TECHNICAL ADVANCE

***effectR*: An Expandable R Package to Predict Candidate RxLR and CRN Effectors in Oomycetes Using Motif Searches**

Javier F. Tabima,¹ and Niklaus J. Grünwald^{2,†}

¹ Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, U.S.A.

² Horticultural Crops Research Laboratory, USDA-ARS, Corvallis, OR 97330, U.S.A.

Accepted 19 March 2019.

Effectors are small, secreted proteins that facilitate infection of host plants by all major groups of plant pathogens. Effector protein identification in oomycetes relies on identification of open reading frames with certain amino acid motifs among additional minor criteria. To date, identification of effectors relies on custom scripts to identify motifs in candidate open reading frames. Here, we developed the R package *effectR*, which provides a convenient tool for rapid prediction of effectors in oomycete genomes, or with custom scripts for any genome, in a reproducible way. The *effectR* package relies on a combination of regular expressions statements and hidden Markov model approaches to predict candidate RxLR and crinkler effectors. Other custom motifs for novel effectors can easily be implemented and added to package updates. The *effectR* package has been validated with published oomycete genomes. This package provides a convenient tool for wet lab researchers interested in reproducible identification of candidate effectors in oomycete genomes.

Keywords: crinkler (CRN), effector, oomycete, plant-pathogen interactions, R package, RxLR

Secreted effector proteins have been reported for all major groups of plant pathogens, including bacteria, fungi, oomycetes, nematodes, and viruses (Jones and Dangl 2006; Toruño et al. 2016). Effector proteins are defined as secreted proteins that manipulate plant processes to the advantage of the parasite, in order to promote infection and generate disease (Pette and

Kamoun 2014). Effector proteins modulate and interfere with the normal physiology of the plant host in order to facilitate disease and infection (Jiang and Tyler 2012; Jones and Dangl 2006; Kamoun 2007). Effector proteins are common in bacterial plant pathogens. Up to 29 effector proteins are injected by the bacterial pathogen *Pseudomonas syringae* into the host plant cells via the type III secretion system (Chang et al. 2005). Effector discovery has predicted over 50 candidate effector proteins for the species *Heterodera glycines*, a plant-parasitic nematode that affects soybean crops across the world (Wang et al. 2001). Multiple effector proteins have been reported for a plethora of plant-pathogenic fungal and oomycete species (Selin et al. 2016).

The discovery of secreted effector proteins in these organisms was a major breakthrough, as these proteins are directly involved in pathogenicity. These effector proteins that disrupt the normal physiology of the plant can be recognized by specific resistance (*R*) genes coding for proteins (Jones and Dangl 2006). The recognition of an effector protein by an *R* gene leads to effector triggered immunity, which generates a signaling cascade that results in programmed cell death via a hypersensitive response. This programmed cell death slows the growth of the pathogen and avoids the proliferation of disease into neighboring plant cells (Jones and Dangl 2006).

Plant-pathogenic oomycetes or water molds are a group of highly devastating plant pathogens. The diseases caused by these organisms can affect hundreds of plant species and have led to high mortality of trees in forest ecosystems, losses of millions of dollars to agriculture, and have been implicated in the Irish potato famine (Fry 2008; Grünwald et al. 2008). Widely recognized plant-pathogen genera found in the water molds include *Phytophthora*, *Pythium*, *Albugo*, and *Peronospora* (Erwin and Ribeiro 1996).

Oomycetes contain a high number of predicted effector proteins (Birch et al. 2006; Kamoun 2006, 2007; Pais et al. 2013). These proteins are secreted by the pathogen haustorium and are translocated into the host plant cell, in which they are transported to different organelles to disrupt physiological functions and facilitate disease (Anderson et al. 2015; Birch et al. 2006; Kamoun 2006). Recent advances in molecular and computational biology have provided new information about the amino acid sequence of these oomycete effectors. Conserved motifs were identified in two key oomycete effector protein families, namely, the RxLR-dEER motif for RxLR effectors and the LFLAK-HVLV motif for crinkler (CRN) effectors. These canonical motifs are located near the N terminus

[†]Corresponding author: N. J. Grünwald; nik.grunwald@ars.usda.gov

Funding: This work was supported by funds from the United States Department of Agriculture Agricultural Research Service project 2072-22000-041-00-D and National Institute of Food and Agriculture project 2010-511001-21649. Mention of trade names or commercial products in this manuscript are solely for the purpose of providing specific information and do not imply recommendation or endorsement.

*The e-Xtra logo stands for “electronic extra” and indicates that supplementary text, one supplementary figure, and one supplementary table are published online.

The author(s) declare no conflict of interest.

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 2019.

of an amino acid sequence following the signal peptide sequence. To date, hundreds of these RxLR and CRN effectors per species have been predicted for important plant-pathogenic oomycetes from the genus *Phytophthora*, such as *Phytophthora infestans* (approximately 580 RxLR and 196 CRN effector genes [Haas et al. 2009]), *P. sojae* (approximately 470 RxLR and 100 CRN effector genes [Tyler et al. 2006]), and *P. ramorum* (approximately 260 RxLR and 19 CRN effector genes [Tyler et al. 2006]). In addition, the downy mildew pathogen *Hyaloperonospora arabidopsidis* contains up to 134 predicted RxLR-coding genes in its genome (Baxter et al. 2010). However, not all oomycete species contain evidence of RxLR-coding genes in their genomes. No RxLR effector proteins have, to date, been found in the genera *Pythium*, *Saprolegnia*, or *Albugo* (Lévesque et al. 2010; Links et al. 2011; Win et al. 2012).

The process of identifying candidate effector-coding genes in sequenced genomes has, to date, been ad hoc and not reproducible, due to the fact that each user can choose slightly varying thresholds. The most common bioinformatic process used for predicting effectors was described by Hass et al. (2009), in which the authors used a combination of pattern matching via regular expressions (REGEX) based on the canonical RxLR and LFLAK motifs, followed by homology searches using Markov models. This method was used for the prediction of effector proteins in *Phytophthora infestans*, resulting in a total of 562 predicted RxLR effectors and 196 CRN effectors (Hass et al. 2009). The approach used by Hass et al. (2009) has been modified to include predicted effector sequences from other species, in order to improve the homology search (Govers and Bouwmeester 2008). Several independent tools are available to perform different steps to recognize effectors (Dalio et al. 2018; Sonah et al. 2016), but, to date, no bioinformatic tool allows the prediction of oomycete effectors in a simple and fast manner.

We developed the expandable *effectR* R package designed to predict effector proteins in a fast and reproducible way. This package uses both REGEX and homology search mechanisms based on hidden Markov models (HMM) to predict candidate effector proteins, using gene models obtained from whole-genome sequences. *effectR* provides functions in the statistical and computer language R (R Core Team 2018), which allows rapid identification and evaluation of candidate effector proteins for any oomycete species with a sequenced genome. The *effectR* package has been developed to facilitate the prediction of effector proteins for researchers with limited expertise in computational biology. We tested the *effectR* package using available oomycete genomes and successfully validated effector prediction for the species *Phytophthora infestans*. The *effectR* package is modular and can be expanded to predict effector proteins with different canonical motifs specified by the user. In fact, we encourage contributions to our GitHub repository of new functionality for effector prediction for other organismal groups. We included an example of a custom

implementation for bacterial proteins that contain PAAR repeats. The *effectR* package is released on CRAN and a brief user tutorial is provided on GitHub.

RESULTS AND DISCUSSION

We developed the R package *effectR* that provides a convenient tool for rapid prediction of effectors in oomycete genomes or with custom scripts for any genome in a reproducible way. The *effectR* package relies on a combination of REGEX statements and HMM approaches to predict candidate RxLR and CRN effectors using several function calls (Table 1). Three steps are required to call effectors: i) A REGEX search to select amino acid sequences translated from all the open reading frames (ORFs) in a genome that match the motifs of interest, ii) a second, broad search of the amino acid sequences that match a Markov chain profile created with the amino acids that matched the motifs of interest in the REGEX step, and iii) a posthoc set of tools combining the results from the two previous steps after filtering for redundancies (Fig. 1). We are using the oomycete *P. infestans* as an example application.

Step 1: Obtaining candidate RxLR and CRN effectors using REGEX.

The following example uses the function `regex.search()` with the test dataset `test_infestans.fasta`. This dataset contains a subset of 28 sequences, including six reference RxLR effectors (PexRD36, PexRD1, ipi01/Avrblb1, Avr1, Avr4, and Avr3a) (Haas et al. 2009), nine randomly selected predicted CRN effector proteins identified by Hass et al. (2009), and 14 in-house translated ORFs from the *P. infestans* reference genome sequence, of which eight translated ORFs contain RxLR and EER motifs. This test dataset is included in the *effectR* package and can be loaded as follows:

```
# Loading the effectR package in R
library("effectR")

# Using the read.fasta function of the seqinr package to
import the translated ORF FASTA file
fasta.file <- system.file
("extdata", "test_infestans.fasta",
 package = "effectR")
ORF <- seqinr::read.fasta(fasta.file)

# Verifying if the length of the loaded FASTA file
correspond to the 28 translated ORF's from P. infestans
length(ORF)

## [1] 28

# Executing step 1: Prediction of RxLR effectors from the
ORF object. Results are saved in the rxlr.cand object
rxlr.cand <- regex.search(ORF, motif = "RxLR")

# What is the number of RxLR effectors predicted by step
1?
length(rxlr.cand)

## [1] 15
```

This code snippet illustrates that `regex.search()` was able to predict all 14 expected RxLR candidates from the test dataset along with one additional candidate out of the 28 reference effectors and translated ORFs from the included `test_infestans.fasta` dataset. In addition, the *effectR* package can also predict CRN effectors or predict other families of interest based on custom REGEX provided by the end user. Changing the `motif="RxLR"` option to "CRN" or "custom" will

Table 1. Functions found in *effectR* and their descriptions

Function	Description
<code>effector.summary()</code>	Returns nonredundant sequences from <code>hmm.search()</code> or <code>regex.search()</code> and generates a motif table
<code>hmm.logo()</code>	Plots the relative frequencies of each position for the <code>hmmsearch()</code> table
<code>hmm.search()</code>	Searching for motifs using hidden Markov model searches
<code>regex.search()</code>	Searching for motifs using regular expressions
<code>shiny.effectR()</code>	A function to run <i>effectR</i> functions in a GUI using the shiny app

allow the prediction of these other motifs of interest without reloading the ORF dataset, thus significantly reducing processing time. Using the “CRN” option successfully predicts the nine CRN candidate effector proteins from Hass et al. (2009) included in the test dataset. Users are also encouraged to submit new functions to the *effectR* GitHub repository for prediction of other motifs for any organismal group (discussed below and in Supplementary Text S1 for an example).

Step 2: Using HMMs to predict additional candidate effectors.

The following code example uses the function “*custom*” on the *test_infestans.fasta* dataset after obtaining the REGEX RxLR candidates.

```
# Loading the effectR package in R
library("effectR")

# Using the read.fasta function of the seqinr package to
import the translated ORF FASTA file
fasta.file <- system.file
("extdata", "test_infestans.fasta",
 package = "effectR")
ORF <- seqinr::read.fasta(fasta.file)

# Step 1 prediction
REGEX <- regex.search(ORF, motif = "RxLR")

# Expanding the search of RxLR effectors using HMM
searches (step 2). All candidate effectors
predicted by step 2 will be saved in the
candidate.rxlr object
```

```
candidate.rxlr <- hmm.search(original.seq =
fasta.file, regex.seq = REGEX)

## Starting MAFFT alignment.
## ---
## Executing MAFFT
## Please be patient
## MAFFT alignment finished!
## Starting HMM
## ---
## Creating HMM profile

[additional console output was removed for brevity]

## HMM search done.
## ---
##
## Total of sequences found in REGEX: 15
## Total of sequences found in HMM: 17
## Total of redundant hits: 15
## Number of effector candidates: 17
```

The previous code snippet shows that *hmm.search()* was able to identify two new RxLR candidates from the *test_infestans.fasta* dataset, increasing our total number of candidate RxLR effectors to 17. The *hmm.search()* function also detects candidate effectors previously predicted by the REGEX step. In the previous example, the *hmm.search()* function detected 15 candidate effectors using the *hmm.search()* step. The combined 17 candidate effectors based on the REGEX and HMM searches can now be evaluated in step 3.

Fig. 1. Workflow representing the steps used in the *effectR* R package. The package uses an amino acid FASTA file as input that contains either all translated gene models from a sequenced genome or the six-frame translation open reading frames (ORF) of a genome to obtain effectors. The genomic ORFs represent six-frame translations of a genome, using the external software *getorf* from EMBOSS. After reading the amino acid FASTA input file into R, the *effectR* package predicts all effector proteins using three steps: i) Searching MOTIF patterns via regular expression (REGEX), ii) searching for additional effectors using a homology search based on hidden Markov models (HMM) using *hmmer*, and iii) a posthoc analysis of the candidate effectors to visualize and summarize the motifs of the candidates obtained by the package. All steps are modular and can be modified for discovery of any kind of effector.

Step 3: Post hoc tools for curation of candidate effector genes.

In the following steps, *effectR* provides functions to summarize and visualize candidate effector genes.

Step 3a: Summarizing the predicted effector proteins and determining the number and position of each motif in the effector sequences. The *effectR* package includes the `effector.summary()` function that combines the results from both REGEX and HMM searches into a list of unique predicted effectors that can be exported into multiple sequence format files. The `effector.summary()` function returns a table of the number of motifs per sequence and the position of the first residue of each motif of interest for all candidate effectors. In addition, the table includes a summary motif column that will define the candidate sequence as complete (includes both motifs of interest), only X motif (if the sequence includes only one of the two motifs of interest), or no motifs (no motifs of interest found).

```
# Loading the effectR package in R
library("effectR")

# Using the read.fasta function of the seqinr package to
import the translated ORF FASTA file
fasta.file <- system.file
("extdata", "test_infestans.fasta",
 package = "effectR")
ORF <- seqinr::read.fasta(fasta.file)

# Step 1 prediction
REGEX <- regex.search(ORF, motif = "RxLR")

# Expanding the search of RxLR effectors using HMM
searches (step 2). All candidate effectors predicted by
step 2 will be saved in the candidate.rxlr object
candidate.rxlr <- hmm.search(original.seq =
fasta.file, regex.seq = REGEX)

# Summarizing the predictions from step 1 and step 2 using
the effectR.summary() function (Step 3a).

# The summary of non-redundant RxLR predicted effectors
will be stored in the RxLR.effector$consensus.
sequences objects

# The table of motif number and position will be stored in
the RxLR.effectors$motif.table object
RxLR.effectors <- effector.summary(candidate.rxlr,
motif = "RxLR")

# What is the number of non-redundant RxLR effectors
predicted by step 1 and step 2?
length(RxLR.effectors$consensus.sequences)

## [1] 17

# Table of the first five motif number and position for
all non-redundant, predicted RxLR effector proteins
head(RxLR.effectors$motif.table, n = 5)

##Sequence ID    RxLR number RxLR position EER number
EER position MOTIF Length
PITG_21388_ipi01/Avrblb1_RxLR    2    51,122    1    70
Complete    152
PITG_23132_PexRD36_RxLR        1        28    1    55
Complete    76
```

```
PITG_15287_PexRD1_RxLR        1    50    1    75
Complete    213
PITG_07387_Avr4_RxLR          1    42    1    53
Complete    287
PITG_14371_Avr3a_RxLR          1    44    1    57
Complete    147
```

Step 3b: Plotting the HMM profile. To visualize the position of the motifs of interest, we include the function `hmm.logo()` that plots the results from the HMM profile obtained in *hmmbuild* in a logo plot (Schneider and Stephens 1990), using the *ggplot2* package (Wickham 2016). The `hmm.logo()` function reads the *hmm* profile table and extracts a bit score of each amino acid at each position. Then, *effectR* plots the bit score on the y axis, the amino acid position in the x axis, and overlays the amino acid with the highest bit score over the plot in order to represent the frequency for each amino acid found at every position of the consensus sequence from the multiple sequence alignment (MSA) step. Here is an example of the `hmm.logo()` function for the *P. infestans* example dataset:

```
# Loading the effectR package in R
library("effectR")

# Using the read.fasta function of the seqinr package to
import the translated ORF FASTA file
fasta.file <- system.file
("extdata", "test_infestans.fasta",
 package = "effectR")
ORF <- seqinr::read.fasta(fasta.file)

# Step 1 prediction
REGEX <- regex.search(ORF, motif = "RxLR")

# Expanding the search of RxLR effectors using HMM
searches (step 2). All candidate effectors predicted by
step 2 will be saved in the candidate.rxlr object
candidate.rxlr <- hmm.search(original.seq =
fasta.file, regex.seq = REGEX)

# Plotting the HMM profile created in the hmm.search()
function using the hmm.logo() function
hmm.logo(candidate.rxlr$HMM_Table)

## R graphical output [NOTE: NOT SHOWN HERE, BUT PRINTED
TO CONSOLE]:
```

Validation.

To test the developed package, we used a set of four sequenced genomes of three oomycete species and one fungal species to predict RxLR effectors. We used the genomes of the oomycete species *Phytophthora infestans* T30-4 version 1 (Haas et al. 2009), *Pythium ultimum* DAOM BR144 version 1 (Lévesque et al. 2010), and *Albugo candida* Ac2VRR version 1 (Links et al. 2011). *P. infestans* has more than 580 predicted RxLR effectors (Haas et al. 2009), while neither *Pythium ultimum* nor *A. candida* have any reported RxLR motifs. Instead, *A. candida* has 26 predicted gene models with a variant RxLR motif called Ac-RXL (Links et al. 2011). The genome of the ascomycete *Fusarium oxysporum* f. sp. *lycopersici* (Ma et al. 2010) was used as a negative control or outgroup with no expectation of finding RxLR motives.

The assembled reference genomes of each of the four species were downloaded from Fungi-DB (Stajich et al. 2012). For each genome assembly, a six-frame translation of all ORFs was predicted using *getorf* from EMBOSS (Rice et al. 2000). Translated ORFs with a length of less than 100 nucleotides were discarded, as none of the reported functional effector proteins in oomycete species are shorter than this size. We

predicted all RxLR effector proteins for each translated sequence from the genome assembly in the *effectR* package. Additionally, prediction of the signal peptide using all predicted candidate effectors was performed in SignalP 3.0 (Bendtsen et al. 2004), as recommended by Sperschneider et al. (2015). A threshold D score > 0.8 was used. Any candidate effector with a predicted signal peptide was considered a high-confidence candidate effector. All high-confidence candidate effector proteins from *P. infestans* were compared with the published list of effectors from Haas et al. (2009). We created a custom blast database using the RxLR effector proteins and identified matches between the amino acid sequences of the *effectR* predictions against the RxLR candidate database in Blastp (Altschul et al. 1990).

As a first step, we identified a total of 174 proteins for *Pythium ultimum*, 47 for *A. candida*, and three for *F. oxysporum* f. sp. *lycopersici* that showed evidence for one or both the RxLR and EER motifs. However, in the second step, none of these proteins were considered high-confidence candidate effectors, as they lacked a signal peptide cleavage site (Table 2). These results are consistent with previous reports of finding no RxLR effector proteins in genomes of these oomycete and fungal species. The Ac-RXL effector proteins of *A. candida* have an independent origin from the RxLR effector proteins found in *Phytophthora* species (Links et al. 2011), and no functional RxLR effector proteins have been reported for *F. oxysporum* f. sp. *lycopersici* (Ellis et al. 2009; Ma et al. 2010).

For *P. infestans*, we predicted 395 candidate effector proteins in the REGEX step and 827 candidate effector proteins in the HMM step, for a total of 900 nonredundant RxLR effector proteins (a number similar to the 831 RxLR effectors predicted by Haas et al. [2009], using only a combination of REGEX and HMM methods). In contrast to the predictions in the previously screened fungal and oomycete species, a high number of high-confidence candidate effectors (631 high-confidence candidate effectors with evidence of a signal peptide cleavage site) were predicted for the translated sequences of *P. infestans* (Table 2). The logo plot shows the prevalence of the RxLR-EER motifs obtained by the REGEX step between residues 49 and 67 (Fig. 2). The high number of effector proteins predicted in the HMM step is a result of the low thresholds used by our package in order to obtain as many candidate effectors as possible. Of the 631 high-confidence candidate effector proteins, we predicted 453 RxLR effectors with a Blastp match of more than 90% identity with the Haas et al. (2009) RxLR predicted effector proteins. The genomic position of each of these 453 predicted proteins also corresponded with the position of the homologous effector protein reported by Haas et al. (2009), indicating that *effectR* successfully predicted previously known RxLR effector proteins from *P. infestans*. The functionally validated effector proteins PexRD36, PexRD1, ipi01/Avrblb1, Avr4, and Avr3a were correctly predicted by the *effectR* package, providing a successful positive control. The prediction of these 631 effector

Table 2. Predicted RxLR effector proteins in *effectR* for the open reading frame translations (ORF) of the assembled genomes of *Phytophthora infestans*, *Pythium ultimum*, *Albugo candida*, and *Fusarium oxysporum* f. sp. *lycopersici*^a

Species	No. predicted ORFs	No. with RxLR/EER motifs ^b			No. with a signal peptide cleavage site ^c
		REGEX	HMM	Total	
<i>Phytophthora infestans</i>	208,272	395	827	900	631
<i>Pythium ultimum</i>	63,882	61	149	174	0
<i>Albugo candida</i>	24,405	13	34	47	0
<i>Fusarium oxysporum</i> f. sp. <i>lycopersici</i>	55,224	3	0	3	0

^a Total number of predicted effector proteins at each step of the package (step 1: REGEX [regular expressions] search; step 2: HMM [hidden Markov model] search) and the total number of unique effector proteins (intersection of step 1 with step 2). If the candidate protein predicted in *effectR* does not have a signal peptide, then the predicted protein is not considered a high-confidence candidate effector protein.

^b Number of proteins with evidence of RxLR/EER motifs predicted by *effectR*.

^c Number of RxLR effectors proteins with a signal peptide cleavage site.

Fig. 2. Sequence logo plot of the candidate RxLRs obtained for *Phytophthora infestans* in the hidden Markov model profile built from the sequences obtained in the REGEX step (step 1). The size of the letters is proportional to the height of the bar and reflects the relative frequency of the particular amino acid. The plot also shows the high relative frequency of the RxLR-EER motifs around the 50th residue of the amino acid sequence.

proteins was performed in under 20 min on a laptop computer using two cores and less than 1 Gb of RAM. This is a modest processing time for a FASTA file with more than 200,000 translated ORFs.

A total of 99 of the RxLR effectors listed by Hass et al. (2009) were not predicted by *effectR*. Of these 99 RxLR effector proteins, we find 18 proteins with no presence of either a RxLR or EER domain, 24 proteins with no evidence of an EER domain, and the remainder 57 proteins with the RxLR or EER domain with higher upstream or downstream distances from the REGEX expected residue position. These results are expected, as Haas et al. (2009) used a combination of Blastp searches, TribeMCL clustering, and homology searches from reference effectors in addition to the REGEX + HMM searches, leading to a higher number of candidate effector proteins than expected just from the searches implemented in this package. *effectR* also predicted 79 candidate effector proteins not present in Hass et al. (2009) that include evidence of a signal peptide cleavage site (Table 3). Of these 79 newly predicted effector proteins, 40 proteins contained both RxLR and EER domains, 14 only contained the RxLR motif, 18 only had the EER motif, and seven proteins did not include any of the motifs of interest (Table 3).

In addition to the prediction of RxLR effector proteins, we predicted the CRN effector proteins for the ORF translations of the species used in the proof of concept (Supplementary Table 1). The results of the CRN prediction are consistent with the expectation of a high number of CRN effector proteins present in the genomes of oomycete species, with 21 candidate CRN effectors predicted for *Pythium ultimum* and 214 CRN effectors predicted for *P. infestans*. Of the 214 CRN effectors of *P. infestans* predicted by *effectR*, we found 159 CRN effector proteins with a Blastp match of more than 90% identity with the Haas et al. (2009) CRN predicted effector proteins. The number of predicted effectors for these two species is similar to the reported number of CRN effectors (196 CRN effectors for *P. infestans* and 26 CRN effectors for *Pythium ultimum*) (Lévesque et al. 2010). However, only four CRN effectors in *A. candida* were predicted. These four predicted CRN proteins for *A. candida* only have the LxLAK motif, as reported by Links et al. (2011), and are not considered canonical CRN effectors (Links et al. 2011). Finally, only three CRN effectors were predicted for the fungal species *Fusarium oxysporum* f. sp. *lycopersici*. No CRN effectors have been reported for this fungal pathogen, and the CRN effectors predicted by *effectR* only contain the LxLAK motif, indicating that these proteins are not canonical CRN effectors, as in *A. candida*.

Custom scripts for other effectors.

effectR can easily be modified for identifying candidate effectors. For example, `regex.search()` can be modified to call bacterial proteins that contain PAAR repeats. These proteins are associated with the VgrG-like spikes found in the type VI secretion system of bacteria and have been shown to be essential in target cell killing by the bacterial species *Vibrio cholerae* and *Acitenobacter baylyi* (Shneider et al. 2013). The PAAR proteins have a homonymous amino acid sequence motif (PAAR) with one or more repeats. We created an example showing how the *effectR* can predict potential candidate proteins in the proteome of the reference strain ATCC 39315 of *V. cholerae* (Heidelberg et al. 2000) by using the PAAR motif as part of the REGEX search. Our results indicate that *effectR* successfully predicted 19 candidate PAAR proteins, two of which are homologous to the previously reported PAAR proteins. These two predicted proteins only differ in one amino acid when compared with PAAR homolog proteins previously described by Shneider et al. (2013) from other strains of *V. cholerae* (Supplementary Fig. 1). These results show that

effectR can correctly predict different proteins with other motifs than the canonical oomycete effectors and, also, illustrate the importance of using manual curation based on homologous proteins to avoid the detection of false positives.

Conclusions.

Our *effectR* package provides a novel tool for reproducible prediction of candidate effectors in oomycete genomes. The package is modular, in which every step can be modified to predict candidate effector proteins. Custom motifs for any new effector family can easily be added via custom scripts or by contributing to the GitHub repository. The package has been successfully tested for translated ORFs predicted from genomes of oomycete plant pathogens and can be used for a quick survey of the effector arsenal involved in any plant-pathogen interactions for any species of interest in which effector motifs are known. However, *effectR* only identifies candidate effectors, and further functional validation in the wet lab is needed.

MATERIALS AND METHODS

The *effectR* package.

The *effectR* package is written in the R computer language (R Core Team 2018). *effectR* allows prediction of oomycete effector proteins. The package requires as input a FASTA file, ideally containing all six-frame amino acid translations for each ORF of the sequenced genome of interest or, at a minimum, all translated gene models. The package then returns the total number of predicted effectors, the amino acid sequence for each of the predicted effectors, the number and position of the motifs of interest for each predicted effector, and the Markov chain profile table, which can be conveniently visualized.

External programs required to execute *effectR*.

The *effectR* package requires installation of additional programs and the recommended use of external software to assure its correct functionality (reference resources provided on GitHub for instructions). *effectR* can use amino acid translations of the gene models predicted in a genome of interest, but we recommend the use of six-frame translations of all ORFs from a genome assembly. Including all six translations for each ORF in a genome will allow prediction of more candidate effectors present in an organism of interest. To generate six-frame ORF translations, we use *getorf* from EMBOSS (Rice et al. 2000). *getorf* can be run locally on the command line or online via the EMBOSS explorer (Rice et al. 2000). Other additional programs required by *effectR* are the MAFFT v7 MSA tool (Katoh and Standley 2013) and HMMER 3.1b2 (Eddy 2011). MAFFT performs a sequence alignment of the candidate effectors in order to build a reference profile to be scanned by HMMER. HMMER executes the searches based on HMMs (Fig. 1, step 2). These two programs are external to R and must be installed on the same machine as *effectR*. The *effectR* package includes functions to detect if both HMMER and MAFFT are available in the default user path or the user can specify the location of each of the binaries when executing *effectR*.

While these additional programs are indispensable to complete step 2 of the *effectR* package, step 1 and step 3 can be executed independently of them. The user can execute step 1 within *effectR*, use the output from step 1 and the original ORF file to perform a MSA using other tools of interest, and execute HMM searches externally. The *effectR* package can import external MSA in FASTA format and HMM results in table format to be used in step 3. The inclusion of step 2 as part of the package was created for a more streamlined process in which the user can identify candidates using REGEX searches, broaden the number of candidate proteins via HMM searches,

Table 3. New candidate RxLR effector proteins ($n = 79$) from *Phytophthora infestans* predicted by the *effectR* package^a

Scaffold	Start	End	RxLR motifs		EER motifs		Motifs present
			Number	Position	Number	Position	
Supercontig_1.2	387,181	387,789	0	NA ^b	0	NA	No motifs
Supercontig_1.2	5,134,349	5,134,744	1	35	1	51	Complete ^c
Supercontig_1.3	4,668,546	4,669,019	1	43	1	66	Complete
Supercontig_1.4	1,947,482	1,947,706	0	NA	0	NA	No motifs
Supercontig_1.5	2,020,297	2,020,617	1	40	0	NA	Only RxLR
Supercontig_1.5	3,399,390	3,400,325	0	NA	1	47	Only EER
Supercontig_1.5	1,365,830	1,366,261	1	54	1	81	Complete
Supercontig_1.5	3,296,675	3,297,394	1	31	1	44	Complete
Supercontig_1.5	1,695,450	1,695,806	1	43	1	56	Complete
Supercontig_1.7	778,069	778,482	1	50	0	NA	Only RxLR
Supercontig_1.8	2,694,681	2,694,974	0	NA	0	NA	No motifs
Supercontig_1.9	417,751	418,602	1	47	0	NA	Only RxLR
Supercontig_1.10	726,219	726,617	0	NA	1	59	Only EER
Supercontig_1.11	1,535,262	1,536,773	0	NA	0	NA	No motifs
Supercontig_1.11	1,187,840	1,188,091	1	50	1	63	Complete
Supercontig_1.11	1,786,307	1,786,816	1	50	1	60	Complete
Supercontig_1.11	529,720	529,944	1	41	0	NA	Only RxLR
Supercontig_1.11	1,902,450	1,903,310	1	53	1	61	Complete
Supercontig_1.13	612,373	612,726	1	50	1	63	Complete
Supercontig_1.14	2,070,861	2,072,003	2	57,322	1	73	Complete
Supercontig_1.15	1,593,048	1,593,419	0	NA	1	62	Only EER
Supercontig_1.15	668,809	669,123	1	23	1	45	Complete
Supercontig_1.15	2,682,075	2,682,359	1	49	0	NA	Only RxLR
Supercontig_1.16	1,715,898	1,716,353	1	65	0	NA	Only RxLR
Supercontig_1.16	1,567,608	1,568,108	1	57	1	76	Complete
Supercontig_1.17	448,981	449,679	1	47	1	83	Complete
Supercontig_1.18	806,113	806,700	1	43	1	51	Complete
Supercontig_1.20	1,412,054	1,413,028	0	NA	1	62	Only EER
Supercontig_1.20	1,339,618	1,340,277	1	52	0	NA	Only RxLR
Supercontig_1.25	1,219,557	1,220,138	0	NA	1	60	Only EER
Supercontig_1.25	1,023,691	1,023,990	0	NA	1	46	Only EER
Supercontig_1.25	1,314,779	1,315,240	1	42	1	56	Complete
Supercontig_1.25	1,633,314	1,633,970	1	48	1	68	Complete
Supercontig_1.26	1,717,796	1,718,344	0	NA	0	NA	No motifs
Supercontig_1.27	816,842	817,168	1	50	1	67	Complete
Supercontig_1.27	638,168	638,521	1	51	1	65	Complete
Supercontig_1.29	807,454	807,804	0	NA	1	56	Only EER
Supercontig_1.29	1,420,380	1,420,796	1	42	1	57	Complete
Supercontig_1.32	40,829	41,050	1	47	1	64	Complete
Supercontig_1.33	857,150	858,472	1	40	1	57	Complete
Supercontig_1.33	1,310,346	1,310,840	1	56	0	NA	Only RxLR
Supercontig_1.33	1,324,452	1,324,946	1	40	1	57	Complete
Supercontig_1.34	80,145	80,615	0	NA	1	64	Only EER
Supercontig_1.34	708,534	709,043	1	47	1	66	Complete
Supercontig_1.35	875,543	876,028	1	46	0	NA	Only RxLR
Supercontig_1.38	848,659	849,336	0	NA	0	NA	No motifs
Supercontig_1.38	794,856	797,288	1	216	1	46	Complete
Supercontig_1.39	1,376,318	1,376,599	1	50	1	63	Complete
Supercontig_1.40	730,075	730,683	0	NA	1	62	Only EER
Supercontig_1.42	695,078	695,443	1	30	0	NA	Only RxLR
Supercontig_1.43	780,095	780,469	1	50	1	69	Complete
Supercontig_1.43	1,272,359	1,272,643	1	53	1	63	Complete
Supercontig_1.43	1,286,094	1,286,504	1	53	1	69	Complete
Supercontig_1.44	865,765	866,316	0	NA	1	76	Only EER
Supercontig_1.44	1,006,809	1,007,273	1	39	1	56	Complete
Supercontig_1.47	1,060,823	1,061,302	1	54	1	74	Complete
Supercontig_1.63	438,688	438,969	0	NA	1	69	Only EER
Supercontig_1.63	549,263	549,847	0	NA	1	68	Only EER
Supercontig_1.66	372,213	372,752	0	NA	1	76	Only EER
Supercontig_1.66	259,626	260,069	1	44	1	56	Complete
Supercontig_1.67	516,713	517,306	0	NA	1	64	Only EER
Supercontig_1.79	318,551	319,021	1	51	0	NA	Only RxLR
Supercontig_1.89	375,480	376,355	1	35	0	NA	Only RxLR
Supercontig_1.89	370,258	370,542	1	52	1	60	Complete
Supercontig_1.124	94,888	95,385	1	47	1	60	Complete
Supercontig_1.129	11,438	16,690	1	53	1	67	Complete
Supercontig_1.130	54,647	54,997	0	NA	1	56	Only EER
Supercontig_1.141	177,814	179,079	0	NA	1	52	Only EER
Supercontig_1.220	51,445	51,846	0	NA	1	61	Only EER
Supercontig_1.231	92,953	93,969	1	38	1	49	Complete
Supercontig_1.235	17,662	20,670	1	52	1	71	Complete
Supercontig_1.241	140,057	140,608	0	NA	1	60	Only EER
Supercontig_1.241	90,165	90,533	1	53	1	67	Complete
Supercontig_1.302	59,522	59,977	2	51,122	1	70	Complete
Supercontig_1.385	9,167	10,102	1	216	1	46	Complete
Supercontig_1.709	2,672	2,869	0	NA	0	NA	No motifs
Supercontig_1.1192	4,030	5,904	3	51,77,310	0	NA	Only RxLR
Supercontig_1.2669	2,032	2,460	1	27	1	43	Complete
Supercontig_1.3033	1,747	2,109	2	74,80	0	NA	Only RxLR

^a None of these candidate proteins were predicted by Haas et al. (2009). The candidate effector proteins are sorted by scaffold and start position.^b NA = not applicable.^c Complete = candidate RxLR effector proteins with both RxLR and EER domains present in the amino acid sequence.

and summarize and quality control the results obtained from the *effectR* package in a fast and seamless manner.

Obtaining candidate RxLR and CRN effectors using REGEX.

To predict the first set of candidate effector proteins, *effectR* searches the ORF translation file to find sequences that match the motifs of interest (Fig. 1). These searches are based on REGEX matching. For the RxLR motif, the REGEX reported by Haas et al. (2009) is used:

```
^\w{10,40}\w{1,96}R\wLR\w{1,40}EER.
```

This REGEX example shows the three parts used in step 1 to identify an RxLR effector candidate (Fig. 3). Part 1A ($\w{10,40}$) reserves the first 10 to 40 positions of the amino acid sequence for the signal peptide. The *effectR* package does

not predict the cleavage site of a signal peptide and relies on external programs to predict the presence of said structural motif. Part 1B ($\w{1,96}R\wLR$) searches for amino acid residues that match the RxLR motif within the following 1 to 96 residues after the signal peptide. Part 1C ($\w{1,40}EER$) searches for the EER motif within the 40 residues following the RxLR motif. Note, that this is modified from Haas et al. (2009), who used $[ED][ED][KR]$ in order to simplify the initial REGEX search. If the user wants to use the original Haas et al. (2009) motif as part of the REGEX search, the `custom` option at the `regex.search` step can be modified in the following manner: `regex.search(seq=ORF, motif = "custom", reg.pat = "^\w{10,40}\w{1,96}R\wLR\w{1,40}[ED][ED][RK]" )`. In addition, to identifying effectors with the canonical W-Y-L motif found in RxLR proteins (Win et al. 2012), the following custom

Fig. 3. Graphical representation of the regular expression (REGEX) used by *effectR* to predict candidate RxLR effectors. This REGEX is integrated in the `hmmsearch()` function of *effectR*. The REGEX search is divided into three parts (gray boxes): Part 1 reserves the space for the signal peptide; part 2 performs a downstream search for the RxLR motif; and part 3 searches for the EER motif downstream from the RxLR motif. If a query amino acid sequence has a match to the two motifs, the sequence is considered a candidate RxLR effector protein and is used to build the hidden Markov model profile for step 2 of the package.

The de novo Effector Pipeline online server: RxLR Module

The RxLR de novo effector pipeline calls effectors for oomycete genomes in a fast and reproducible way.

Instructions and guidelines:

Please read the instructions very carefully. Follow each step and be patient, the processes might take some time to start. **CLICK ONLY ONE IN EACH BUTTON.**

To obtain a better assessment of the effectors in your genome please use a 6-frame translation of the genome of interest (Use [getorf](#) to obtain the 6-frame translation and use the output file as input in this pipeline.)

Example file: Subset of Twenty-five sequences of *Phytophthora infestans* obtained from [FungiDB](#) and processed in [getorf](#)

[Example dataset](#)

Upload FASTA formatted files (.fna,.fasta,.fas,.orf).

[Browse...](#) No file selected

Step 1

Regular expression search

Step 1 uses a regular expression search to identify the sequences with the motifs of interest (RxLR+EER as based on Haas et al., 2009).

Select motif to search

RxLR

[Do REGEX search](#)

[Download REGEX candidates](#)

Download will be available once the processing is completed.

Step 2

HMM search

Step 2 aligns the candidate genes obtained in Step 1 to create a Hmmer search to identify additional potential effectors.

[Align and do HMM search](#)

[Download HMM files](#)

Download will be available once the processing is completed.

Fig. 4. Screen capture of the Shiny R web application for *effectR*. All functions available in the package are included in this web application. The web application can be initiated within an active R session by running the `shiny.effectR()` function.

script can be implemented: `regex.search(seq=ORF, motif="custom", reg.pat = "^\\w{10,40}\\w{1,96}R\\wLR\\w{1,40}[ED][ED][RK]\\w{1-20}[WYL]")`.

A limitation of *effectR* is that it does not directly predict the presence of a signal peptide. Other programs (i.e., SignalP version 3.0 [Bendtsen et al. 2004]), rather than higher versions, as recommended by Sperschneider et al. (2015) for oomycetes, can be used to predict the presence of the signal peptide.

Step 1 is summarized in the `regex.search()` function of the *effectR* package. Step 1 determines if a translated ORF is a candidate effector after identifying the presence of a RxLR/CRN motif via REGEX (RxLR motif patterns: `^\\w{10,40}\\w{1,96}R\\wLR\\w{1,40}EER`; CRN motif pattern: `^\\w{1,90}LFLAK\\w+` [Haas et al. 2009]). In addition, we have added a custom option that allows user specification of a custom REGEX, in order to identify candidate genes for other effectors of interest.

Using HMMs to predict additional candidate effectors.

A second independent method of identifying candidate effectors relies on HMM (Fig. 1, step 2). This search allows the identification of additional sequences that match a HMM profile. In *effectR*, the HMM profile is built using the candidate effectors predicted in the REGEX step. The HMM profile includes the probabilities of any amino acid occurring in a given position of the consensus sequence. The consensus sequence is the product of a MSA of the sequences used to build the HMM profile (Eddy 1998).

To create the HMM profile, *effectR* aligns the candidate effectors to identify common motifs and builds the HMM profile based on these common motifs. *effectR* uses MAFFT (Katoh and Standley 2013) to calculate the MSA of the candidate effectors predicted from the REGEX step. The package *effectR* uses the E-INS-i iterative refinement algorithm for creating the alignment. The E-INS-i algorithm is suitable for sequences containing common domains of interest flanked by large unalignable regions (Katoh and Standley 2013). These large, unalignable regions are typically observed in RxLR and CRN motifs. After generating the MSA from REGEX candidate effectors, *effectR* creates a HMM profile using HMMER `hmmbuild` and `hmmsearch` modules. After the HMM profile is built, *effectR* searches the original ORF file for sequences with hits against the HMM profile, using `hmmsearch`. To encourage the user to perform manual curation, *effectR* does not apply any significance thresholds in the `hmmsearch` step and returns all the sequences that match the HMM profile to the user by default. However, we have included an option within `hmm.search`, called `hmm.thres`, that establishes a threshold of the bit score cutoff to return candidate effector proteins. The `hmm.thres` option is equivalent to the `-T` option used as bit score cutoff for the per-sequence ranked hit `hmmsearch` program. The result from the HMM search is a list of translated ORFs that match the HMM profile. This list can be used for manual curation steps (Fig. 1, step 3) or can be exported using the `write.fasta` function of the `seqinr` package (Charif and Lobry 2007). This step has been summarized in the `hmm.search()` function of the *effectR* package.

Post hoc tools for combining candidate effector genes from REGEX and HMM searches.

The *effectR* package allows for manual curation of the predicted effector proteins to screen for highly heterogeneous sequences or for sequences that match the HMM profile but do not include any of the motifs of interest in their amino acid sequence. The manual curation functions created for the *effectR* package include the `effector.summary()` function. This function summarizes the results from steps 1 and 2

and generates a list of unique candidate effectors. The `effector.summary()` function also returns the numbers of motifs and position in the amino acid sequence for all predicted candidate effectors. Finally, we created the `hmm.logo()` function to plot logo-like figures based on the candidate effectors predicted in steps 1 and 2. Each step is modular and fully compatible with the results of either REGEX or HMM searching.

Software availability.

The *effectR* package can be downloaded from CRAN or GitHub. *effectR* can be used on the command line or can be deployed in a user-friendly, point-and-click graphical interface, via the shiny R framework (Chang et al. 2017), using the `shiny.effectR()` function (Fig. 4). The end user must install the packages *shiny* (Chang et al. 2017) and *shinyjs* (Attali 2018) in order to deploy the shiny R graphic interface.

ACKNOWLEDGMENTS

We thank B. Knaus, Z. Kamvar, and Z. Foster for their technical advice, extensive code review, and vast expertise in R programming.

AUTHOR-RECOMMENDED INTERNET RESOURCE

GitHub *effectR* page: <https://github.com/grunwaldlab/effectR>

LITERATURE CITED

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Anderson, R. G., Deb, D., Fedkenheuer, K., and McDowell, J. M. 2015. Recent progress in RXLR effector research. *Mol. Plant-Microbe Interact* 28:1063-1072.
- Attali, D. 2018. shinyjs: Easily improve the user experience of your shiny apps in seconds. R package version 1.0. CRAN Project. Published online. <https://cran.r-project.org/web/packages/shinyjs/index.html>
- Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A., Anderson, R., Badejoko, W., Bittner-Eddy, P., Boore, J. L., Chibucos, M. C., Coates, M., Dehal, P., Delehaunty, K., Dong, S., Downton, P., Dumas, B., Fabro, G., Fronick, C., Fuerstenberg, S. I., Fulton, L., Gaulin, E., Govers, F., Hughes, L., Humphray, S., Jiang, R. H. Y., Judelson, H., Kamoun, S., Kyung, K., Meijer, H., Minx, P., Morris, P., Nelson, J., Phuntumart, V., Qutob, D., Rehmany, A., Rougon-Cardoso, A., Ryden, P., Torto-Alalibo, T., Studholme, D., Wang, Y., Win, J., Wood, J., Clifton, S. W., Rogers, J., Van den Ackerveken, G., Jones, J. D. G., McDowell, J. M., Beynon, J., and Tyler, B. M. 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* 330:1549-1551.
- Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340:783-795.
- Birch, P. R. J., Rehmany, A. P., Pritchard, L., Kamoun, S., and Beynon, J. L. 2006. Trafficking arms: Oomycete effectors enter host plant cells. *Trends Microbiol.* 14:8-11.
- Chang, J. H., Urbach, J. M., Law, T. F., Arnold, L. W., Hu, A., Gombar, S., Grant, S. R., Ausubel, F. M., and Dangel, J. L. 2005. A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*. *Proc. Natl. Acad. Sci. U.S.A.* 102:2549-2554.
- Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. 2017. Shiny: web application framework for R. R package version 1.0.5. CRAN Project. Published online. <https://cran.r-project.org/web/packages/shiny/index.html>
- Charif, D., and Lobry, J. R. 2007. SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. Pages 207-232 in: *Structural Approaches to Sequence Evolution*. U. Bastioli and M. Porto, E. Roman, and M. Vendruscolo. Springer-Verlag, Berlin.
- Dalio, R. J. D., Herlihy, J., Oliveira, T. S., McDowell, J. M., and Machado, M. 2018. Effector biology in focus: A primer for computational prediction and functional characterization. *Mol. Plant-Microbe Interact* 31:22-33.
- Eddy, S. R. 1998. Profile hidden Markov models. *Bioinformatics* 14: 755-763.
- Eddy, S. R. 2011. Accelerated profile HMM searches. *PLOS Comput. Biol.* 7:e1002195.

- Ellis, J. G., Rafiqi, M., Gan, P., Chakrabarti, A., and Dodds, P. N. 2009. Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. *Curr. Opin. Plant Biol.* 12: 399-405.
- Erwin, D. C., and Ribeiro, O. K. 1996. *Phytophthora* diseases worldwide. APS Press, St. Paul, MN, U.S.A.
- Fry, W. 2008. *Phytophthora infestans*: The plant (and *R* gene) destroyer. *Mol. Plant Pathol.* 9:385-402.
- Govers, F., and Bouwmeester, K. 2008. Effector trafficking: RXLR-dEER as extra gear for delivery into plant cells. *Plant Cell* 20:1728-1730.
- Grünwald, N. J., Goss, E. M., and Press, C. M. 2008. *Phytophthora ramorum*: A pathogen with a remarkably wide host range causing sudden oak death on oaks and ramorum blight on woody ornamentals. *Mol. Plant Pathol.* 9:729-740.
- 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., Silver, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J., 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., 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.
- Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Umayam, L., Gill, S. R., Nelson, K. E., Read, T. D., Tettelin, H., Richardson, D., Ermolaeva, M. D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R. D., Nierman, W. C., White, O., Salzberg, S. L., Smith, H. O., Colwell, R. R., Mekalanos, J. J., Venter, J. C., and Fraser, C. M. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406:477-483.
- Jiang, R. H. Y., and Tyler, B. M. 2012. Mechanisms and evolution of virulence in oomycetes. *Annu. Rev. Phytopathol.* 50:295-318.
- Jones, J. D. G., and Dangl, J. L. 2006. The plant immune system. *Nature* 444:323-329.
- Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu. Rev. Phytopathol.* 44:41-60.
- Kamoun, S. 2007. Groovy times: Filamentous pathogen effectors revealed. *Curr. Opin. Plant Biol.* 10:358-365.
- Katoh, K., and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30:772-780.
- Lévesque, C. A., Brouwer, H., Cano, L., Hamilton, J. P., Holt, C., Huitema, E., Raffaele, S., Robideau, G. P., Thines, M., Win, J., Zerillo, M. M., Beakes, G. W., Boore, J. L., Busam, D., Dumas, B., Ferreira, S., Fuerstenberg, S. I., Gachon, C. M. M., Gaulin, E., Govers, F., Grenville-Briggs, L., Horner, N., Hostetler, J., Jiang, R. H. Y., Johnson, J., Krajaeun, T., Lin, H., Meijer, H. J. G., Moore, B., Morris, P., Phuntumart, V., Puiu, D., Shetty, J., Stajich, J. E., Tripathy, S., Wawra, S., van West, P., Whitty, B. R., Coutinho, P. M., Henrissat, B., Martin, F., Thomas, P. D., Tyler, B. M., De Vries, R. P., Kamoun, S., Yandell, M., Tisserat, N., and Buell, C. R. 2010. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol.* 11:R73.
- Links, M. G., Holub, E., Jiang, R. H. Y., Sharpe, A. G., Hegedus, D., Beynon, E., Sillito, D., Clarke, W. E., Uzuhashi, S., and Borhan, M. H. 2011. De novo sequence assembly of *Albugo candida* reveals a small genome relative to other biotrophic oomycetes. *BMC Genomics* 12:503.
- Ma, L.-J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P. M., Kang, S., Shim, W.-B., Woloshuk, C., Xie, X., Xu, J.-R., Antoniw, J., Baker, S. E., Bluhm, B. H., Breakspear, A., Brown, D. W., Butchko, R. A. E., Chapman, S., Coulson, R., Coutinho, P. M., Danchin, E. G. J., Diener, A., Gale, L. R., Gardiner, D. M., Goff, S., Hammond-Kosack, K. E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C. D., Koehrsen, M., Kumar, L., Lee, Y.-H., Li, L., Manners, J. M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.-Y., Proctor, R. H., Regev, A., Ruiz-Roldan, M. C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D. C., Turgeon, B. G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C. A., Kistler, H. C., and Rep, M. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464:367-373.
- Pais, M., Win, J., Yoshida, K., Etherington, G. J., Cano, L. M., Raffaele, S., Banfield, M. J., Jones, A., Kamoun, S., and Saunders, D. G. 2013. From pathogen genomes to host plant processes: The power of plant parasitic oomycetes. *Genome Biol.* 14:211.
- Petre, B., and Kamoun, S. 2014. How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biol.* 12:e1001801.
- R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European molecular biology open software suite. *Trends Genet.* 16:276-277.
- Schneider, T. D., and Stephens, R. M. 1990. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* 18:6097-6100.
- Selin, C., de Kievit, T. R., Belmonte, M. F., and Fernando, W. G. 2016. Elucidating the role of effectors in plant-fungal interactions: Progress and challenges. *Front. Microbiol.* 7:600.
- Shneider, M. M., Buth, S. A., Ho, B. T., Basler, M., Mekalanos, J. J., and Leiman, P. G. 2013. PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* 500:350-353.
- Sonah, H., Deshmukh, R. K., and Bélanger, R. R. 2016. Computational prediction of effector proteins in fungi: Opportunities and challenges. *Front. Plant Sci.* 7:126.
- Sperschneider, J., Williams, A. H., Hane, J. K., Singh, K. B., and Taylor, J. M. 2015. Evaluation of secretion prediction highlights differing approaches needed for oomycete and fungal effectors. *Front. Plant Sci.* 6:1168.
- Stajich, J. E., Harris, T., Brunk, B. P., Brestelli, J., Fischer, S., Harb, O. S., Kissinger, J. C., Li, W., Nayak, V., Pinney, D. F., Jr., Stoeckert, C. J., Jr., and Roos, D. S. 2012. FungiDB: An integrated functional genomics database for fungi. *Nucleic Acids Res.* 40:D675-D681.
- Toruño, T. Y., Stergiopoulos, I., and Coaker, G. 2016. Plant-pathogen effectors: Cellular probes interfering with plant defenses in spatial and temporal manners. *Annu. Rev. Phytopathol.* 54:419-441.
- Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grünwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K. H., Lee, M. K., McDonald, W. H., Medina, M., Meijer, H. J., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K., Sakiham, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z., Zhang, H., Grigoriev, I. V., Rokhsar, D. S., and Boore, J. L. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313:1261-1266.
- Wang, X., Allen, R., Ding, X., Goellner, M., Maier, T., de Boer, J. M., Baum, T. J., Hussey, R. S., and Davis, E. L. 2001. Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. *Mol. Plant-Microbe Interact* 14:536-544.
- Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer, New York.
- 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 Pathog.* 8:e1002400.