protein translocation into fungal haustoria.

Fungi employ a highly selective repertoire of plant proteins to the extrahaustorial membrane, and as such it may be unlikely that GPM encodes all the molecular machinery it uses for effector delivery or potential effector-like protein import. Barley powdery mildew caused by *Blumeria graminis* has been shown to aggregate endoplasmic reticulum membrane to the extrahaustorial membrane and likely co-opt *HvSec61βa*, a barley translocon responsible for shuttling peptides between the cytosol and endoplasmic reticulum, for virulence-related activity (Zhang et al., 2013). This supports the notion that while *Erysiphe necator* only has one gene model for clathrin light chain present in its genome annotations, there are likely various plant-derived proteins that can facilitate clathrin-mediated endocytosis at the haustorial interface (Jones et al., 2014). In addition to vesicular trafficking, the encasement of plant-derived extracellular vesicles and multivesicular bodies during the infection process likely has a part to play in protein shuttling between pathogen and host (Micali et al., 2011).

`

IV.   Perspectives on *in vivo* Validation

One immense challenge in studying obligate biotrophs, particularly haustoria-forming fungi, is the inability to cultivate the pathogen outside of its host. This study set out to independently validate effector-like grape proteins localizing within *E. necator* haustoria through haustorial extraction and discovery-based mass spectrometry, relying on peptide mass fingerprinting to identify plant proteins following identification of fungal proteins through a custom-made haustorial proteome database of *E. necator* isolate 'LNYM'. Over the course of several months, a biphasic extraction protocol was adopted and optimized for the extraction of *E. necator*, a protocol that to the author's knowledge did not exist for this pathosystem before this point. The protocol consisted of a homogenization step followed by Percoll density gradient extraction, fluor-labeling, and FACS to generate samples highly enriched in fungal haustoria. In this project we took a multifaceted approach to confirm that our protocol, modified from an original protocol for the extraction of wheat stripe rust haustoria, did indeed produce a sample enriched in *Erysiphe necator* haustorial structures (Garnica & Rathjen, 2014). This approach included fungal particle staining with lactophenol cotton blue and benchtop microscopy, including measurement of particle length with a micrometer, confocal microscopy of WGA-AlexaFluor488-bound haustoria, and FACS sorting of samples indicating distinct chloroplast and haustorial cell populations (Figures 12-17).

The availability of this protocol, along with several optimizations conducted in this study such as determination of the haustorial layer in the Percoll density gradient, sample volume adjustments, and fluor compatibility have implications beyond the scope of this study and its further directions. Pure haustorial samples could be used to investigate myriad pathogenically relevant aspects of complex *E. necator* haustoria that would not be possible in complex with the host, including but not limited to the presence of effector-like proteins of host origin within the haustorial cytoplasm. It is important to note that this protocol strips the two outermost layers of the fungal haustoria during the homogenization step, namely the extrahaustorial membrane and the extrahaustorial matrix, leaving only the fungal cell wall and cell membrane behind. This notably limits further studies using this protocol to excluding conclusions about these two layers of the haustorial boundary without additional modifications to the protocol.

`

A number of additional changes are recommended to subsequent users of the haustorial extraction protocol as a result of factors that negatively impacted our capability in this study to complete *in vitro* validation through mass spectrometry. The major issue was the inability to pellet haustorial cells following FACS. This is believed to be due to a prohibitively small starting material of 20 grams of leaf tissue. It is strongly recommended that subsequent attempts at haustorial extraction use at least 100 grams of leaf tissue to ensure that enough cells are sorted to be pelleted in the final sample. Alternatively, different approaches to pellet the cells, collect them on a filter, or otherwise concentrate them could be attempted prior to protein extraction. In our experiments, we achieved a cell sorting efficiency of 33%, meaning that around a third of the sample run through the sorter consisted of haustoria that were retained in the collection tube. Between 12 and 18 billion cells were sorted in a five-hour time frame with eight microliters of sample following Percoll density gradient extraction, following 1:1 dilution in 1X PBS buffer.

Another major challenge in our implementation of the haustorial extraction protocol was the use of wheat germ agglutinin conjugated to AlexaFluor488. This fluor was found to nonspecifically bind to both haustoria and chloroplast, which complicated downstream sorting attempts (Figure 12). Wheat germ agglutinin binds N-acetyl-D-glucosamine and sialic acid, and one of these species appear to be present on chloroplast surfaces as well as haustorial surfaces. In subsequent haustorial extraction attempts, it is recommended that a different conjugate species is determined for haustorial fluorescent tagging, ensuring that the attached fluor does not overlap with chloroplast autofluorescence. Concanavalin A, which binds to a-mannopyranosyl and a-glucopyranosyl residues, is the lectin used in the original protocol in wheat stripe rust and may be a valuable tool in better separating chloroplasts and haustoria in the *Vitis vinifiera / Erysiphe necator* pathosystem (Garnica & Rathjen, 2014). However, Concanavalin A has been found to bind more strongly to rust haustoria than to powdery mildew haustoria, suggesting that using this lectin instead may trade efficient fluor binding for specificity (Hahn & Mendgen, 1992). Future work following the above changes to the haustorial extraction protocol should focus on fluorescent imaging of protein

`

candidates, further predictive and computational considerations, and characterizing the haustorial proteome.

V.   **Future Directions**

Subsequent work in this domain should be focused on a targeted analysis of the proteins identified in this study for relevant structural and functional characteristics. The global haustorial protein repertoire can be better understood through mass spectrometry approaches. *Vitis* protein localization can be confirmed through fluorescent imaging, and computational predictions can be improved through the consideration of additional endocytic motifs as they are discovered in plants and fungi. Various intermediate confirmatory experiments should also be conducted on candidate ELPs to confirm their secretion and uptake into fungal haustoria.

The value of determining a global proteomic inventory in host-pathogen interactions cannot be understated, and as such it is one of the paramount future directions for this project suggested by the author. Previous whole-proteome analysis of barley powdery mildew haustoria revealed hundreds of proteins and their functions within haustoria, including the missing primary metabolism enzyme pyruvate decarboxylase (Godfrey et al., 2009). The extent to which pathogenesis or metabolism-relevant proteins are missing within *E. necator* haustoria can be best ascertained by a top-down whole-proteome approach. The notion of plant effector-like proteins and plant proteins co-opted by the fungus during pathogenesis give an entirely new perspective on what previously could be discarded as contamination in whole-proteome studies. Furthermore, protein quantification and targeted analysis can reveal the dose-dependent details behind effector-like protein function.

In this study, we provide a shortlist of anticipated grapevine effector-like proteins in the SecretePEPPr + Motif datasets that show sequence-level evidence of uptake into the fungus through clathrin-mediated endocytosis. Visual confirmation of this phenomenon should be conducted through fluorescent labeling and imagery through transgenic or other means. The value of this further experiment is twofold, as it would confirm both the localization of plant effector-like proteins within haustoria and

`

suggestive evidence of their translocation mechanism: be it clathrin-mediated endocytosis, a different vesicle-based transport, classical secretion into the extrahaustorial matrix, or translocon activity across the extrahaustorial membrane. With the added haustorial extraction protocol, this can be conducted either *in situ* or post-extraction to confirm localization within the haustorial cytoplasm and reduce background noise from secondary localization within plant subcellular compartments.(Plett et al., 2017) additionally confirmed the secretion of *Populus trichocarpa* candidate effector-like proteins through a yeast expression system where the survival of yeast was dependent on the secretion of the candidate protein. This study also performed additional fluorescent labeling of specific candidate proteins to confirm nuclear localization. The successful isolation of haustorial cells in this project solves an important obstacle arising from the obligate biotrophic nature of *Erysiphe necator*, namely studying its haustorium without host contamination or interference. That being said, validation of candidate effector-like proteins should be approached from multiple perspectives including confirming their secretion or uptake within clathrin-coated vesicles and amino acid scrambling to ensure the necessity of N-terminal signal peptides in that process, where relevant.

Computational approaches to peptide classification are always lacking some form of information, and as such SecretePEPPr is written in a way that new endocytic motifs can be searched for at any point in the future. If new endosome, multivesicular body, or translocon-based transportation motifs are discovered in the future, they can be incorporated into further analyses. Sequence-based methods can often be limited in their scope and accuracy, and the additional consideration for protein structure and protein-protein interactions could greatly improve the predictive power of the pipeline.

In the future, it is also important for this work to be repeated in a better understood model organism such as maize or *Arabidopsis thaliana* to further confirm if effector-like proteins of plant origin are present in other pathosystems involving plants and haustoria-forming fungi. Host-originating protein targets for the purpose of combating fungal infection are especially useful in agriculture, as we have more control over modulating host biological functions through technologies like transgenics and transient gene expression than we do pathogens, even when applying fungicides frequently.

`

CONCLUSIONS

Erysiphe necator, the causal pathogen of grapevine powdery mildew on *Vitis* spp., is one of many fungal pathogens which harbor a repertoire of effector proteins capable of translocating to various subcellular compartments of its host. Recent evidence from other plant species has suggested a role for host effector-like proteins, which are produced within the plant and translocate to the fungal cytoplasm in mutualistic interactions. In this study, we designed and implemented a multifaceted, systematic computational methodology for predicting candidate effector-like proteins from a whole genome annotation based on predicted secretion through classical and non-classical means, size, and the prevalence and presence of clathrin-mediated endocytic motifs. In this project, a shortlist of four candidate effector-like protein classes is provided across four *Vitis vinifera* cultivars, identifying high confidence targets for follow up investigations with FP-tagging, targeted mass spectrometry, or discovery-based mass spectrometry methods. Finally, a model is conceptualized for the translocation of plant effector-like proteins and their novel functions. Understanding and validating plant effector-like proteins will open an entirely new set of functional characterizations for defense-related secreted plant proteins.

`

REFERENCES

Afzal, A. J., Wood, A. J., & Lightfoot, D. A. (2008). Plant receptor-like serine threonine kinases: Roles in signaling and plant defense. *Molecular Plant-Microbe Interactions: MPMI*, *21*(5), 507–517. https://doi.org/10.1094/MPMI-21-5-0507

Agrawal, G. K., Jwa, N.-S., Lebrun, M.-H., Job, D., & Rakwal, R. (2010). Plant secretome: Unlocking secrets of the secreted proteins. *Proteomics*, *10*(4), 799–827. https://doi.org/10.1002/pmic.200900514

Alexandersson, E., Ashfaq, A., Resjö, S., & Andreasson, E. (2013). Plant secretome proteomics. *Frontiers in Plant Science*, *4*. https://doi.org/10.3389/fpls.2013.00009

Alscher, R. G., Erturk, N., & Heath, L. S. (2002). Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *Journal of Experimental Botany*, *53*(372), 1331–1341. https://doi.org/10.1093/jexbot/53.372.1331

Amselem, J., Vigouroux, M., Oberhaensli, S., Brown, J. K. M., Bindschedler, L. V., Skamnioti, P., Wicker, T., Spanu, P. D., Quesneville, H., & Sacristán, S. (2015). Evolution of the EKA family of powdery mildew avirulence-effector genes from the ORF 1 of a LINE retrotransposon. *BMC Genomics*, *16*(1), 917. https://doi.org/10.1186/s12864-015-2185-x

Barman, A. R., & Banerjee, J. (2015). Versatility of germin-like proteins in their sequences, expressions, and functions. *Functional & Integrative Genomics*, *15*(5), 533–548. https://doi.org/10.1007/s10142-015-0454-z

Barnett, M. (2015). Computational identification of conserved haustorial-expressed genes in the grapevine powdery mildew fungus Erysiphe necator. *Theses*. https://scholarworks.rit.edu/theses/8725

Belchí-Navarro, S., Almagro, L., Sabater-Jara, A. B., Fernández-Pérez, F., Bru, R., & Pedreño, M. A. (2013). Induction of trans-resveratrol and extracellular pathogenesis-related proteins in elicited suspension cultured cells of Vitis vinifera cv Monastrell. *Journal of Plant Physiology*, *170*(3), 258–264. https://doi.org/10.1016/j.jplph.2012.10.003

Beracochea, V. C., Almasia, N. I., Peluffo, L., Nahirñak, V., Hopp, E. H., Paniego, N., Heinz, R. A., Vazquez-Rovere, C., & Lia, V. V. (2015). Sunflower germin-like protein HaGLP1 promotes ROS accumulation and enhances protection against fungal pathogens in transgenic Arabidopsis thaliana. *Plant Cell Reports*, *34*(10), 1717–1733. https://doi.org/10.1007/s00299-015-1819-4

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome: Tair: Making and Mining the "Gold Standard" Plant Genome. *Genesis*, *53*(8), 474–485. https://doi.org/10.1002/dvg.22877

Bouzarelou, D., Billini, M., Roumelioti, K., & Sophianopoulou, V. (2008). EglD, a putative endoglucanase, with an expansin like domain is localized in the conidial cell wall of Aspergillus nidulans. *Fungal Genetics and Biology*, *45*(6), 839–850. https://doi.org/10.1016/j.fgb.2008.03.001

Bozkurt, T. O., Schornack, S., Win, J., Shindo, T., Ilyas, M., Oliva, R., Cano, L. M., Jones, A. M. E., Huitema, E., van der Hoorn, R. A. L., & Kamoun, S. (2011). Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proceedings of the National Academy of Sciences*, *108*(51), 20832–20837. https://doi.org/10.1073/pnas.1112708109

`

Bozkurt, Tolga O., Belhaj, K., Dagdas, Y. F., Chaparro-Garcia, A., Wu, C.-H., Cano, L. M., & Kamoun, S. (2015). Rerouting of plant late endocytic trafficking toward a pathogen interface. *Traffic (Copenhagen, Denmark)*, *16*(2), 204–226. https://doi.org/10.1111/tra.12245

Bozkurt, Tolga O., & Kamoun, S. (2020). The plant–pathogen haustorial interface at a glance. *Journal of Cell Science*, *133*(5). https://doi.org/10.1242/jcs.237958

Brewer, M. T., & Milgroom, M. G. (2010). Phylogeography and population structure of the grape powdery mildew fungus, Erysiphe necator, from diverse Vitis species. *BMC Evolutionary Biology*, *10*, 268. https://doi.org/10.1186/1471-2148-10-268

Cadle-Davidson, L.E. (2019). A perspective on breeding and implementing durable powdery mildew resistance. Acta Hortic. 1248, 541-548 https://doi.org/10.17660/ActaHortic.2019.1248.72

Chaudhari, P., Ahmed, B., Joly, D. L., & Germain, H. (2014). Effector biology during biotrophic invasion of plant cells. *Virulence*, *5*(7), 703–709. https://doi.org/10.4161/viru.29652

Cheng, F., Blackburn, K., Lin, Y., Goshe, M. B., & Williamson, J. D. (2009). Absolute Protein Quantification by LC/MSE for Global Analysis of Salicylic Acid-Induced Plant Protein Secretion Responses. *Journal of Proteome Research*, *8*(1), 82–93. https://doi.org/10.1021/pr800649s

Christensen, A. B., Thordal-Christensen, H., Zimmermann, G., Gjetting, T., Lyngkjær, M. F., Dudler, R., & Schweizer, P. (2004). The Germinlike Protein GLP4 Exhibits Superoxide Dismutase Activity and Is an Important Component of Quantitative Resistance in Wheat and Barley. *Molecular Plant-Microbe Interactions®*, *17*(1), 109–117. https://doi.org/10.1094/MPMI.2004.17.1.109

Coleman, C., Copetti, D., Cipriani, G., Hoffmann, S., Kozma, P., Kovács, L., Morgante, M., Testolin, R., & Di Gaspero, G. (2009). The powdery mildew resistance gene REN1 co-segregates with an NBS-LRR gene cluster in two Central Asian grapevines. *BMC Genetics*, *10*, 89. https://doi.org/10.1186/1471-2156-10-89

Cosgrove, D. J. (2015). Plant expansins: Diversity and interactions with plant cell walls. *Current Opinion in Plant Biology*, *25*, 162–172. https://doi.org/10.1016/j.pbi.2015.05.014

Delaunois, B., Colby, T., Belloy, N., Conreux, A., Harzen, A., Baillieul, F., Clément, C., Schmidt, J., Jeandet, P., & Cordelier, S. (2013). Large-scale proteomic analysis of the grapevine leaf apoplastic fluid reveals mainly stress-related proteins and cell wall modifying enzymes. *BMC Plant Biology*, *13*(1), 24. https://doi.org/10.1186/1471-2229-13-24

Délye, C., Laigret, F., & Corio-Costet, M. F. (1997). A mutation in the 14 alpha-demethylase gene of Uncinula necator that correlates with resistance to a sterol biosynthesis inhibitor. *Applied and Environmental Microbiology*, *63*(8), 2966–2970.

Feechan, A., Kocsis, M., Riaz, S., Zhang, W., Gadoury, D. M., Walker, M. A., Dry, I. B., Reisch, B., & Cadle-Davidson, L. (2015). Strategies for RUN1 Deployment Using RUN2 and REN2 to Manage Grapevine Powdery Mildew Informed by Studies of Race Specificity. *Phytopathology®*, *105*(8), 1104–1113. https://doi.org/10.1094/PHYTO-09-14-0244-R

Ficke, A., Gadoury, D. M., & Seem, R. C. (2002). Ontogenic resistance and plant disease management: A case study of grape powdery mildew. *Phytopathology*, *92*(6), 671–675. https://doi.org/10.1094/PHYTO.2002.92.6.671

Ficke, A., Gadoury, D. M., Seem, R. C., Godfrey, D., & Dry, I. B. (2004). Host Barriers and Responses to *Uncinula necator* in Developing Grape Berries. *Phytopathology®*, *94*(5), 438–445. https://doi.org/10.1094/PHYTO.2004.94.5.438

`

Floerl, S., Majcherczyk, A., Possienke, M., Feussner, K., Tappe, H., Gatz, C., Feussner, I., Kües, U., & Polle, A. (2012). Verticillium longisporum infection affects the leaf apoplastic proteome, metabolome, and cell wall properties in Arabidopsis thaliana. *PloS One*, *7*(2), e31435. https://doi.org/10.1371/journal.pone.0031435

Fuller, K. B., Alston, J. M., & Sambucci, O. S. (2014). The value of powdery mildew resistance in grapes: Evidence from California. *Wine Economics and Policy*, *3*(2), 90–107. https://doi.org/10.1016/j.wep.2014.09.001

Gadoury, D. M., Cadle-Davidson, L., Wilcox, W. F., Dry, I. B., Seem, R. C., & Milgroom, M. G. (2012). Grapevine powdery mildew (Erysiphe necator): A fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. *Molecular Plant Pathology*, *13*(1), 1–16. https://doi.org/10.1111/j.1364-3703.2011.00728.x

Garnica, D. P., & Rathjen, J. P. (2014). Purification of fungal haustoria from infected plant tissue by flow cytometry. *Methods in Molecular Biology (Clifton, N.J.)*, *1127*, 103–110. https://doi.org/10.1007/978-1-62703-986-4_8

Godfrey, D., Böhlenius, H., Pedersen, C., Zhang, Z., Emmersen, J., & Thordal-Christensen, H. (2010). Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *BMC Genomics*, *11*(1), 317. https://doi.org/10.1186/1471-2164-11-317

Godfrey, D., Zhang, Z., Saalbach, G., & Thordal-Christensen, H. (2009). A proteomics study of barley powdery mildew haustoria. *PROTEOMICS*, *9*(12), 3222–3232. https://doi.org/10.1002/pmic.200800645

Gupta, R., Lee, S. E., Agrawal, G. K., Rakwal, R., Park, S., Wang, Y., & Kim, S. T. (2015). Understanding the plant-pathogen interactions in the context of proteomics-generated apoplastic proteins inventory. *Frontiers in Plant Science*, *6*, 352. https://doi.org/10.3389/fpls.2015.00352

Hahn, M., & Mendgen, K. (1992). Isolation by ConA binding of haustoria from different rust fungi and comparison of their surface qualities. *Protoplasma*, *170*(3–4), 95–103. https://doi.org/10.1007/BF01378785

Han, Y., Chen, Y., Yin, S., Zhang, M., & Wang, W. (2015). Over-expression of TaEXPB23, a wheat expansin gene, improves oxidative stress tolerance in transgenic tobacco plants. *Journal of Plant Physiology*, *173*, 62–71. https://doi.org/10.1016/j.jplph.2014.09.007

Herbers, K., Mönke, G., Badur, R., & Sonnewald, U. (1995). A simplified procedure for the subtractive cDNA cloning of photoassimilate-responding genes: Isolation of cDNAs encoding a new class of pathogenesis-related proteins. *Plant Molecular Biology*, *29*(5), 1027–1038. https://doi.org/10.1007/BF00014975

Himmelbach, A., Liu, L., Zierold, U., Altschmied, L., Maucher, H., Beier, F., Müller, D., Hensel, G., Heise, A., Schützendübel, A., Kumlehn, J., & Schweizer, P. (2010). Promoters of the Barley Germin-Like *GER4* Gene Cluster Enable Strong Transgene Expression in Response to Pathogen Attack. *The Plant Cell*, *22*(3), 937–952. https://doi.org/10.1105/tpc.109.067934

Hoffmann, S., Di Gaspero, G., Kovács, L., Howard, S., Kiss, E., Galbács, Z., Testolin, R., & Kozma, P. (2008). Resistance to Erysiphe necator in the grapevine 'Kishmish vatkana' is controlled by a single locus through restriction of hyphal growth. *Theoretical and Applied Genetics*, *116*(3), 427–438. https://doi.org/10.1007/s00122-007-0680-4

Hooper, C. M., Castleden, I. R., Aryamanesh, N., Jacoby, R. P., & Millar, A. H. (2016). Finding the Subcellular Location of Barley, Wheat, Rice and Maize Proteins: The Compendium of Crop

`

Proteins with Annotated Locations (cropPAL). *Plant & Cell Physiology*, *57*(1), e9. https://doi.org/10.1093/pcp/pcv170

Hooper, C. M., Castleden, I. R., Tanz, S. K., Aryamanesh, N., & Millar, A. H. (2017). SUBA4: The interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Research*, *45*(Database issue), D1064–D1074. https://doi.org/10.1093/nar/gkw1041

Ilyas, M., Rasheed, A., & Mahmood, T. (2016). Functional characterization of germin and germin-like protein genes in various plant species using transgenic approaches. *Biotechnology Letters*, *38*(9), 1405–1421. https://doi.org/10.1007/s10529-016-2129-9

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., … The French–Italian Public Consortium for Grapevine Genome Characterization. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, *449*(7161), 463–467. https://doi.org/10.1038/nature06148

Jayasankar, S., Li, Z., & Gray, D. J. (2003). Constitutive expression of Vitis vinifera thaumatin-like protein after in vitro selection and its role in anthracnose resistance. *Functional Plant Biology*, *30*(11), 1105–1115. https://doi.org/10.1071/fp03066

Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K. C., McGuire, B., Gubler, W. D., Walker, M. A., & Cantu, D. (2014). Adaptive genomic structural variation in the grape powdery mildew pathogen, Erysiphe necator. *BMC Genomics*, *15*(1), 1081. https://doi.org/10.1186/1471-2164-15-1081

Kaffarnik, F. A. R., Jones, A. M. E., Rathjen, J. P., & Peck, S. C. (2009). Effector proteins of the bacterial pathogen Pseudomonas syringae alter the extracellular proteome of the host plant, Arabidopsis thaliana. *Molecular & Cellular Proteomics: MCP*, *8*(1), 145–156. https://doi.org/10.1074/mcp.M800043-MCP200

Kim, S. G., Wang, Y., Lee, K. H., Park, Z.-Y., Park, J., Wu, J., Kwon, S. J., Lee, Y.-H., Agrawal, G. K., Rakwal, R., Kim, S. T., & Kang, K. Y. (2013). In-depth insight into in vivo apoplastic secretome of rice-Magnaporthe oryzae interaction. *Journal of Proteomics*, *78*, 58–71. https://doi.org/10.1016/j.jprot.2012.10.029

Knecht, K., Seyffarth, M., Desel, C., Thurau, T., Sherameti, I., Lou, B., Oelmüller, R., & Cai, D. (2010). Expression of BvGLP-1 encoding a germin-like protein from sugar beet in Arabidopsis thaliana leads to resistance against phytopathogenic fungi. *Molecular Plant-Microbe Interactions: MPMI*, *23*(4), 446–457. https://doi.org/10.1094/MPMI-23-4-0446

Kozik, P., Francis, R. W., Seaman, M. N. J., & Robinson, M. S. (2010). A screen for endocytic motifs. *Traffic (Copenhagen, Denmark)*, *11*(6), 843–855. https://doi.org/10.1111/j.1600-0854.2010.01056.x

Kumar, M., Brar, A., Yadav, M., Chawade, A., Vivekanand, V., & Pareek, N. (2018). Chitinases—Potential Candidates for Enhanced Plant Resistance towards Fungal Pathogens. *Agriculture*, *8*(7), 88. https://doi.org/10.3390/agriculture8070088

Langner, T., Kamoun, S., & Belhaj, K. (2018). CRISPR Crops: Plant Genome Editing Toward Disease Resistance. *Annual Review of Phytopathology*, *56*(1), 479–512. https://doi.org/10.1146/annurev-phyto-080417-050158

Leborgne-Castel, N., Adam, T., & Bouhidel, K. (2010). Endocytosis in plant-microbe interactions. *Protoplasma*, *247*(3–4), 177–193. https://doi.org/10.1007/s00709-010-0195-8

`

Liu, H., Hu, M., Wang, Q., Cheng, L., & Zhang, Z. (2018). Role of Papain-Like Cysteine Proteases in Plant Development. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.01717

Liu, J.-J., Sturrock, R., & Ekramoddoullah, A. K. M. (2010). The superfamily of thaumatin-like proteins: Its origin, evolution, and expression towards biological function. *Plant Cell Reports*, *29*(5), 419–436. https://doi.org/10.1007/s00299-010-0826-8

Lo Presti, L., & Kahmann, R. (2017). How filamentous plant pathogen effectors are translocated to host cells. *Current Opinion in Plant Biology*, *38*, 19–24. https://doi.org/10.1016/j.pbi.2017.04.005

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., & Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. *Annual Review of Plant Biology*, *66*(1), 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Lucas, J. A., Hawkins, N. J., & Fraaije, B. A. (2015). Chapter Two—The Evolution of Fungicide Resistance. In S. Sariaslani & G. M. Gadd (Eds.), *Advances in Applied Microbiology* (Vol. 90, pp. 29–92). Academic Press. https://doi.org/10.1016/bs.aambs.2014.09.001

Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–W641. https://doi.org/10.1093/nar/gkz268

Matsubayashi, Y., Ogawa, M., Morita, A., & Sakagami, Y. (2002). An LRR Receptor Kinase Involved in Perception of a Peptide Plant Hormone, Phytosulfokine. *Science*, *296*(5572), 1470–1472. https://doi.org/10.1126/science.1069607

McHale, L., Tan, X., Koehl, P., & Michelmore, R. W. (2006). Plant NBS-LRR proteins: Adaptable guards. *Genome Biology*, *7*(4), 212. https://doi.org/10.1186/gb-2006-7-4-212

Mendgen, K., Bachem, U., Stark-Urnau, M., & Xu, H. (2011). Secretion and endocytosis at the interface of plants and fungi. *Canadian Journal of Botany*. https://doi.org/10.1139/b95-306

Micali, C. O., Neumann, U., Grunewald, D., Panstruga, R., & O'Connell, R. (2011). Biogenesis of a specialized plant-fungal interface during host cell internalization of Golovinomyces orontii haustoria. *Cellular Microbiology*, *13*(2), 210–226. https://doi.org/10.1111/j.1462-5822.2010.01530.x

Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A., & Cantu, D. (2019). Diploid Genome Assembly of the Wine Grape Carménère. *G3: Genes, Genomes, Genetics*, *9*(5), 1331–1337. https://doi.org/10.1534/g3.119.400030

Minio, A., Massonnet, M., Figueroa-Balderas, R., Vondras, A. M., Blanco-Ulate, B., & Cantu, D. (2019). Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry Development. *G3: Genes, Genomes, Genetics*, *9*(3), 755–767. https://doi.org/10.1534/g3.118.201008

Monteiro, S., Barakat, M., Piçarra-Pereira, M. A., Teixeira, A. R., & Ferreira, R. B. (2003). Osmotin and Thaumatin from Grape: A Putative General Defense Mechanism Against Pathogenic Fungi. *Phytopathology®*, *93*(12), 1505–1512. https://doi.org/10.1094/PHYTO.2003.93.12.1505

Morgan, W., & Kamoun, S. (2007). RXLR effectors of plant pathogenic oomycetes. *Current Opinion in Microbiology*, *10*(4), 332–338. https://doi.org/10.1016/j.mib.2007.04.005

`

Nickel, W., & Seedorf, M. (2008). Unconventional Mechanisms of Protein Transport to the Cell Surface of Eukaryotic Cells. *Annual Review of Cell and Developmental Biology*, *24*(1), 287–308. https://doi.org/10.1146/annurev.cellbio.24.110707.175320

Nikolaidis, N., Doran, N., & Cosgrove, D. J. (2014). Plant Expansins in Bacteria and Fungi: Evolution by Horizontal Gene Transfer and Independent Domain Fusion. *Molecular Biology and Evolution*, *31*(2), 376–386. https://doi.org/10.1093/molbev/mst206

Patnaik, D., & Khurana, P. (2001). Germins and germin like proteins: An overview. *Indian Journal of Experimental Biology*, *39*(3), 191–200.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Plett, J. M., Yin, H., Mewalal, R., Hu, R., Li, T., Ranjan, P., Jawdy, S., De Paoli, H. C., Butler, G., Burch-Smith, T. M., Guo, H.-B., Ju Chen, C., Kohler, A., Anderson, I. C., Labbé, J. L., Martin, F., Tuskan, G. A., & Yang, X. (2017). Populus trichocarpa encodes small, effector-like secreted proteins that are highly induced during mutualistic symbiosis. *Scientific Reports*, *7*(1), 382. https://doi.org/10.1038/s41598-017-00400-8

Qiu, W., Feechan, A., & Dry, I. (2015). Current understanding of grapevine defense mechanisms against the biotrophic fungus (Erysiphe necator), the causal agent of powdery mildew disease. *Horticulture Research*, *2*(1), 15020. https://doi.org/10.1038/hortres.2015.20

Rafiqi, M., Ellis, J. G., Ludowici, V. A., Hardham, A. R., & Dodds, P. N. (2012). Challenges and progress towards understanding the role of effectors in plant-fungal interactions. *Current Opinion in Plant Biology*, *15*(4), 477–482. https://doi.org/10.1016/j.pbi.2012.05.003

Rafiqi, M., Gan, P. H. P., Ravensdale, M., Lawrence, G. J., Ellis, J. G., Jones, D. A., Hardham, A. R., & Dodds, P. N. (2010). Internalization of Flax Rust Avirulence Proteins into Flax and Tobacco Cells Can Occur in the Absence of the Pathogen. *The Plant Cell*, *22*(6), 2017–2032. https://doi.org/10.1105/tpc.109.072983

Roach, M. J., Johnson, D. L., Bohlmann, J., Vuuren, H. J. J. van, Jones, S. J. M., Pretorius, I. S., Schmidt, S. A., & Borneman, A. R. (2018). Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLOS Genetics*, *14*(11), e1007807. https://doi.org/10.1371/journal.pgen.1007807

Sambucci, O., Alston, J. M., Fuller, K. B., & Lusk, J. (2019). The Pecuniary and Nonpecuniary Costs of Powdery Mildew and the Potential Value of Resistant Grape Varieties in California. *American Journal of Enology and Viticulture*, *70*(2), 177–187. https://doi.org/10.5344/ajev.2018.18032

Saucet, S. B., & Shirasu, K. (2016). Molecular Parasitic Plant–Host Interactions. *PLoS Pathogens*, *12*(12). https://doi.org/10.1371/journal.ppat.1005978

Sauter, M. (2015). Phytosulfokine peptide signalling. *Journal of Experimental Botany*, *66*(17), 5161–5169. https://doi.org/10.1093/jxb/erv071

Savojardo, C., Martelli, P. L., Fariselli, P., & Casadio, R. (2018). DeepSig: Deep learning improves signal peptide detection in proteins. *Bioinformatics (Oxford, England)*, *34*(10), 1690–1696. https://doi.org/10.1093/bioinformatics/btx818

Sels, J., Mathys, J., De Coninck, B. M. A., Cammue, B. P. A., & De Bolle, M. F. C. (2008). Plant pathogenesis-related (PR) proteins: A focus on PR peptides. *Plant Physiology and Biochemistry: PPB*, *46*(11), 941–950. https://doi.org/10.1016/j.plaphy.2008.06.011

`

Sperschneider, J., Catanzariti, A.-M., DeBoer, K., Petre, B., Gardiner, D. M., Singh, K. B., Dodds, P. N., & Taylor, J. M. (2017). LOCALIZER: Subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports*, *7*(1), 44598. https://doi.org/10.1038/srep44598

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology*, *19*(9), 2094–2110. https://doi.org/10.1111/mpp.12682

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., & Taylor, J. M. (2016). EffectorP: Predicting fungal effector proteins from secretomes using machine learning. *The New Phytologist*, *210*(2), 743–761. https://doi.org/10.1111/nph.13794

Stergiopoulos, I., & de Wit, P. J. G. M. (2009). Fungal Effector Proteins. *Annual Review of Phytopathology*, *47*(1), 233–263. https://doi.org/10.1146/annurev.phyto.112408.132637

Su, C., Liu, H., Wafula, E. K., Honaas, L., Pamphilis, C. W. de, & Timko, M. P. (2020). SHR4z, a novel decoy effector from the haustorium of the parasitic weed Striga gesnerioides, suppresses host plant immunity. *New Phytologist*, *226*(3), 891–908. https://doi.org/10.1111/nph.16351

Sun, Y., Qiao, Z., Muchero, W., & Chen, J.-G. (2020). Lectin Receptor-Like Kinases: The Sensor and Mediator at the Plant Cell Surface. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.596301

Szabo, L. J., & Bushnell, W. R. (2001). Hidden robbers: The role of fungal haustoria in parasitism of plants. *Proceedings of the National Academy of Sciences*, *98*(14), 7654–7655. https://doi.org/10.1073/pnas.151262398

Tan, P. K., Howard, J. P., & Payne, G. S. (1996). The sequence NPFXD defines a new class of endocytosis signal in Saccharomyces cerevisiae. *The Journal of Cell Biology*, *135*(6), 1789–1800.

Tsirigos, K. D., Peters, C., Shu, N., Käll, L., & Elofsson, A. (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Research*, *43*(W1), W401-407. https://doi.org/10.1093/nar/gkv485

Van loon, L. C., & Van strien, E. A. (1999). The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiological and Molecular Plant Pathology*, *55*(2), 85–97. https://doi.org/10.1006/pmpp.1999.0213

Waters, M. G., Serafini, T., & Rothman, J. E. (1991). "Coatomer": A cytosolic protein complex containing subunits of non-clathrin-coated Golgi transport vesicles. *Nature*, *349*(6306), 248–251. https://doi.org/10.1038/349248a0

Weldon, W. A., Palumbo, C. D., Kovaleski, A. P., Tancos, K., Gadoury, D. M., Osier, M. V., & Cadle-Davidson, L. (2019). Transcriptomic Profiling of Acute Cold Stress-Induced Disease Resistance (SIDR) Genes and Pathways in the Grapevine Powdery Mildew Pathosystem. *Molecular Plant-Microbe Interactions®*, *33*(2), 284–295. https://doi.org/10.1094/MPMI-07-19-0183-R

Weng, K., Li, Z.-Q., Liu, R.-Q., Wang, L., Wang, Y.-J., & Xu, Y. (2014). Transcriptome of Erysiphe necator -infected Vitis pseudoreticulata leaves provides insight into grapevine resistance to powdery mildew. *Horticulture Research*, *1*(1), 1–12. https://doi.org/10.1038/hortres.2014.49

Yan, X., Qiao, H., Zhang, X., Guo, C., Wang, M., Wang, Y., & Wang, X. (2017). Analysis of the grape (Vitis vinifera L.) thaumatin-like protein (TLP) gene family and demonstration that TLP29

`

contributes to disease resistance. *Scientific Reports, 7*(1), 4269. https://doi.org/10.1038/s41598-017-04105-w

Yang, Y., Yang, X., Dong, Y., & Qiu, D. (2018). The Botrytis cinerea Xylanase BcXyl1 Modulates Plant Immunity. *Frontiers in Microbiology, 9*. https://doi.org/10.3389/fmicb.2018.02535

Yin, K., & Qiu, J.-L. (2019). Genome editing for plant disease resistance: Applications and perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences, 374*(1767), 20180322. https://doi.org/10.1098/rstb.2018.0322

Zeilinger, S., Gupta, V. K., Dahms, T. E. S., Silva, R. N., Singh, H. B., Upadhyay, R. S., Gomes, E. V., Tsui, C. K.-M., & Nayak S, C. (2016). Friends or foes? Emerging insights from fungal interactions with plants. *FEMS Microbiology Reviews, 40*(2), 182–207. https://doi.org/10.1093/femsre/fuv045

Zhang, W.-J., Hanisch, S., Kwaaitaal, M., Pedersen, C., & Thordal-Christensen, H. (2013). A component of the Sec61 ER protein transporting pore is required for plant susceptibility to powdery mildew. *Frontiers in Plant Science, 4*. https://doi.org/10.3389/fpls.2013.00127

Zhou, B., Benbow, H. R., Brennan, C. J., Arunachalam, C., Karki, S. J., Mullins, E., Feechan, A., Burke, J. I., & Doohan, F. M. (2020). Wheat Encodes Small, Secreted Proteins That Contribute to Resistance to Septoria Tritici Blotch. *Frontiers in Genetics, 11*. https://doi.org/10.3389/fgene.2020.00469

`