EffectorO: motif-independent prediction of effectors in oomycete genomes using machine learning and lineage specificity

Munir Nur^{1#}, Kelsey Wood^{1,2#}, and Richard Michelmore^{1,3*}

¹The Genome Center, University of California, Davis, USA

²Integrative Genetics & Genomics Graduate Group, University of California, Davis, USA

³Departments of Plant Sciences, Molecular & Cellular Biology, Medical Microbiology & Immunology, University of California, Davis, USA.

Co-first authors

* Corresponding author e-mail: rwmichelmore@ucdavis.edu

Abstract

Oomycete plant pathogens cause a wide variety of diseases, including late blight of potato, sudden oak death, and downy mildew of many plants. These pathogens are major contributors to losses in many food crops. Oomycetes secrete “effector” proteins to manipulate their hosts to the advantage of the pathogen. Plants have evolved to recognize effectors, resulting in an evolutionary cycle of defense and counter-defense in plant-microbe interactions. This selective pressure results in highly diverse effector sequences that can be difficult to computationally identify using sequence similarity. We developed a pipeline, EffectorO, that uses two complementary approaches to predict effectors in oomycete pathogen genomes: (1) a machine learning-based pipeline that predicts effector probability based on the biochemical properties of the N-terminal amino acid sequence of a protein and

is trained on experimentally verified oomycete effectors and (2) a pipeline based on lineage-specificity to find proteins that are unique to one species or genus, a sign of evolutionary divergence due to adaptation to the host. We tested EffectorO on *Bremia lactucae*, which causes lettuce downy mildew, and *Phytophthora infestans*, which causes late blight of potato and tomato, and predicted many novel effector candidates, while still recovering the majority of known effector candidates. EffectorO will be useful for discovering novel families of oomycete effectors without relying on sequence similarity to known effectors.

Introduction

Oomycetes are filamentous eukaryotic organisms that resemble fungi but are more closely related to brown algae (Baldauf 2003). Pathogenic oomycetes cause huge losses in agriculture and aquaculture worldwide (Derevnina et al. 2016; Haverkort et al. 2008). Oomycete pathogens and other parasites must remain undetected by their hosts in order to successfully infect and complete their life cycle. This is accomplished through the action of effectors, secreted proteins that interfere with host immune signaling and physiology to promote pathogen growth (Wang et al. 2019). To thwart the action of effectors and infection by the pathogen, plants have evolved receptors encoded by resistance genes (R genes) to detect effectors and trigger a defense response resulting in disease resistance (Wang et al. 2019). Therefore, identification of effectors is important for both understanding how pathogens infect their hosts and for understanding and breeding for disease resistance.

For some oomycetes, such as *Phytophthora* species and downy mildew pathogens, the discovery of conserved amino acid motifs and protein domains in effector proteins has assisted in the prediction of effectors from genomes. One such motif is the RXLR motif, which

is located in the N-terminus of the protein just following the secretion-related signal peptide (Rehmany et al. 2005). The RXLR motif is often associated with a downstream EER motif, and both motifs have been implicated in secretion and/or translocation of effectors into host cells (Birch et al. 2009; Whisson et al. 2007). Although their role in effector movement remains unresolved (Ellis and Dodds 2011), the predictive value of the RXLR and EER motifs has been demonstrated by the discovery of numerous effectors predicted in pathogen genomes using the motif and validated by testing the candidate proteins for recognition by resistant genotypes (avirulence activity) or immune suppression (virulence activity) *in planta* (Deb et al. 2018; Fabro et al. 2011; Pecrix et al. 2019; Pel et al. 2009; Stassen et al. 2013; Zheng et al. 2014). While the RXLR motif is clearly a conserved effector motif in *Phytophthora* species, effectors in the related downy mildew pathogens show divergence and diversification of this motif (Bailey et al. 2011; Sharma et al. 2015).

Due to the sequence divergence of the RXLR motif in downy mildew species, both string- and Hidden Markov Model (HMM)-based searches recover fewer effector candidates in these pathogens compared to *Phytophthora* spp. (Baxter et al. 2010; McGowan and Fitzpatrick 2017). Another approach to predict effectors in downy mildew species (as well as in *Phytophthora*) is to search for the structurally conserved WY domain, which is present in ~25–50% of RXLR effectors in *Phytophthora* and downy mildew pathogens (Bos et al. 2010; Dou et al. 2008; He et al. 2019; King et al. 2014; Win et al. 2012) and has been implicated in effector function (Bos et al. 2010; Dou et al. 2008; He et al. 2019; King et al. 2014). While HMM searches for the WY domain have a low false positive rate for predicting candidate effectors (Wood et al. 2020), they likely have a high false negative rate, as not all RXLR effectors have WY domains and there may be *bona fide* effectors with novel N-terminal

motifs and uncharacterized effector domains. Unlike in oomycetes, there are no widely conserved effector motifs across all fungi. Degenerate RXLR-like motifs have been found in some fungal pathogens (Gu et al. 2011; Kale et al. 2010) and others have an N-terminal Y/F/WxC motif (Godfrey et al. 2010).

Recently, EffectorP, a machine learning based approach to identify effectors in fungal genomes using biochemical characteristics, successfully expanded the pool of candidate effectors in fungal species (Sperschneider et al. 2016, 2018). Machine learning-based effector prediction pipelines, such as EffectorP, work by building effector classification models trained on known effectors and secreted non-effectors from the target organisms. Once a suitable model is determined, the classifier calculates the biochemical features from the amino acid sequence of the protein and outputs a probability that the protein is an effector based on those features. To date, machine learning has not been applied to effector prediction in oomycetes.

Due to the constant evolutionary cycle between pathogen and host, pathogenicity-related proteins, such as effectors (as well as their corresponding R genes), often show increased rates of sequence divergence (Dong et al. 2015). This is especially true for organisms that have narrow host ranges, such as downy mildew pathogens. For example, over 6,000 genes in the downy mildew pathogen *Hyaloperonospora arabidopsidis* had no identified ortholog in *Phytophthora* species (Baxter et al. 2010) and over 10,000 genes were lineage-specific in the animal pathogen *Saprolegnia parasitica* (Jiang et al. 2013). This lineage specificity can be accounted for in part by the lack of genomic resources in closely related organisms, but for many of these proteins, it is likely due to co-adaptation to specific hosts of these pathogens. We reasoned that by using lineage specificity in combination with a

machine learning-based effector classifier, we would be able to expand the candidate effector repertoires of oomycete pathogens, especially those with narrow host ranges.

Here we present EffectorO: a pipeline to predict novel effectors in oomycete genomes that avoids the limitations of motif/domain-based searches. EffectorO consists of a machine learning based effector classifier trained on a diverse set of oomycete effectors (EffectorO-ML) combined with a lineage-specific protein sequence identification pipeline (EffectorO-LSP). We used two economically important oomycete pathogens as model organisms: *Bremia lactucae*, an obligate biotrophic downy mildew pathogen with complete host-specificity to *Lactuca* spp., and *Phytophthora infestans*, which causes late blight of potato and tomato. EffectorO had high cross-validation accuracy for classifying protein sequences as effectors or non-effectors and predicted thousands of novel effectors in these two oomycete genomes, while recovering known effectors. This analysis was expanded to 27 additional oomycete genomes. One EffectorO prediction was validated by demonstrating the elicitation of cell death on downy mildew resistant genotypes by an effector of *B. lactucae* with no obvious canonical sequence domains. Thus, a machine learning approach, especially when combined with lineage-specificity analysis, can be useful in gene annotation and effector prediction in oomycete genomes.

Materials and Methods

Training sets: Positive (effectors) and negative (non-effectors)

Our positive training set consisted of oomycete effectors experimentally shown to be either recognized by R genes (avirulence activity) or to have an immune-suppressing effect on the host (virulence activity) (Supplemental Table 1). Seventy-seven of these sequences

were obtained from the Pathogen-Host Interaction database (PHI-base) (Urban et al. 2019) and 15 effector candidates were added from *B. lactucae* that had avirulence (Giesbers et al. 2017; Pelgrom et al. 2019; Stassen et al. 2013; Wood et al. 2020) or virulence activity (Wood et al. 2020) resulting in an initial set of 92 sequences. We then used CD-Hit at a sequence identity threshold of 70% and word length of five amino acids in order to identify and remove closely related sequences (Fu et al. 2012; Li and Godzik 2006), resulting in a final positive training set of 88 unique oomycete effectors.

Our negative training set consisted of representative oomycete sequences that were predicted to be secreted and highly conserved throughout the genomes of 28 oomycete species (Supplemental Table 2). To find these sequences, we first performed an Orthofinder search on the oomycete genomes at an e-value cutoff of 1e-10 and a query coverage cutoff of 90% (Emms and Kelly 2015, 2019). We then retrieved ortholog centroids, with the species chosen at random, that were present in at least 26 genomes. We predicted signal peptides using SignalP v4.1 (sensitive mode) (Petersen et al. 2011). To further reduce potential model bias and redundancy in the set of non-effectors, we ran CD-hit and removed sequences with greater than 40% identity. This identified 320 sequences, from which a training set of 88 sequences was chosen at random to obtain a balanced 1:1 positive to negative training set ratio.

Feature selection and encoding for classical machine learning models

To calculate biochemical characteristics of effectors and non-effectors, we used published amino acid scales from ExPASyProtScale (ExPASy; Expert Protein Analysis System 2004) and evaluated six features that could be informative for effector classification. We

focused on the first 100 amino acids of the N-terminus of protein sequences, with the rationale that (1) secretion and translocation signals are typically encoded in the N-terminus and (2) effectors may differ in their C-terminal domains depending on their specific functions (for example, not all validated oomycete effectors have WY domains). To ensure the feature scales were equivalent to each other, we transformed the calculated feature distributions to zero mean and unit variance (Z score). We calculated the average values for six N-terminal features: the grand average of hydropathy (GRAVY), hydrophobicity, surface exposure, disorder propensity, bulkiness, and residue interface propensity of the first 100 amino acids using data from ExPASyProtScale (ExPaSy; Expert Protein Analysis System 2004).

PONDR VSL2 (Peng et al. 2006) was used as an additional method to visualize levels of intrinsic disorder for effectors and secreted non-effectors across the N-terminal sequence. The average sequence disorder was calculated at each amino acid position for the first 160 amino acids including the signal peptide in order to visualize the differences between candidate effectors and non-effectors.

Feature encoding for Convolutional Neural Network models

In order to account for signals that may be concealed through net average calculations, we also tried using a convolutional neural network (CNN) to extract positional-dependent differences in amino acid sequence between effectors and non-effectors. We used the first 100 amino acid values for each feature and padded the ends with zeros for sequences shorter than 100 aa. For each protein in our positive and negative training sets, this feature encoding resulted in a 6 x 100 matrix to include every positional encoding for all six biochemical features in the first 100 amino acids.

Machine learning architectures

For the classical machine learning pipeline, we tested various parametric and non-parametric models and evaluated what worked best for effector classification. We tested and compared the following models using multiple cross-validation trials (resampling procedures that rotate training and testing splits): Random Forest, Naïve Bayes (Bernoulli and Gaussian), Logistic Regression, and K-Nearest Neighbors. To reduce overfitting, we monitored training and testing scores (accuracy, sensitivity, specificity, false positive rate, and area under the precision-recall curve). Models were used from the scikit-learn machine learning library in Python (Pedregosa et al. 2011). We tested the Logistic Regression, and Gaussian/Bernoulli Naive-Bayes models with default parameters, as well as the Random Forest classifier with 400 estimators and a max depth of 4, and the K-Nearest Neighbors classifier with the neighbors parameter set to 5.

Lineage Specificity

Proteins encoded by lineage specific genes were obtained using the predicted secretome of a species as a BLASTp query against the translated open reading frames (ORFs) of other oomycete species (Supplemental Table 2). Conserved proteins were defined as proteins with a BLASTp hit in another species with an e-value cutoff of $<10e-7$, percent identity of $>30\%$, and a query coverage of $>30\%$. Protein sequences that were not conserved in another oomycete species were considered to be lineage-specific. Similar BLASTp cutoffs are used in other pipelines to obtain species-specific genes for numerous species, including other oomycetes (Rujirawat et al. 2019; Zhou et al. 2015). ORFs were used instead of gene

models because most oomycete effectors are encoded by a single exon and because gene prediction algorithms often misannotate effector genes. Orthofinder (Emms and Kelly 2015) was tested as well but was found to predict a much higher number of lineage specific proteins (and thus likely contained more false positives) so we used our more conservative BLASTp-based pipeline for EffectorO-LSP.

Prediction of RXLR-EER motifs and WY domains

To predict the RXLR-EER motif, we used a string search for [RQGH]XLR and [DE][DE][KR] for EER as well as the HMM for RXLR-EER developed by Whisson et al. (2007) for *Phytophthora* species. To predict WY domains, we used the HMM training set containing three *Phytophthora* species by Boutemy et al. (2011), with proteins with a HMMer bitscore >0 classified as having a WY domain. HMMer v3.1 was used for all HMM searches (Eddy 2011).

Scripts

All scripts for running EffectorO-ML and EffectorO-LSG can be found on Github at <https://github.com/mjnur/oomycete-effector-prediction>. A webapp for EffectorO-ML is available at <https://bremia.ucdavis.edu/effectorO.php>.

Candidate gene cloning and agroinfiltration

The Avr6 candidate *BLE01* was cloned from *B. lactucae* genomic DNA using Gateway cloning into pEG100 (35S promoter) as described previously (Wood et al. 2020). The signal

peptide was excluded from the coding sequence to ensure expression inside the plant cell. Two distinct alleles were obtained. *Agrobacterium tumefaciens* strain C58Rif⁺ was transformed with pEG100-BLE01 and infiltrations were performed on three- to five-week-old lettuce plants as described previously (Wood et al. 2020). Leaves were scored for necrosis 4-5 days after infiltration. *A. tumefaciens* containing pEG100 (empty vector) and pEG100-GFP were used as negative controls, and *A. tumefaciens* containing pBAV139-35S:HopM1 and pBAV139-35S:AvrPto were used as positive controls for cell death (Wroblewski et al. 2009).

Results

Effector and non-effector proteins differ in their biochemical characteristics

To determine whether various N-terminal amino acid characteristics would be sufficient to distinguish effectors from non-effector proteins, we compared predicted biochemical parameters between validated oomycete effectors and secreted, conserved proteins (presumed non-effectors). We found that validated effectors had higher intrinsic disorder and had lower scores for GRAVY, hydrophobicity, exposed residues, bulkiness, and interface, compared with secreted, conserved proteins (Fig. 1A). The intrinsic disorder score was significantly higher in validated effectors, starting from just after the signal peptide (first ~20 aa) until approximately 80 aa after the start codon (Fig. 1B).

Evaluation of machine learning effector classifier models

To find the effector classifier with the best performance, we evaluated several machine learning models (Random Forest, Linear SVCs, Bernoulli and Gaussian Naïve Bayes,

Logistic Regression, and K-Nearest Neighbors) as well as a convolutional neural network model that takes into account positional information for classification. The Random Forest classifier initially outperformed other models, so we attempted several combinations of parameters, adjusting the number of estimators (`n_estimators`) and maximum tree depth (`max_depth`), while monitoring the cross-validation accuracy. The best receiver-operator curve (ROC) was achieved by the Random Forest classifier (with parameters `n_estimators = 400` and `max_depth = 4`) with an area under the curve (AUC) of 0.89 (Figure 2) and an overall 5-fold cross-validation accuracy of 84.26% (Table 1). We also tested a convolutional neural network (CNN), which takes position into account, to predict effectors; however, it was found to perform similarly but not better than the most accurate Random Forest model. Therefore, we used the Random Forest model for our EffectorO-ML pipeline to classify effectors based on the biochemical characteristics of the first 100 amino acids of the N-terminus for all subsequent analyses.

Figure 1. (A) Amino acid biochemical characteristics for validated oomycete effectors and secreted non-effectors as calculated by the ExPASyProtScale and used as features for the machine learning classifier. Violin plot shows the distribution of the N-terminal Z scores for each group of proteins, with the p-value shown above plot. **(B)** Intrinsic disorder as calculated by PONDR VSL2 for validated oomycete effectors and secreted non-effectors with the *B. lactucae* secretome for reference. The average and standard error are shown for each group of proteins at each amino acid position.

Figure 2. Receiver-operator curves (ROCs) for the various machine learning models tested on a cross validation fold, using 1:1 effector to non-effector testing and training sets.

The Random Forest effector classifier uses a default probability threshold of 0.5 to classify a protein as an effector; this probability threshold can be modified depending on preference for either inclusivity/sensitivity or specificity. For the top performing Random Forest model, we calculated and averaged accuracy, specificity, sensitivity (true positive rate), and false positive rate along the five cross-validation folds for different probability thresholds, illustrating the tradeoffs between rates of false positives and true positives when varying probability thresholds (Table 2). To obtain another estimate of the false positive and false negative rate, we ran EffectorO-ML on a set of 634 predicted secreted proteins that were conserved in at least 24 of the 28 sequenced oomycete species used for EffectorO-LSP

(presumed non-effectors) as well as on the 41 secreted *B. lactucae* WY effectors (presumed true effectors). Secreted conserved orthologs and secreted WY proteins had significantly different distributions (Figure 3). We found that a probability cutoff of 0.5 resulted in approximately 28% of these conserved proteins being classified as effectors while capturing 90% of the WY effectors (Table 3). At a probability cutoff of 0.6, only 16% of the conserved proteins were classified as effectors, with 85% of the WY proteins classified as effectors (Table 3). Since we presume most of the conserved secreted proteins are not effectors due to lineage specificity, 16% would be a good estimate of the false positive rate at the 0.6 threshold, which is similar to the false positive rate for string searches for the RXLR motif (Wood et al. 2020). The false positive rate can be decreased further if desired by increasing the probability threshold, but at the loss of true effectors. For inclusive searches (especially if combined with other lines of evidence), we recommend between 0.5 and 0.6 as the threshold for analyses of oomycete genomes.

Figure 3. Violin plots of effector probabilities for conserved secreted oomycete orthologs and secreted *B. lactucae* WY domain containing proteins, generated by the top-performing Random Forest model.

Table 1. Average cross validation metrics for various machine learning models for classifying proteins as effectors or non-effectors

	Random Forest	Logistic Regression	K-Nearest Neighbors	Bernoulli Naïve Bayes	Gaussian Naïve Bayes
Accuracy	84.26%	65.90%	70.16%	63.61%	67.87%
Specificity	82.97%	63.31%	69.12%	30.95%	61.14%
Sensitivity	86.95%	68.73%	71.10%	96.09%	74.59%
False Positive Rate	17.03%	36.69%	30.88%	69.05%	38.86%

Table 2. Metrics at different probability thresholds for the most accurate Random Forest model

	Random Forest Model Effector Probability Threshold =		
	0.7	0.6	0.5
Accuracy	76.23%	79.02%	84.26%
Specificity	89.84%	85.09%	82.97%
Sensitivity	62.74%	73.00%	86.95%
False Positive Rate	10.16%	14.91%	17.03%

Table 3. Number and percent of oomycete conserved secreted orthologous proteins (634 total) and *B. lactucae* secreted WY proteins (41 total) classified as an effector by the EffectorO-ML Random Forest model at different effector probability thresholds

	Random Forest Model Effector Probability Threshold =				
	0.9	0.8	0.7	0.6	0.5
Secreted Orthologs	2 (0.3%)	12 (1.9%)	54 (8.5%)	103 (16%)	180 (28%)
Secreted WY Proteins	1 (2.5%)	5 (12.5%)	20 (50%)	34 (85%)	36 (90%)

EffectorO predicts novel effectors in oomycete species and recovers known effectors

To find novel effectors from *P. infestans* and *B. lactucae*, we used our EffectorO-ML and LSP pipelines on the predicted secreted ORFs of these genomes. For *P. infestans*, we found 8,685 proteins (41% of the secretome) to be lineage-specific and 5,814 proteins (30%) to be classified as candidate effectors by machine learning, with 1,235 proteins (6%) present in both categories (Figure 4). For *B. lactucae*, we found 2,478 proteins (34% of the secretome) to be lineage-specific and 1,777 proteins (24%) to be classified as candidate effectors by machine learning, with 597 proteins (8%) present in both categories (Figure 4). EffectorO was largely able to recover proteins containing effector signatures, such as the RXLR-EER motif and the WY domain (Figure 4 and Supplemental Table 3), which were not used as classifiers in our analysis but were contained in proteins in the positive training set.

Figure 4. Venn diagrams showing the number of open reading frames (ORFs) identified in the predicted secretomes of *P. infestans* and *B. lactucae* that are lineage specific proteins (LSPs) and/or predicted to be an effector by the machine learning classifier (ML). The number of RXLR-EER and WY proteins that are included or excluded from these datasets are also shown.

The majority of EffectorO candidates are expressed in B. lactucae

To investigate whether EffectorO candidates are expressed at the transcript level, we used an available RNA-seq dataset of *B. lactucae* grown on lettuce seedlings (NCBI BioProject PRJNA523226). We found that transcripts that encoded proteins predicted to be effectors by EffectorO-ML and EffectorO-LSP had similar mean expression levels as transcripts encoding proteins containing WY domains and slightly lower expression than those containing RXLR-EER motifs (Figure 5). Transcripts that encoded proteins predicted to be effectors by either the ML or LSP pipeline, but not both, had lower mean expression levels and more proteins with zero reads present in the RNA-seq dataset (Figure 5). The full secretome had the lowest mean expression and also had 1,716 predicted transcripts with zero reads, suggesting that many of the predicted ORFs may not represent real genes (Figure 5).

Figure 5. RNA-seq expression levels (log of the length-normalized read counts) for candidate effector transcripts encoding proteins with the RXLR-EER motif (RXLR-EER), WY domain (WY), EffectorO-ML -LSP (ML and LSP), EffectorO-ML only (ML), EffectorO-LSP only (LSP), or the full secretome.

EffectorO-ML predicts candidate effectors in diverse oomycete pathogens

To determine whether EffectorO could predict effectors in other pathogens across the phylum Oomycota, we tested EffectorO-ML on an additional 27 published oomycete genomes. We found that even though EffectorO was trained primarily on effectors from downy mildew and *Phytophthora* pathogens, the algorithm predicted hundreds to thousands of candidate effectors in distantly related oomycetes infecting animals (e.g., *Achlya hypogyna*, *Aphanomyces* spp., *Saprolegnia* spp.), as well as oomycetes from the genera *Albugo* and *Pythium* (Figure 6). Critically, EffectorO-ML predicted a similar number of effectors in the free-living oomycete species *Thraustotheca clavata* (14.6% of the genome) as the false positive rate for EffectorO (17.0%, Table 1).

Figure 6. Number of predicted secreted proteins (translated ORFs) classified as effectors using EffectorO-ML only (cutoff = 0.5) (dark blue bars), EffectorO-LSP only (pink bars), EffectorO-ML and LSP (purple bars), and those classified as non-effectors (grey bars) in 28 species of oomycetes, including Peronosporales (*Phytophthora* and downy mildew species), *Pythium* species, animal pathogens (*Achlya*, *Saprolegnia*, and *Aphanomyces* spp.), *Albugo*

spp., and a free-living oomycete (*Thraustotheca clavata*). Numbers are estimated by subtracting the number of EffectorO-ML hits from the false positive rate for each genome (14.6%, estimated using *T. clavata*, which had fewer predicted effectors than the estimated false positive rate of 17.0%).

An EffectorO candidate causes cell death when expressed in lettuce

To see whether any EffectorO candidates are recognized effectors (i.e., have avirulence activity), we cloned a gene from *B. lactucae* that encoded a protein that does not have RXLR or EER motifs nor predicted WY domains, but was classified as a probable effector by both EffectorO-ML and Effector-LSP. The gene encoding this protein was not found in the gene models (i.e. was not predicted to be a gene by the annotation software), but did encode a predicted ORF. The candidate effector, named BLE01 for (*Bremia lactucae* EffectorO candidate 1), was within the mapping region for *Avr6*, as determined by sequencing of segregating sexual progeny of *B. lactucae* isolates (R.J. Gil, unpublished data). We found that transient expression of *BLE01* caused cell death on the lettuce cultivar Sabine, which contains *Dm6*, but not Cobham Green, which does not contain any known *Dm* genes (Fig 7A). When we screened a larger panel of lettuce cultivars with various *Dm* genes, we found that it also elicited cell death in RYZ2164, but not in the other seven cultivars with other *Dm* genes or in the non-host plant *Nicotiana benthamiana* (Fig 7B).

Figure 7. (A) Example transient expression results from lettuce infiltrated with *A. tumefaciens* containing two candidate effector alleles from *B. lactucae* (BLE01-H1 and BLE01-H2) as well as with the negative controls empty vector (EV) and green fluorescent protein (GFP) and the positive controls for cell death, *Pseudomonas syringae* pv. *tomato* effectors HopM1 and AvrPto. Leaves are representative of four replicates and photos were taken 4 days after infiltration. **(B)** Necrosis scores four days post-agroinfiltration of the candidate effectors and controls in a panel of lettuce cultivars and the non-host *Nicotiana benthamiana*. Necrosis was scored as follows: 0 = no necrosis or chlorosis, 1 = mild chlorosis, 2 = chlorosis, 3 = chlorosis and necrosis, 4 = full necrosis. Number of replicates (N), averages, and standard deviations are shown for each line and effector combination.

Discussion

We demonstrated that our EffectorO pipeline can predict novel candidate effectors using a machine learning-based approach combined with a lineage-specificity criterion for oomycete effector classification. Therefore, EffectorO can be used to mine oomycete

genomes to expand the set of predicted effectors without relying on canonical N-terminal motifs (such as RXLR and EER) or C-terminal domains (such as the WY domain); thus, it may reveal effectors that were previously missed by only searching for these features. EffectorO predicts an additional ~600 ML+LSP candidate effectors from *B. lactucae* and ~1,200 additional ML+LSP candidate effectors from *P. infestans*. The percent of effectors predicted in the secretome by EffectorO-ML was approximately 25%, which is similar to what was predicted for fungal secretomes with EffectorP (Sperschneider et al. 2018). The number of potentially novel effectors in *B. lactucae* and *P. infestans* that were identified is much larger than previous reports of effector repertoires from these two species (Fletcher et al. 2019; Haas et al. 2009). The large size of these candidate sets should not preclude functional analysis when combined with additional data from the pathogen of interest such as avirulence gene mapping, population genomics, and gene expression data.

EffectorO-ML was trained on the N-terminal sequences of validated effectors compared to conserved secreted proteins (presumed non-effectors), which we show differ in several biochemical characteristics and especially in the degree of N-terminal disorder. Sequences in the N-terminus of effector proteins are known to be important for secretion and translocation of effectors; here we demonstrate that they have distinct biochemical signatures that can be used to predict the probability that a given sequence is an effector. EffectorO-ML was able to recover the majority of predicted RXLR-EER and WY domain containing proteins from *B. lactucae* and *P. infestans*, without relying on using those motifs for prediction. By focusing on the N-terminus for effector classification, our pipeline allows for a diversity of C-terminal domains, which are typically considered to be the functional region of effector proteins. EffectorO may help to uncover new effector families, which can

then be bioinformatically characterized using motif/domain-based analysis, as well as structurally characterized using cryo-EM or x-ray crystallography of effector proteins, as has been done for WY-containing effectors (Boutemy et al. 2011). The predicted effector repertoires should also be analyzed with the latest ML structural prediction algorithms such as AlphaFold (Senior et al. 2020) and trRosetta (Yang et al. 2020). Recently, trRosetta was used to predict effectors based on structure for the fungal species *Magnaporthe oryzae* (Seong and Krasileva 2021) and a similar analysis should be done with EffectorO candidates.

The most accurate model for EffectorO-ML was the Random Forest classifier. Random Forest classifiers work by building many decision trees and then taking the consensus of all of the decision trees to calculate an overall probability for a given classification (Pavlov 2000). Because the final output is the average of many independent models, Random Forest classifiers are robust to individual errors in any given single model. The fungal effector classifier, EffectorP 2.0, uses a similar approach by utilizing an ensemble classifier that takes the consensus predictions of multiple Naive-Bayes and decision trees to obtain an accuracy of 89% for predicting fungal effectors. While EffectorP is trained on a larger set of effectors than EffectorO and relies on additional biochemical characteristics (16 features vs. our six), we achieved a similar accuracy to EffectorP 2.0 in predicting oomycete effectors. Our Random Forest model had an estimated 84% accuracy and 82% specificity at a probability cutoff of 0.5. The false positive rate can be decreased by increasing the probability cutoff for classification of an effector; however, this will also result in more false negatives (missed true effectors). We recommend 0.5 as an inclusive threshold for casting a wide net to catch effectors; however, depending on the application the researcher may want to use a more stringent threshold to find higher confidence effectors. It may also be possible to further

improve the accuracy and specificity of EffectorO by adding additional proteins to the training set or by adding more biochemical and structural characteristics. Gene models may be used instead of ORFs to reduce false positives, although gene prediction algorithms may miss true effectors (as was the case for the *Avr6* candidate *BLE01*).

Similar to EffectorO-ML, EffectorO-LSP will predict additional proteins besides effectors, not only due to false positives but also due to the biological reality that there may be lineage specific proteins that fulfill other functions. Thus, EffectorO-LSP predictions are most useful in combination with other lines of evidence of effector activity (such as EffectorO-ML, expression during infection, RXLR motifs and/or WY domains). Approximately 41% of the *P. infestans* secretome and 34% of the *B. lactucae* secretome were predicted to be lineage specific. Some of these predicted proteins may not actually be lineage-specific; some may be encoded by pseudogenes or result from failure to detect homology (Weisman et al. 2020). We found that many were expressed in *B. lactucae* during infection, suggesting that some of these represent true genes. In addition to effectors, these lineage-specific secreted proteins could also be related to the lifestyles of these organisms, their sexual reproduction, or may be products of co-evolution with other microbes rather than co-evolution with the host. The amount of lineage specificity expected for a given organism will depend on whether or not closely related organisms are used in the analysis and how closely pathogens are adapted to their hosts. Depending on the relatedness of a given species to those in the database, we recommend that BLAST results from very closely related species be filtered out in order to find proteins that are truly lineage specific.

EffectorO predictions were validated for an effector from *B. lactucae* that was within the region co-segregating with *Avr6* but did not have RXLR or EER motifs nor any WY

domains. Here we have shown that this effector, *BLE01*, causes cell death when expressed in lettuce with the *Dm6* gene but not on cultivars containing other known *Dm* genes, making it a good candidate for the protein encoded by *Avr6*. Expression of *BLE01* also caused cell death on RYZ2164. The background of RYZ2164 is proprietary, so it may have *Dm6* in its lineage or it may contain another *Dm* gene that recognizes the effector.

EffectorO-ML and -LSP can expand the set of predicted effectors for identification of genes for resistance to agriculturally important oomycetes. These predictions can be used to choose candidates for “effectoromics” (Vleeshouwers and Oliver 2015) screens to identify effectors with avirulence and virulence activities. EffectorO predictions will be especially useful when combined with mapping of avirulence phenotypes for the cloning of specific avirulence genes enabling an effector-driven approach to R gene identification.

Acknowledgements

The authors would like to thank John Emerson for greenhouse support, Elizabeth Georgian for manuscript editing and formatting, Kyle Fletcher for very helpful comments and suggestions, and Mary Corrigan for additional minor editing.

Literature Cited

Bailey, K., Cevik, V., Holton, N., Byrne-Richardson, J., Sohn, K. H., Coates, M., Woods-Tör, A., Aksoy, H. M., Hughes, L., Baxter, L., Jones, J. D. G., Beynon, J., Holub, E. B., and Tör, M. 2011. Molecular cloning of ATR5(Emoy2) from *Hyaloperonospora arabidopsidis*, an avirulence determinant that triggers RPP5-mediated defense in Arabidopsis. *Molecular Plant-Microbe Interactions*. 24:827–838

478 Baldauf, S. L. 2003. The Deep Roots of Eukaryotes. *Science*. 300:1703–1706

479 Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A.,
480 Anderson, R., Badejoko, W., Bittner-Eddy, P., Boore, J. L., Chibucos, M. C., Coates, M.,
481 Dehal, P., Delehaunty, K., Dong, S., Downton, P., Dumas, B., Fabro, G., Fronick, C.,
482 Fuerstenberg, S. I., Fulton, L., Gaulin, E., Govers, F., Hughes, L., Humphray, S., Jiang, R.
483 H. Y., Judelson, H., Kamoun, S., Kyung, K., Meijer, H., Minx, P., Morris, P., Nelson, J.,
484 Phuntumart, V., Qutob, D., Rehmany, A., Rougon-Cardoso, A., Ryden, P., Torto-
485 Alalibo, T., Studholme, D., Wang, Y., Win, J., Wood, J., Clifton, S. W., Rogers, J., Van den
486 Ackerveken, G., Jones, J. D. G., McDowell, J. M., Beynon, J., and Tyler, B. M. 2010.
487 Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora*
488 *arabidopsidis* genome. *Science*. 330:1549–1551

489 Birch, P. R. J., Armstrong, M., Bos, J., Boevink, P., Gilroy, E. M., Taylor, R. M., Wawra, S.,
490 Pritchard, L., Conti, L., Ewan, R., Whisson, S. C., van West, P., Sadanandom, A., and
491 Kamoun, S. 2009. Towards understanding the virulence functions of RXLR effectors
492 of the oomycete plant pathogen *Phytophthora infestans*. *Journal of Experimental*
493 *Botany*. 60:1133–1140

494 Bos, J. I. B., Armstrong, M. R., Gilroy, E. M., Boevink, P. C., Hein, I., Taylor, R. M., Zhendong, T.,
495 Engelhardt, S., Vetukuri, R. R., Harrower, B., Dixelius, C., Bryan, G., Sadanandom, A.,
496 Whisson, S. C., Kamoun, S., and Birch, P. R. J. 2010. *Phytophthora infestans* effector
497 AVR3a is essential for virulence and manipulates plant immunity by stabilizing host
498 E3 ligase CMPG1. *Proceedings of the National Academy of Sciences*. 107:9909–9914

499 Boutemy, L. S., King, S. R. F., Win, J., Hughes, R. K., Clarke, T. A., Blumenschein, T. M. A.,
500 Kamoun, S., and Banfield, M. J. 2011. Structures of *Phytophthora* RXLR effector

proteins: a conserved but adaptable fold underpins functional diversity. *Journal of Biological Chemistry*. 286:35834–35842

Deb, D., Anderson, R. G., How-Yew-Kin, T., Tyler, B. M., and McDowell, J. M. 2018. Conserved RxLR Effectors From Oomycetes *Hyaloperonospora arabidopsidis* and *Phytophthora sojae* Suppress PAMP- and Effector-Triggered Immunity in Diverse Plants. *Molecular Plant-Microbe Interactions*. 31:374–385

Derevnina, L., Petre, B., Kellner, R., Dagdas, Y. F., Sarowar, M. N., Giannakopoulou, A., De la Concepcion, J. C., Chaparro-Garcia, A., Pennington, H. G., van West, P., and Kamoun, S. 2016. Emerging oomycete threats to plants and animals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371

Dong, S., Raffaele, S., and Kamoun, S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development*. 35:57–65

Dou, D., Kale, S. D., Wang, X., Jiang, R. H. Y., Bruce, N. A., Arredondo, F. D., Zhang, X., and Tyler, B. M. 2008. RXLR-mediated entry of *Phytophthora sojae* effector Avr1b into soybean cells does not require pathogen-encoded machinery. *Plant Cell*. 20:1930–1947

Eddy, S. R. 2011. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* 7:e1002195

Ellis, J. G., and Dodds, P. N. 2011. Showdown at the RXLR motif: Serious differences of opinion in how effector proteins from filamentous eukaryotic pathogens enter plant cells. *Proceedings of the National Academy of Sciences*. 108:14381–14382

Emms, D. M., and Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*. 16:157

524 Emms, D. M., and Kelly, S. 2019. OrthoFinder: phylogenetic orthology inference for
525 comparative genomics. *Genome Biology*. 20:238

526 ExPaSy (Expert Protein Analysis System). 2004. *Encyclopedic Dictionary of Genetics,*
527 *Genomics and Proteomics.*

528 Fabro, G., Steinbrenner, J., Coates, M., Ishaque, N., Baxter, L., Studholme, D. J., Körner, E.,
529 Allen, R. L., Piquerez, S. J. M., Rougon-Cardoso, A., Greenshields, D., Lei, R., Badel, J. L.,
530 Caillaud, M.-C., Sohn, K.-H., Van den Ackerveken, G., Parker, J. E., Beynon, J., and
531 Jones, J. D. G. 2011. Multiple Candidate Effectors from the Oomycete Pathogen
532 *Hyaloperonospora arabidopsidis* Suppress Host Plant Immunity. *PLOS Pathogens*.
533 7:e1002348

534 Fletcher, K., Gil, J., Bertier, L. D., Kenefick, A., Wood, K. J., Zhang, L., Reyes-Chin-Wo, S.,
535 Cavanaugh, K., Tsuchida, C., Wong, J., and Micheltore, R. 2019. Genomic signatures
536 of heterokaryosis in the oomycete pathogen *Bremia lactucae*. *Nature*
537 *Communications*. 10:2645

538 Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. 2012. CD-HIT: accelerated for clustering the next-
539 generation sequencing data. *Bioinformatics*. 28:3150–3152

540 Giesbers, A. K. J., Pelgrom, A. J. E., Visser, R. G. F., Niks, R. E., Van den Ackerveken, G., and
541 Jeuken, M. J. W. 2017. Effector-mediated discovery of a novel resistance gene against
542 *Bremia lactucae* in a nonhost lettuce species. *New Phytologist*. 216:915–926

543 Godfrey, D., Böhlenius, H., Pedersen, C., Zhang, Z., Emmersen, J., and Thordal-Christensen, H.
544 2010. Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif.
545 *BMC Genomics*. 11:317

Gu, B., Kale, S. D., Wang, Q., Wang, D., Pan, Q., Cao, H., Meng, Y., Kang, Z., Tyler, B. M., and Shan, W. 2011. Rust Secreted Protein Ps87 Is Conserved in Diverse Fungal Pathogens and Contains a RXLR-like Motif Sufficient for Translocation into Plant Cells. PLOS ONE. 6:e27217

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M. V., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I. B., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J., Grünwald, N. J., Horn, K., Horner, N. R., Hu, C.-H., Huitema, E., Jeong, D.-H., Jones, A. M. E., Jones, J. D. G., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L., Maclean, D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J. G., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J. I., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F., Birch, P. R. J., Whisson, S. C., Judelson, H. S., and Nusbaum, C. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature. 461:393–398

Haverkort, A. J., Boonekamp, P. M., Hutten, R., Jacobsen, E., Lotz, L. A. P., Kessel, G. J. T., Visser, R. G. F., and van der Vossen, E. A. G. 2008. Societal Costs of Late Blight in

Potato and Prospects of Durable Resistance Through Cisgenic Modification. *Potato Research*. 51:47–57

He, J., Ye, W., Choi, D. S., Wu, B., Zhai, Y., Guo, B., Duan, S., Wang, Y., Gan, J., Ma, W., and Ma, J. 2019. Structural analysis of *Phytophthora* suppressor of RNA silencing 2 (PSR2) reveals a conserved modular fold contributing to virulence. *Proceedings of the National Academy of Sciences*. 116:8054–8059

Jiang, R. H. Y., de Bruijn, I., Haas, B. J., Belmonte, R., Löbach, L., Christie, J., van den Ackerveken, G., Bottin, A., Bulone, V., Díaz-Moreno, S. M., Dumas, B., Fan, L., Gaulin, E., Govers, F., Grenville-Briggs, L. J., Horner, N. R., Levin, J. Z., Mammella, M., Meijer, H. J. G., Morris, P., Nusbaum, C., Oome, S., Phillips, A. J., van Rooyen, D., Rzeszutek, E., Saraiva, M., Secombes, C. J., Seidl, M. F., Snel, B., Stassen, J. H. M., Sykes, S., Tripathy, S., van den Berg, H., Vega-Arreguin, J. C., Wawra, S., Young, S. K., Zeng, Q., Dieguez-Uribeondo, J., Russ, C., Tyler, B. M., and van West, P. 2013. Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen *Saprolegnia parasitica*. *PLOS Genetics*. 9:e1003272

Kale, S. D., Gu, B., Capelluto, D. G. S., Dou, D., Feldman, E., Rumore, A., Arredondo, F. D., Hanlon, R., Fudal, I., Rouxel, T., Lawrence, C. B., Shan, W., and Tyler, B. M. 2010. External Lipid PI3P Mediates Entry of Eukaryotic Pathogen Effectors into Plant and Animal Host Cells. *Cell*. 142:284–295

King, S. R. F., McLellan, H., Boevink, P. C., Armstrong, M. R., Bukharova, T., Sukarta, O., Win, J., Kamoun, S., Birch, P. R. J., and Banfield, M. J. 2014. *Phytophthora infestans* RXLR effector PexRD2 interacts with host MAPKKK ϵ to suppress plant immune signaling. *Plant Cell*. 26:1345–1359

592 Li, W., and Godzik, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of
593 protein or nucleotide sequences. *Bioinformatics*. 22:1658–1659

594 McGowan, J., and Fitzpatrick, D. A. 2017. Genomic, Network, and Phylogenetic Analysis of
595 the Oomycete Effector Arsenal. *mSphere*. 2

596 Pavlov, Y. L. 2000. Random Forests. VSP Publishers, Utrecht, the Netherlands.

597 Pecrix, Y., Buendia, L., Penouilh-Suzette, C., Maréchaux, M., Legrand, L., Bouchez, O., Rengel,
598 D., Gouzy, J., Cottret, L., Vear, F., and Godiard, L. 2019. Sunflower resistance to
599 multiple downy mildew pathotypes revealed by recognition of conserved effectors
600 of the oomycete *Plasmopara halstedii*. *Plant J*. 97:730–748

601 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel, M.,
602 Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D.,
603 Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine Learning in Python. *J*
604 *Mach Learn Res*. 2011;12: 2825–2830.

605 Pelgrom, A. J. E., Eikelhof, J., Elberse, J., Meisrimler, C.-N., Raedts, R., Klein, J., and Van den
606 Ackerveken, G. 2019. Recognition of lettuce downy mildew effector BLR38 in
607 *Lactuca serriola* LS102 requires two unlinked loci. *Molecular Plant Pathology*.
608 20:240–253

609 Pel, M. A., Foster, S. J., Park, T.-H., Rietman, H., van Arkel, G., Jones, J. D. G., Van Eck, H. J.,
610 Jacobsen, E., Visser, R. G. F., and Van der Vossen, E. A. G. 2009. Mapping and cloning
611 of late blight resistance genes from *Solanum venturii* using an interspecific candidate
612 gene approach. *Molecular Plant-Microbe Interactions*. 22:601–615

613 Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K., and Obradovic, Z. 2006. Length-dependent
614 prediction of protein intrinsic disorder. *BMC Bioinformatics*. 7:208

615 Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: discriminating
616 signal peptides from transmembrane regions. *Nat. Methods.* 8:785–786

617 Rehmany, A. P., Gordon, A., Rose, L. E., Allen, R. L., Armstrong, M. R., Whisson, S. C., Kamoun,
618 S., Tyler, B. M., Birch, P. R. J., and Beynon, J. L. 2005. Differential recognition of highly
619 divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two
620 *Arabidopsis* lines. *Plant Cell.* 17:1839–1850

621 Rujirawat, T., Patumcharoenpol, P., Kittichotirat, W., and Krajaejun, T. 2019. Oomycete
622 Gene Table: an online database for comparative genomic analyses of the oomycete
623 microorganisms. *Database.* 2019

624 Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Židek, A., Nelson,
625 A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P.,
626 Jones, D. T., Silver, D., Kavukcuoglu, K., and Hassabis, D. 2020. Improved protein
627 structure prediction using potentials from deep learning. *Nature.* 577:706–710

628 Seong, K., and Krasileva, K. V. 2021. Computational structural genomics unravels common
629 folds and predicted functions in the secretome of fungal phytopathogen
630 *Magnaporthe oryzae*. *bioRxiv* doi: 10.1101/2021.01.25.427855

631 Sharma, R., Xia, X., Cano, L. M., Evangelisti, E., Kemen, E., Judelson, H., Oome, S., Sambles, C.,
632 van den Hoogen, D. J., Kitner, M., Klein, J., Meijer, H. J. G., Spring, O., Win, J., Zipper, R.,
633 Bode, H. B., Govers, F., Kamoun, S., Schornack, S., Studholme, D. J., Van den
634 Ackerveken, G., and Thines, M. 2015. Genome analyses of the sunflower pathogen
635 *Plasmopara halstedii* provide insights into effector evolution in downy mildews and
636 *Phytophthora*. *BMC Genomics.* 16:741

637 Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., and Taylor, J. M. 2018. Improved
638 prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular*
639 *Plant Pathology*. 19:2094–2110

640 Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J.
641 M., and Taylor, J. M. 2016. EffectorP: predicting fungal effector proteins from
642 secretomes using machine learning. *New Phytol.* 210:743–761

643 Stassen, J. H. M., den Boer, E., Vergeer, P. W. J., Andel, A., Ellendorff, U., Pelgrom, K., Pel, M.,
644 Schut, J., Zonneveld, O., Jeuken, M. J. W., and Van den Ackerveken, G. 2013. Specific In
645 Planta Recognition of Two GCLR Proteins of the Downy Mildew *Bremia lactucae*
646 Revealed in a Large Effector Screen in Lettuce. *Molecular Plant-Microbe*
647 *Interactions*. 26:1259–1270

648 Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S. Y., De Silva, N.,
649 Martinez, M. C., Pedro, H., Yates, A. D., Hassani-Pak, K., and Hammond-Kosack, K. E.
650 2019. PHI-base: the pathogen–host interactions database. *Nucleic Acids Research*.

651 Vleeshouwers, V. G. A. A., and Oliver, R. P. 2015. Effectors as Tools in Disease Resistance
652 Breeding Against Biotrophic, Hemibiotrophic, and Necrotrophic Plant Pathogens.
653 *Molecular Plant-Microbe Interactions*. 2015:17–27

654 Wang, Y., Tyler, B. M., and Wang, Y. 2019. Defense and Counterdefense During Plant-
655 Pathogenic Oomycete Infection. *Annu. Rev. Microbiol.* 73:667–696

656 Weisman, C. M., Murray, A. W., and Eddy, S. R. 2020. Many but not all lineage-specific genes
657 can be explained by homology detection failure. *PLOS Biology* 18(11):e3000862.

658 Whisson, S. C., Boevink, P. C., Moleleki, L., Avrova, A. O., Morales, J. G., Gilroy, E. M.,
659 Armstrong, M. R., Grouffaud, S., van West, P., Chapman, S., Hein, I., Toth, I. K.,

- Pritchard, L., and Birch, P. R. J. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature*. 450:115–118
- Win, J., Krasileva, K. V., Kamoun, S., Shirasu, K., Staskawicz, B. J., and Banfield, M. J. 2012. Sequence Divergent RXLR Effectors Share a Structural Fold Conserved across Plant Pathogenic Oomycete Species. *PLOS Pathogens*. 8:e1002400
- Wood, K., Nur, M., Gil, J., Fletcher, K., Lakeman, K., Gothberg, A., Khuu, T., Kopetzky, J., Pandya, A., Pel, M., and Micheltore, R. 2020. Effector prediction and characterization in the oomycete pathogen *Bremia lactucae* reveal host-recognized WY domain proteins that lack the canonical RXLR motif. *PLOS Pathogens* 10: e1009012
- Wroblewski T., Caldwell K .S., Piskurewicz U., Cavanaugh K. A., Xu H., Kozik A., Ochoa O., McHale L. K., Lahre K., Jelenska J., Castillo J. A., Blumenthal D., Vinatzer B. A., Greenberg J. T., and Micheltore R. W. 2009. Comparative large-scale analysis of interactions between several crop species and the effector repertoires from multiple pathovars of *Pseudomonas* and *Ralstonia*. *Plant Physiology*. 150: 1733–49. doi:10.1104/pp.109.140251
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. 2020. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*. 117:1496–1503
- Zheng, X., McLellan, H., Fraiture, M., Liu, X., Boevink, P. C., Gilroy, E. M., Chen, Y., Kandel, K., Sessa, G., Birch, P. R. J., and Brunner, F. 2014. Functionally Redundant RXLR Effectors from *Phytophthora infestans* Act at Different Steps to Suppress Early flg22-Triggered Immunity. *PLOS Pathogens*. 10:e1004057

683 Zhou, K., Huang, B., Zou, M., Lu, D., He, S., and Wang, G. 2015. Data in support of genome-
684 wide identification of lineage-specific genes within *Caenorhabditis elegans*. Data in
685 Brief. 4:595–601