



Evaluation of Secretion Prediction Highlights Differing Approaches Needed for Oomycete and Fungal Effectors

Jana Sperschneider^{1*}, Angela H. Williams^{1,2}, James K. Hane^{3,4}, Karam B. Singh^{1,2} and Jennifer M. Taylor⁵

¹ CSIRO Agriculture Flagship, Centre for Environment and Life Sciences, Perth, WA, Australia, ² The Institute of Agriculture, The University of Western Australia, Crawley, WA, Australia, ³ Department of Environment and Agriculture, CCDM Bioinformatics, Centre for Crop and Disease Management, Curtin University, Perth, WA, Australia, ⁴ Curtin Institute for Computation, Curtin University, Perth, WA, Australia, ⁵ CSIRO Agriculture, Black Mountain Laboratories, Canberra, ACT, Australia

OPEN ACCESS

Edited by:

Marc-Henri Lebrun,
Institut National de la Recherche
Agronomique, France

Reviewed by:

Guus Bakkeren,
Agriculture & Agri-Food Canada,
Canada
Gregor Langen,
University of Cologne, Germany
Marc-Henri Lebrun,
Institut National de la Recherche
Agronomique, France

*Correspondence:

Jana Sperschneider
jana.sperschneider@csiro.au

Specialty section:

This article was submitted to
Plant Biotic Interactions,
a section of the journal
Frontiers in Plant Science

Received: 05 June 2015

Accepted: 07 December 2015

Published: 23 December 2015

Citation:

Sperschneider J, Williams AH,
Hane JK, Singh KB and Taylor JM
(2015) Evaluation of Secretion
Prediction Highlights Differing
Approaches Needed for Oomycete
and Fungal Effectors.
Front. Plant Sci. 6:1168.
doi: 10.3389/fpls.2015.01168

The steadily increasing number of sequenced fungal and oomycete genomes has enabled detailed studies of how these eukaryotic microbes infect plants and cause devastating losses in food crops. During infection, fungal and oomycete pathogens secrete effector molecules which manipulate host plant cell processes to the pathogen's advantage. Proteinaceous effectors are synthesized intracellularly and must be externalized to interact with host cells. Computational prediction of secreted proteins from genomic sequences is an important technique to narrow down the candidate effector repertoire for subsequent experimental validation. In this study, we benchmark secretion prediction tools on experimentally validated fungal and oomycete effectors. We observe that for a set of fungal SwissProt protein sequences, SignalP 4 and the neural network predictors of SignalP 3 (*D*-score) and SignalP 2 perform best. For effector prediction in particular, the use of a sensitive method can be desirable to obtain the most complete candidate effector set. We show that the neural network predictors of SignalP 2 and 3, as well as TargetP were the most sensitive tools for fungal effector secretion prediction, whereas the hidden Markov model predictors of SignalP 2 and 3 were the most sensitive tools for oomycete effectors. Thus, previous versions of SignalP retain value for oomycete effector prediction, as the current version, SignalP 4, was unable to reliably predict the signal peptide of the oomycete Crinkler effectors in the test set. Our assessment of subcellular localization predictors shows that cytoplasmic effectors are often predicted as not extracellular. This limits the reliability of secretion predictions that depend on these tools. We present our assessment with a view to informing future pathogenomics studies and suggest revised pipelines for secretion prediction to obtain optimal effector predictions in fungi and oomycetes.

Keywords: signal peptide prediction, effectors, protein secretion, fungi, oomycetes, plant pathogens

INTRODUCTION

The growing number of sequenced fungal and oomycete plant pathogen genomes has enabled detailed reverse genetics studies into molecular pathogen-host interactions (Dean et al., 2012; Kamoun et al., 2014). Though fungi and oomycetes belong to phylogenetically distinct microbial taxa, they both use a diverse class of molecules, termed effectors, to promote pathogenicity through subversion of host defenses or impairment of normal host-cell function (Dodds and Rathjen, 2010; Lo Presti et al., 2015). Effector molecules may be the products of either secondary metabolite or protein synthesis, however the majority of effectors identified in fungi and oomycetes are the latter. Proteinaceous effectors are initially synthesized intracellularly and require relocation to the extracellular space (apoplastic effectors) or subsequent import into the host cell cytoplasm or specific organelles (cytoplasmic effectors). The classical endoplasmic reticulum (ER)/Golgi-dependent secretion pathway in eukaryotes is well-defined and involves recognition of an N-terminal signal peptide that is cleaved off as the protein is translocated across the membrane (Von Heijne, 1990). Classical signal peptides can be predicted computationally with high accuracy (Menne et al., 2000; Klee and Ellis, 2005; Choo et al., 2009; Min, 2010; Melhem et al., 2013), and the majority of experimentally verified fungal and oomycete effectors are predicted to be secreted in this manner. However, reports are emerging for yet unknown, non-classical secretion pathways to also play a role in fungal and oomycete effector externalization (Ridout et al., 2006; Liu et al., 2014). Numerous eukaryotic plant pathogen effectors have been found to be active inside the host cell cytoplasm; however the knowledge of how effectors are delivered into the plant cells after secretion is fragmentary. In oomycete effectors, conserved amino acid motifs such as RXLR, CHXC, or LFLAK are positioned in N-terminal domains and define oomycete effector superfamilies (Petre and Kamoun, 2014). Although mechanisms have been proposed as to how the RXLR motif may facilitate cell entry through the host cell membrane phospholipid bilayer, the results are still controversial (Tyler et al., 2013; Wawra et al., 2013). Conserved sequence motifs associated with translocation have thus far not been found for fungal effector proteins, which makes their computational prediction from secretomes challenging (Sperschneider et al., 2015). A conserved Y/F/WxC-motif has been identified in the N-terminus of effector candidates in the barley powdery mildew fungus (Godfrey et al., 2010), however the role of this motif in cell entry or pathogenicity remains undetermined.

Several studies have exploited proteomics to experimentally identify secreted proteins involved in pathogenicity. For example, an early proteomics study of extracellular proteins of the wheat-infecting fungus *Fusarium graminearum* identified 120 candidates secreted *in planta*, of which only 56% possessed a predicted signal peptide motif (Paper et al., 2007). A later study in the same species identified only 69 secreted proteins, following growth in barley or wheat flour-based liquid cultures to mimic host-pathogen interactions (Yang et al., 2012). Of these, 70% possessed a predicted signal peptide. A recent study in the oomycete potato pathogen *Phytophthora infestans*

predicted 80% of its extracellular proteome to contain a signal peptide (Meijer et al., 2014). Thus, there appears to be wide variability (both between species and experiments) in the number of extracellular proteins identified through experimental proteomics that are also predicted to be secreted *in silico*. A high percentage of proteins lacking a classical signal peptide may be due to contamination of extracellular samples with intracellular proteins, due to rupture of the fungal cells during the protein extraction procedure. Furthermore, protein extraction may be complicated in species where there is a low or variable pathogen biomass relative to the host, or that selectively secrete different proteins when grown in different *in vitro* or infection-mimicking cultures. Computational limitations also have the potential to complicate proteomics experiments. This may come from variability between species in their use of non-classical secretion mechanisms, which cannot yet be accurately predicted. Gene annotation is also an important determining factor for the reliability of both experimental proteomics and computational prediction of secretion. Proteomic identification of genes is dependent on the completeness and accuracy of translated gene annotations that are used to generate a searchable database of predicted trypsin-digested proteins, to which peptide mass-spectra are matched (Bringans et al., 2009). Thus, missing or incorrect gene annotations may exclude or confuse identification of extracellular proteins. Prediction of secretion also relies strongly on the presence and accurate annotation of the 5' exons of genes, which encode N-terminal signal peptides. Due to these technical difficulties, deriving accurate computational predictions of secreted proteins from whole genome sequences remains an important pursuit in plant pathology, with a view toward efficient identification of secreted proteins for subsequent effector prediction.

The apparent ease of secretion prediction has led to its common use in pathogenomic studies as a first pass filter in narrowing down a whole proteome dataset into a short-list of potential effector candidates (Kämper et al., 2006; Raffaele et al., 2010; Rouxel et al., 2011; Hane et al., 2014; Nemri et al., 2014). A variety of software tools exists for eukaryotes that can predict whether proteins are secreted into the extracellular environment (Emanuelsson et al., 2007). Typically, this involves recognition of the N-terminal secretory signal peptide motif that directs proteins through the classical ER/Golgi-dependent pathway using tools such as SignalP (Petersen et al., 2011). Whilst this is a robust approach for defining a set of potential effector candidates, typically far more candidates are predicted for experimental validation than is feasible. Furthermore, proteins that are predicted to be secreted via a classical pathway might be retained in the ER/Golgi or fulfill roles as part of the cell wall. Therefore, subcellular localization prediction is an important tool that can point toward the functional role or interaction partners of a protein based on its amino acid sequence and can be used to assess if a protein is indeed secreted into the extracellular space (Emanuelsson, 2002). Transmembrane proteins are also commonly predicted and removed from the secretome as these are likely to fulfill functions in the pathogen cell wall. Whilst *in silico* methods for secretome prediction are under active development and show robust performance, their

reported predictive accuracy strongly depends on the selection of the test set and independent benchmarking studies are important for an unbiased tool evaluation. For example, a comprehensive benchmark of secretion prediction tools found that predictive accuracy was in many cases lower than those initially reported by the developers (Klee and Ellis, 2005). Although an evaluation on a large test set covering a wide taxonomic spectrum gives a good indication of a tool's performance, it provides limited insight into its expected performance on a specialized set of proteins, such as effector proteins of fungal and oomycete pathogens.

This study set out to reveal the strengths and weaknesses of existing protein secretion and subcellular localization prediction methods, as applied to the identification of effector proteins produced by fungi or oomycete plant pathogens. Prediction pipelines that have been used in previous studies for defining secretomes and subsequently effector candidates of eukaryotic plant pathogens are diverse and highly parameterized, as exemplified in **Table 1**. For example, SignalP (Nielsen et al., 1997; Nielsen and Krogh, 1998; Bendtsen et al., 2004b; Petersen et al., 2011) or Phobius (Käll et al., 2004) are utilized by the majority of pipelines to extract proteins that are likely to be secreted via a classical pathway. Despite the availability of the latest version of SignalP 4, which was designed to discriminate between signal peptides and N-terminal transmembrane (TM) regions, previous versions (2 and 3) are still frequently used due to their increased sensitivity. Phobius was designed to predict secretion and N-terminal TM domains separately, predicting both the presence of a signal peptide and the number and location of TM helices.

Furthermore, there are also discrepancies in how tools are used and how thresholds for secretion are set (**Table 1**). For example, some studies have used the neural network scores from SignalP 2 and 3 with custom thresholds, whereas others rely on the hidden Markov model probability for predicting the presence of a signal peptide. SignalP 2 and 3 employ predictions from both a neural network (SignalP-NN) and a hidden Markov model (SignalP-HMM), whilst the latest version SignalP 4 is purely based on neural networks. SignalP 2 returns three neural network scores for each position in the sequence: a raw cleavage site score (*C*-score), the signal peptide score (*S*-score), and the combined cleavage site score (*Y*-score). For each sequence, it reports the maximal *C*-, *S*-, and *Y*-scores as well as the mean *S*-score between the N-terminus and the predicted cleavage site that it used to assess whether a sequence contains a signal peptide. Furthermore, it returns two hidden Markov model scores, the *C*-score as well as the probability that the sequence contains a signal peptide (*S*-probability). SignalP 3 replaces the previously used mean *S*-score for classification with the *D*-score, which is calculated as the average of the mean *S*-score, and the maximal *Y*-score. It still uses both neural network scores and calculates the signal peptide probability with a hidden Markov model. SignalP 4 is a neural network based method designed to discriminate between signal peptides and transmembrane regions. Prediction of signal peptides is based entirely on the *D*-score. For all scores, Boolean flags are provided which are either "Y" for a signal peptide or "N" for no signal peptide.

Subcellular localization tools such as TargetP (Emanuelsson et al., 2000), WoLF PSORT (Horton et al., 2007), or ProtComp

are frequently used to complement the predictions made by SignalP or Phobius, either through a union or intersection of predictions made by these methods (**Table 1**). This can serve to filter proteins that may be predicted to contain a signal peptide, yet that might not be fully secreted into the extracellular space due to being retained within the ER/Golgi. TargetP predicts if a protein is secreted or localized to the mitochondria, chloroplast, or another unknown location. It reports reliability class scores from 1 to 5, where 1 corresponds to the strongest prediction. Another tool WoLF PSORT, an updated version of PSORT II, has been trained separately on fungi, animal, and plant data. It reports predicted subcellular locations (nuclear, mitochondria, cytosol, cytoskeleton, endoplasmic reticulum, plasma membrane, extracellular, chloroplast, peroxisome, Golgi apparatus, lysosome, and vacuolar membrane) in terms of respective scores based on a weighted *k*-nearest neighbor classifier. The output format is similar to a sequence similarity search, with scores assigned for each predicted localization site based on the number of nearest neighbors to the query protein. In most studies that employ WoLF PSORT, proteins have been predicted as secreted where extracellular predictions score higher than other locations (**Table 1**). Less commonly, the prediction of non-classically secreted proteins has been reported using SecretomeP, which has been trained on a very small set of verified non-classically secreted proteins derived from mammalian and bacterial sequences (Bendtsen et al., 2004a). Consequently, the relevance of SecretomeP to fungal and oomycete proteins is questionable. Finally, ProtComp is a web-server based tool combining several methods for protein localization, ranging from neural networks to sequence homology searches. Its lack of a publicly distributed version for local installation precludes it from routine use for whole-genome analysis. Predicted transmembrane proteins are typically removed from the set of predicted extracellular proteins using programs such as TMHMM (Krogh et al., 2001) or Phobius. However, most pipelines allow for the presence of one transmembrane domain in the N-terminus, as this can correspond to the signal peptide as both are predicted based on the presence of hydrophobic residues. Additionally, TargetP is often employed to eliminate proteins predicted to be targeted to mitochondria or chloroplasts. In some fungal studies, predicted GPI-anchored proteins are also removed from the set of secreted effector candidates.

The diversity of prediction pipelines shown in **Table 1** illustrates an overall lack of consensus used to predict extracellular pathogen proteins, in particular for effector candidates, and presents difficulties when comparing secretome sizes across different species. Herein we benchmark the performance of individual secretion prediction tools on experimentally verified fungal and oomycete effectors and use the best-performing tools to predict extracellular proteins across fungal and oomycete pathogens. In particular, we show that for cytoplasmic effector proteins that are first secreted into the extracellular space and subsequently translocated to the host cell, protein subcellular localization predictors suffer from poor accuracy. We highlight differences in performance for secretion prediction between fungal effectors and oomycete effectors and conclude by providing practical recommendations for the

TABLE 1 | Examples for approaches used in eukaryotic plant pathogen genomic studies that predict secreted proteins.

References	Species/taxon	Secretome size (total number of predicted genes)	Positive evidence	Negative evidence	Remarks
(A) FUNGAL PATHOGENS					
Kämper et al., 2006	<i>Ustilago maydis</i>	426 (6902)	SignalP 3.0, TargetP, ProtComp 6.0		
Ma et al., 2010	<i>Fusarium oxysporum</i> f. sp. <i>lycopersici</i>	1803 (17,735)	SignalP 3.0	TMHMM, Phobius, TargetP	SignalP <i>D</i> -score, TM allowed in the first 50 amino acids, TargetP no mitochondrial targeting with RC = 1 or 2
	<i>F. verticillioides</i>	1549 (14,179)			
Spanu et al., 2010	<i>Blumeria graminis</i> f. sp. <i>hordei</i>	NA (5854)	SignalP, SecretomeP	–	S-prob > 0.9, SecretomeP > 0.5
Rouxel et al., 2011	<i>Leptosphaeria maculans</i>	NA (12,469)	SignalP 3.0, TargetP	TMHMM	Both SignalP-NN and SignalP-HMM predict SP, TargetP predicts “secreted,” 1 TM in SP allowed
Duplessis et al., 2011	<i>Melampsora larici-populina</i>	NA (16,399)	SignalP	TargetP, TMHMM, big-PI	S-prob > 0.9, TargetP no mitochondrial targeting with RC = 1 or 2, one TM ≥ 18 in first 60 aas allowed
	<i>Puccinia graminis</i> f. sp. <i>tritici</i>	NA (17,773)			
Klosterman et al., 2011	<i>Verticillium dahliae</i>	780 (10,535)	SignalP 3.0, WoLF PSORT	TMHMM, Phobius	<i>D</i> -score > 0.5, WoLF PSORT predicted to be extracellular, no TM predicted by either TMHMM or Phobius
	<i>V. albo-atrum</i>	759 (10,221)			
Cantu et al., 2011	<i>P. striiformis</i> f. sp. <i>tritici</i>	1088 (20,423)	SignalP 3.0	TMHMM, TargetP	S-prob ≥ 0.9, average cleavage site position: 24 ± 9, TM length ≥ 18 aas and not in the first 60 aas, TargetP no mitochondrial proteins (RC = 1 or 2)
O’Connell et al., 2012	<i>Colletotrichum higginsianum</i>	2142 (16,172)	WoLF PSORT	–	WoLF PSORT predicted to be extracellular
	<i>C. graminicola</i>	1650 (12,006)			
Ohm et al., 2012	Dothideomycetes	NA	SignalP 3.0	TMHMM 2.0	One TM allowed in first 40 aas
Brown et al., 2012	<i>Fusarium graminearum</i>	574 (13,937)	SignalP 3.0, TargetP 1.1, ProtComp 8.0, WoLF PSORT 0.2	TMHMM 2.0, big-PI	SignalP <i>D</i> -score = Y, TargetP predicts “secreted,” ProtComp extracellular or unknown, no TM in mature protein, WoLF PSORT extracellular score > 17, no GPI-anchor
de Wit et al., 2012	<i>Cladosporium fulvum</i>	1200 (14,127)	SignalP 3.0, WoLF PSORT	Phobius, TMHMM 2.0, TargetP 1.1, PredGPI	<i>D</i> -score > 0.5, TargetP no mitochondrial targeting, no GPI anchor, no TM predicted by either TMHMM or Phobius
	<i>Dothistroma septosporum</i>	905 (12,580)			
Wiemann et al., 2013	<i>Fusarium fujikuroi</i>	1336 (14,813)	SignalP 4.0, SecretomeP, WoLF PSORT	TargetP, TMHMM	SecretomeP score > 0.6, TargetP no mitochondrial targeting, TargetP RC-score < 4, SignalP <i>S</i> -score > 0.5
Manning et al., 2013	<i>Pyrenophora tritici-repentis</i>	1146 (12,141)	SignalP 3.0, WoLF PSORT	TMHMM	WoLF PSORT predicted to be extracellular, one TM allowed unless it starts in the first 10 aas
Hane et al., 2014	<i>Rhizoctonia solani</i> AG8	1959 (13,964)	SignalP 4.1, Phobius 1.01, WoLF PSORT 0.2	Phobius 1.01	SignalP secreted or Phobius secreted or WoLF PSORT predicted to be extracellular, only one TM domain allowed
Nemri et al., 2014	<i>Melampsora lini</i>	1085 (26,443)	SignalP 2.1-HMM, SignalP 4.1	TMHMM 2.0, TargetP 1.1	No TM, TargetP no mitochondrial targeting, <i>D</i> -score > 0.36, S-prob > 0.88, cleavage site predicted between amino-acid 10 and 40 using SignalP 2

(Continued)

TABLE 1 | Continued

References	Species/taxon	Secretome size (total number of predicted genes)	Positive evidence	Negative evidence	Remarks
Guyon et al., 2014	<i>Sclerotinia sclerotiorum</i>	745 (14,503)	SignalP 2, SignalP 4	TMHMM, GPIsom	TM after SP removed, no GPI anchor
(B) OOMYCETE PATHOGENS					
Haas et al., 2009	<i>Phytophthora infestans</i>	NA (17,797)	SignalP 3.0	–	–
Raffaele et al., 2010	<i>P. infestans</i>	1415 (18,155)	SignalP 2.0, SignalP 3.0, TargetP, PSort	TMHMM	SignalP-HMM 2.0 score ≥ 0.9 , SignalP-NN 3.0 Y-max score ≥ 0.5 , SignalP-NN 3.0 D-score ≥ 0.5 , SignalP-HMM 3.0 S-prob ≥ 0.9 , TargetP predicted localization “Secreted” (S), most probable PSort location “extracellular” (extr.) and no TMHMM predicted TM domain after signal peptide cleavage site
Lévesque et al., 2010	<i>Pythium ultimum</i>	747 (15,297)	SignalP 2.0	TMHMM, TargetP	SignalP-HMM predicts signal peptide, SignalP-NN predicts a cleavage site between amino acids 10 and 40
Links et al., 2011	<i>Albugo candida</i>	939 (15,824)	SignalP 3.0	–	Either SignalP-NN or SignalP-HMM predict SP, SignalP predicts a cleavage site between amino acids 10 and 30
Kemen et al., 2011	<i>A. laibachii</i>	1636 (13,032)	SignalP 3.0	MEMSAT3	Both neural network and hidden Markov model predict signal peptide. Proteins were considered to be without a TM domain with $p_{\text{non-TM}} > 0.0004$ or, for high stringency, $p_{\text{non-TM}} > 0.01$.
(C) SECRETOME STUDIES					
Lowe and Howlett, 2012	Fungi	–	SignalP	–	–
Sperschneider et al., 2015	Fungi	–	SignalP 4.1	–	–
Lo Presti et al., 2015	Fungi	–	SignalP 4.0	TMHMM	No TMs as predicted by TMHMM 2.0c (TMHMM score < 2)

If provided in the original paper, version numbers of prediction tools are given.

computational secretion prediction for effector candidate mining from eukaryotic pathogen genomes.

MATERIALS AND METHODS

Various datasets were chosen for the purpose of comparing the performance of secretion prediction software tools, in the context of plant pathogenomics. Experimentally validated fungal and oomycete effector protein sequences were collected from PHIbase version 3.6 (Urban et al., 2015) and from manual literature searches (Supplementary Data Sheet 1, 2, Supplementary Table 1). For further benchmarking, representative datasets for both extracellular and intracellular proteins of the fungi were obtained by searching SwissProt database records created between 2011 and 2015 for: (1) fungal proteins that have been manually annotated as secreted (taxonomy: “Fungi [4751]” locations: (location: “Secreted [SL-0243]” evidence: manual) created: [20110101 TO 20150101]) (Supplementary Data Sheet 3); and (2) fungal proteins that

have been manually annotated as localized to the nucleus (taxonomy: “Fungi [4751]” locations: (location: “Nucleus [SL-0191]” evidence: manual) created: [20110101 TO 20150101]) (Supplementary Data Sheet 4). Sequences that did not start with “M” or were shorter than 30 aas were removed. Both sets only cover proteins for which entries were created after 2011, to avoid an overlap with the training sets used for secretion prediction tools. We could not extract an equivalent set for oomycete proteins from SwissProt due to the very low number of entries for manually curated secreted proteins (four entries). Secretion prediction tools were run on a local machine, or using web servers where indicated, as in Table 2 (Results given in Supplementary Data Sheet 5). Sensitivity was calculated as $TP/(TP + FN)$ and specificity as $TN/(TN + FP)$, where TP is the number of true positives, TN the number of true negatives, FP the number of false positives and FN the number of false negatives. The Matthews correlation coefficient (MCC) was calculated as $\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$.

TABLE 2 | Software tested in this study and the parameters under which proteins were predicted to be secreted.

Software	References	Secretion or transmembrane prediction criteria
SignalP 2.0	Nielsen et al., 1997 Nielsen and Krogh, 1998	SignalP-HMM: labeled “Y” based on <i>S</i> -probability (default threshold 0.5). SignalP-NN: labeled “Y” based on <i>S</i> -score (default threshold 0.47).
SignalP 3.0	Bendtsen et al., 2004b	SignalP-HMM: labeled “Y” based on <i>S</i> -probability (default threshold 0.5). SignalP-NN: labeled “Y” based on <i>D</i> -score (default threshold 0.43).
SignalP 4.1	Petersen et al., 2011	Labeled “Y” based on <i>D</i> -score (default thresholds 0.45 for SignalP-noTM networks and 0.5 for SignalP-TM networks).
Phobius	Käll et al., 2004	Predicted as secreted if presence of a signal peptide (SP) labeled as “Y.” Transmembrane protein if the number of predicted transmembrane segments (TM) is ≥ 1 .
TargetP 1.1	Emanuelsson et al., 2000	Labeled “S” for signal peptide (“secreted”), regardless of the reliability class score.
WoLF PSORT 0.2	Horton et al., 2007	Secreted if the best score in the ranked localization list is “extracellular.”
ProtComp 9.0*	http://www.softberry.com/berry.phtml	Secreted if the integral prediction of protein location contains “extracellular (secreted).”
TMHMM 2.0	Krogh et al., 2001	Transmembrane protein if one or more transmembrane helices beginning outside the first 60 aas.

All tools were run with default parameters and settings.

*Run using web-server.

RESULTS AND DISCUSSION

Signalp 2, 3 and 4 Show the Best Performance for Secretion Prediction on a Set of Fungal Protein Sequences

Several independent benchmark analyses have been published that compare the accuracy of secretion prediction tools. For example, Klee and Ellis (2005) evaluated a range of secretion prediction methods (SignalP 3.0, SignalP 2.0, TargetP 1.01, PrediSi, Phobius, and ProtComp 6.0) on 372 proteins from five vertebrate organisms and found that TargetP, the SignalP 3 maximum *S*-score and SignalP 3 *D*-score were the most accurate single scores. Choo et al. (2009) found that most of the tested tools were capable of reliably distinguishing secreted from non-secreted proteins, as indicated by the high specificities that were achieved. SignalP 4 has been reported by the authors to outperform previous versions of SignalP for a test set spanning eukaryotic and bacterial sequences (Petersen et al., 2011).

Min (2010) evaluated eukaryotic secretion prediction using Phobius, SignalP 3.0, TargetP, and WoLF PSORT individually and in combination with TMHMM and PS-Scan and found that for fungi the most reliable individual predictor of secretion was WoLF PSORT, but a combination of tools produced the most accurate predictions. A follow-up study including SignalP 4.0 reported WoLF PSORT as the best individual tool for fungal data and also made the general recommendation of using SignalP 4.0 over SignalP 3.0 (Melhem et al., 2013). However, the authors assign a protein as predicted to be secreted by WoLF PSORT if it features “extracellular” in the ranked localization list whereas other studies (Table 1), including ours, have used this tool quite differently requiring more stringently that the “extracellular” score is higher than that of all other sub-cellular locations. Notably, WoLF PSORT stands out amongst the tools compared in that it has been trained on a relatively extensive set of fungal proteins. However, while it performs well for fungal secreted proteins overall, when restricted to known secreted effectors its performance is markedly poorer.

For the evaluation of secretion prediction performance we utilized two data sets from the SwissProt database: one that contained fungal proteins that were manually annotated as secreted (409 proteins) and the other that contained non-secreted fungal proteins that were manually annotated as nuclear (1113 proteins). We could not extract an equivalent set for oomycete proteins from SwissProt due to the very low number of entries for manually curated secreted or nuclear proteins. All tools tested achieved high specificity in the range of 97.2–99.8%, whereas sensitivity varied more dramatically (Table 3). All versions of SignalP, Phobius, and TargetP achieved high sensitivity of more than 94.9%. In contrast, the proportion of proteins that are predicted to be extracellular by WoLF PSORT and ProtComp showed lower sensitivity at 88 and 63.3%, respectively. In terms of the Matthews correlation coefficient (MCC), SignalP 4, SignalP-NN 3 (*D*-score), and SignalP-NN 2 perform best (MCC = 0.96), with SignalP 2 and 3 showing slightly more sensitivity than SignalP 4, which in turn achieves marginally higher specificity. These results confirm the strong predictive performance of SignalP for secreted fungal proteins.

Differences in Sensitivity of Secretion Prediction Tools for Effectors from Fungi and Oomycetes

In line with previous studies (Menne et al., 2000; Klee and Ellis, 2005; Choo et al., 2009; Min, 2010), we found that all tools tested achieved high specificity in secretion prediction. For effector prediction in particular, the use of a sensitive method can be desirable to obtain the most complete candidate effector set. To test the sensitivity of secretion prediction tools for effector proteins from eukaryotic plant pathogens, we collected two sets of experimentally verified fungal and oomycete effectors from the literature. In total, the test set of fungal and oomycete effectors contain 69 and 53 proteins, respectively, (Supplementary Table 1). Interestingly, the sensitivity of secretion prediction tools varied between the fungal and oomycete effector sets (Figure 1). The neural network predictors of SignalP 3 and SignalP 2

TABLE 3 | Performance of secretion prediction tools applied to secreted fungal proteins sourced from SwissProt.

Measure	SignalP 4.1	SignalP-NN 3.0 (D-Score)	SignalP-HMM 3.0	SignalP-NN 2.0	Signal-HMM 2.0	Phobius	TargetP 1.1	WoLF PSORT	ProtComp 9.0
Sensitivity	94.9%	95.6%	95.8%	95.6%	95.8%	95.4%	95.1%	88%	63.3%
Specificity	99.7%	99.6%	98.4%	99.4%	98.2%	98.7%	98.3%	99.8%	97.2%
MCC	0.96	0.96	0.94	0.96	0.94	0.94	0.93	0.91	0.68

Sensitivity, specificity and the Matthews correlation coefficient (MCC) are shown for evaluating the performance of secretion prediction tools. All tools were run with the settings and parameters given in **Table 2**. The best performance in terms of MCC is marked in bold.

(SignalP-NN 2, SignalP-NN 3) as well as TargetP (“S” for secreted with RC scores ranging from 1 to 5) were found to be the most sensitive for fungal effectors (95.7%). In contrast, the hidden Markov model predictors of SignalP 2 and SignalP 3 (SignalP-HMM 2, SignalP-HMM 3) achieved highest sensitivity for oomycete effectors (98.1%). In general, neural networks and hidden Markov models have different strengths in pattern recognition tasks. Whereas neural networks are powerful for correlating features over a longer range, hidden Markov models are advantageous for modeling sequential regions or patterns found in signal peptides (Nielsen et al., 1999). How this could relate to the prediction of signal peptides in fungal and oomycete effectors remains to be determined.

From the fungal effector set, all secretion predictors, including the best-performing tools SignalP-NN 2, SignalP-NN 3, and TargetP, were consistently unable to predict a signal peptide for only three effectors: Avra10, Avr1, and Vdls1 (**Table 4**). Similarly for the oomycete effector set, all secretion predictors including the best-performing tools SignalP 2-HMM and SignalP-HMM 3 were unable to predict a signal peptide for only a single oomycete effector (Psls1). These four effector proteins have been demonstrated to be secreted via non-classical pathways (Ridout et al., 2006; Liu et al., 2014). This suggests that the most sensitive methods are only likely to fail to predict the secretion of non-classically secreted effectors and that using a union of multiple methods would not necessarily improve sensitivity for this test set. At this stage the computational identification of non-classically secreted effectors remains challenging and these types of effectors require experimental validation of their secretion. In the future, an increased understanding of non-classical secretion mechanisms of fungal and oomycete effectors might lead to improved computational prediction of these effectors. Protein tribe clustering with subsequent examination of high-priority effector candidate families (Saunders et al., 2012) or the presence of conserved protein domains has been effectively applied to identify related effector candidates lacking a predicted signal peptide. However, as the vast majority of fungal effectors share little sequence homology, the utility of this method is limited. Furthermore, orthologs of a secreted protein are not necessarily also secreted (Poppe et al., 2015). Therefore, secretomes predicted through the additional use of reciprocal BLASTs and/or tribe analysis are likely to include a high number of false positives.

TargetP predicted signal peptides with the highest reliability class (RC = 1) for only 63.8% of fungal effectors and for

56.6% of oomycete effectors (**Figure 2**). Without a restriction on the reliability class (RC from 1 to 5), TargetP predicted “secreted” as the localization for 95.6% of the fungal effectors (three effectors were predicted as “unknown”), whereas it returned “secreted” for 92.4% of the oomycete effectors (two effectors were predicted as “unknown” and two were predicted as “mitochondrial”). Therefore, a restriction on the predicted reliability class should not be used for predicting the secretion of effectors and the exclusion of proteins predicted to be localized to mitochondria has to be used with caution for oomycete effectors.

The relatively poor performance of SignalP 4 for oomycete effectors (**Figure 1**, sensitivity 83%) is surprising and suggests that previous versions of SignalP (SignalP 2, SignalP 3) should be used for effector mining in oomycete genomes instead. In particular, SignalP 4 does not predict a signal peptide for six out of seven Crinkler effectors in the test set (**Table 4**; CRN1, CRN2, CRN8, CRN15, CRN16, CRN63, CRN115). Crinkler effectors are a large family of modular proteins that are translocated into host cells, featuring a signal peptide followed by a LXLFLAK sequence motif and C-terminal domains (Haas et al., 2009; Schornack et al., 2010). On the set of seven Crinkler effectors, SignalP 4 achieves the lowest sensitivity, whereas the hidden Markov model predictors of SignalP 2 and SignalP 3 (SignalP-HMM 2, SignalP-HMM 3) correctly predict the signal peptide in all seven Crinklers (**Table 4**). This exemplifies the substantial benefits of using previous versions of SignalP (SignalP 2, SignalP 3) for oomycete effector mining.

Signal peptide prediction tools such as SignalP return the set of proteins that are likely to carry a signal peptide for the classical pathway, but do not necessarily imply that a protein will be extracellular. Many proteins with a signal peptide are retained in various cellular compartments and thus, signal peptide prediction is often combined with additional evidence for extracellular protein secretion, such as the absence of transmembrane domains, GPI anchors or retention signals (**Table 1**). We found that no transmembrane regions outside the signal peptide region (first 60 aas) were predicted for any of the 69 fungal effectors using TMHMM or Phobius. For the 53 oomycete effectors, TMHMM and Phobius both return one transmembrane helix outside the signal peptide region for the RXLR effector PITG_03192. This might be an indication that TMHMM and Phobius can be used as a preliminary filter to exclude proteins with multiple, non-N-terminal transmembrane domains for effector mining in fungi. However, these tools should

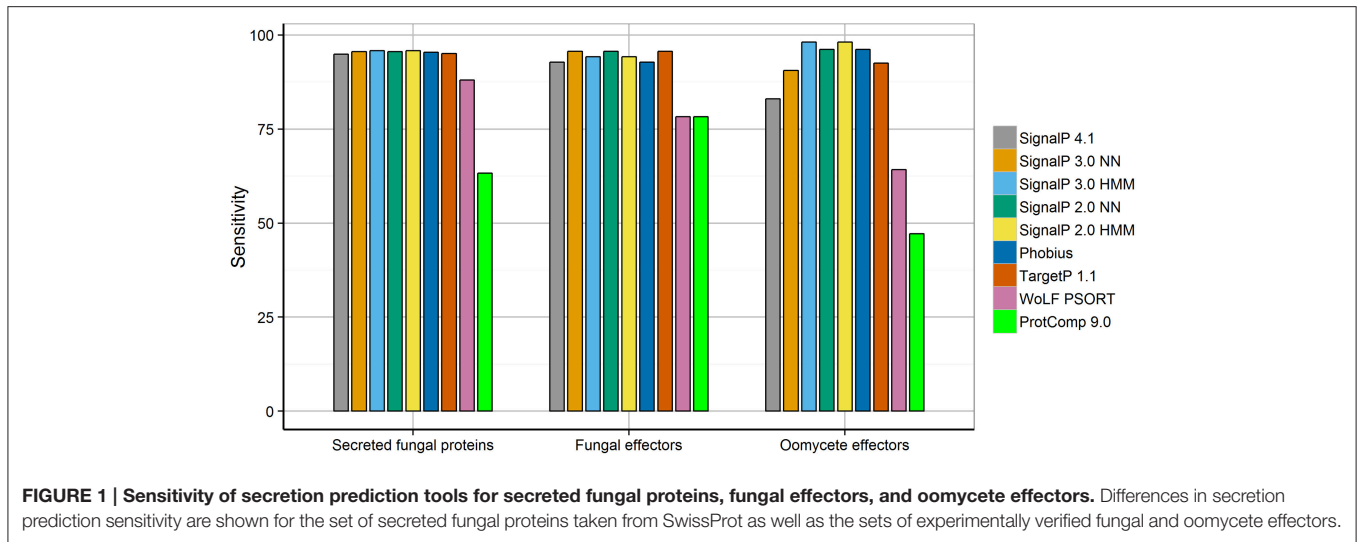


TABLE 4 | Fungal and oomycete effectors that were not predicted to be secreted by the prediction tools tested.

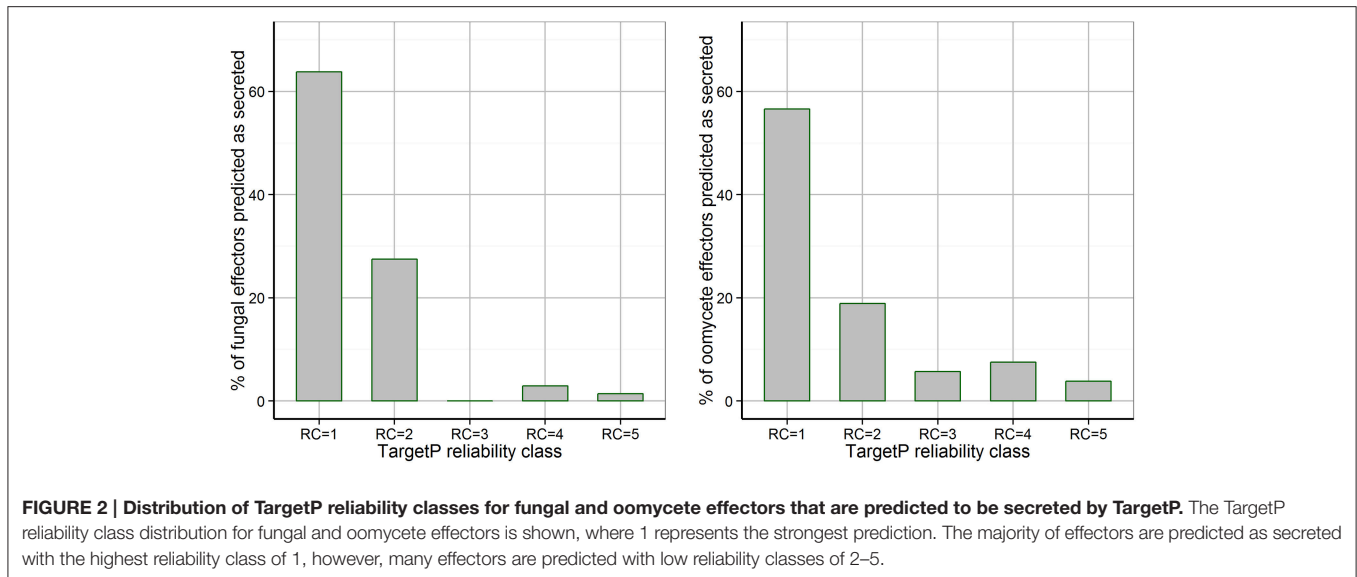
	SignalP 4.1	SignalP-NN 3.0 (D-Score)	SignalP-HMM 3.0	SignalP-NN 2.0	Signal-HMM 2.0	Phobius	TargetP 1.1
Fungal effectors	Avra10 Avrk1 Vdlscl AvrP4 SIX5	Avra10 Avrk1 Vdlscl	Avra10 Avrk1 Vdlscl Avr-Pita2	Avra10 Avrk1 Vdlscl	Avra10 Avrk1 Vdlscl Avr-Pita2	Avra10 Avrk1 Vdlscl Avr-Pita2 AvrP123	Avra10 Avrk1 Vdlscl
Oomycete effectors	Pslsc1 CRN1 CRN2 CRN15 CRN16 CRN63 CRN115 Avr1k ATR5	Pslsc1 CRN1 CRN15 CRN16 ATR5	Pslsc1	Pslsc1 CRN16	Pslsc1	Pslsc1 CRN16	Pslsc1 CRN16 Avr1k ATR5

be used with less stringent requirements for effector prediction in oomycetes.

Subcellular Localization Prediction Tools Should not be used for Predicting Effector Secretion

Prediction of subcellular localization is important for inferring hints about a protein's function. In eukaryotes, a number of compartments exist to which proteins may be localized, e.g., the extracellular space, mitochondria, chloroplast, nucleus, peroxisome, cytosol or plasma membrane. Several plant pathogenomics studies have used the subcellular localization of "extracellular" as a criterion for predicting secretion, commonly using WoLF PSORT which has been trained separately on fungi, animal and plant data. However, we found that applying WoLF PSORT (fungi) to the sets of experimentally verified fungal and oomycete effectors returned 25 cytoplasmic effectors that are

not predicted to be extracellular (34.2% of cytoplasmic effectors, **Figure 3**). This could be explained as follows. First, the estimated sensitivity and specificity of WoLF PSORT is fairly low at around 70% (Horton et al., 2007), which might lead to a high number of false predictions. However, we found that false predictions occurred in particular for non-apoplastic effectors (**Figure 3**). It is possible that WoLF PSORT may have predicted a signal for host cell localization in effectors rather than for the extracellular secretion of the effector from the pathogen cell. Thus, WoLF PSORT should be used with caution when predicting secretomes and its "extracellular" predictions should not be solely relied upon for effector prediction. An alternative approach is to impose a high level of stringency to WoLF PSORT predictions, as was the case for the *F. graminearum* secretome in which proteins were reported as secreted if the extracellular score was >17 (Brown et al., 2012). Whilst this practice is likely to drastically reduce the number of false positives in the secretome, it is prone to miss *bona fide* effectors that are not predicted to be



extracellularly localized. In this study, of the oomycete Crinkler effectors CRN1, CRN2, CRN8, CRN15 and CRN16 which are known to localize to the host cell nucleus (Schornack et al., 2010), WoLF PSORT only predicted a nuclear localization for CRN16. Therefore, the predictions of subcellular localization tools may need to be used with caution in effector prediction studies.

Practical Recommendations for Prediction of Extracellular Proteins in Fungi and Oomycetes

In this study, we have assessed the performance of various secretion and subcellular localization prediction tools, when applied to datasets derived from known fungal and oomycete effectors, as well as extracellular and intracellular fungal proteins. Based on our benchmarking, we deduce recommendations for extracellular protein prediction in fungal and oomycete pathogen genomes.

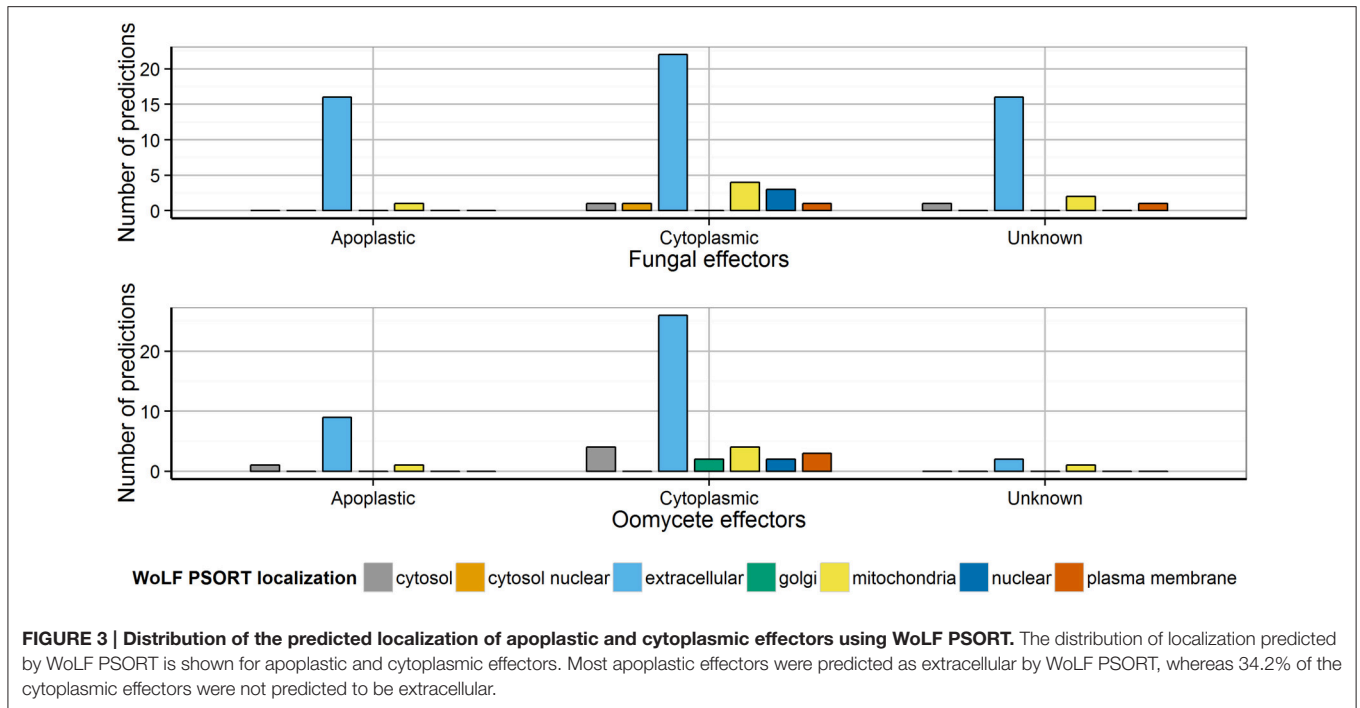
We observe that previous versions of SignalP (2, 3) demonstrate increased sensitivity over the latest version (4.1) for predicting signal peptides of oomycete effectors, with the HMM-based methods outperforming the NN-based methods. Indeed, this has formed the basis for the pipeline PexFinder (Phytophthora Extracellular Proteins Finder), which automates identification of oomycete extracellular proteins from EST data (Torto et al., 2003). PexFinder uses SignalP 2.0 but applies an additional logical filter that predicts a protein to be secreted only if both the hidden Markov model predicts a signal peptide and the neural network predicts a cleavage site between amino acids 10 and 40. Whilst this pipeline was proposed over a decade ago, it still retains its value for mining effectors from oomycete genomes.

In contrast with oomycete effectors, the NN predictors of SignalP 2 and 3, as well as TargetP, were observed to be the most sensitive for predicting signal peptides of fungal effector proteins. Unlike oomycete effectors, no TM domains were predicted outside the N-terminal signal peptide region using

TMHMM or Phobius. Therefore, we propose that for fungal effector mining the requirement of a predicted signal peptide using either SignalP-NN 2 or 3, a TargetP localization prediction of “secreted” or “unknown” (with no restriction on the RC score) and a lack of transmembrane domains outside the signal peptide region (TMHMM/Phobius) would be a robust method. Applying this proposed pipeline to publicly available fungal genomes (some with secretome predictions given in Table 1) highlights the wide variability in the number of predicted secreted proteins produced by the different techniques used in previously published studies (Figure 4). In line with previous reports, we observe a higher percentage of proteins that are predicted to be secreted in pathogens with a biotrophic phase, compared to necrotrophs and saprophytes (Lowe and Howlett, 2012; Lo Presti et al., 2015). By our method, similar numbers of predicted secreted proteins were predicted across multiple species of the same trophic class, whereas reported numbers were highly variable in genome survey publications for these species (Figure 4).

CONCLUSION

Prediction of effector proteins is of vital importance to the field of plant pathology, and relies heavily on the strengths or weaknesses of secretion prediction software. In this study, we assess the performance of popular software tools against known effectors of both the fungi and oomycetes and offer recommendations on which may be better suited to specialized applications. However, such performance evaluations inevitably vary based on the test data sets used, and therefore, we advise readers to carefully consider the suitability of these recommendations to their own data. Based on the results discussed herein, we recommend the use of the neural network predictors of SignalP 2 or 3, a TargetP localization prediction of “secreted” as well as transmembrane protein removal using either TMHMM or Phobius as a robust choice for predicting the secretion of fungal effectors. In comparison, the hidden Markov model predictors of

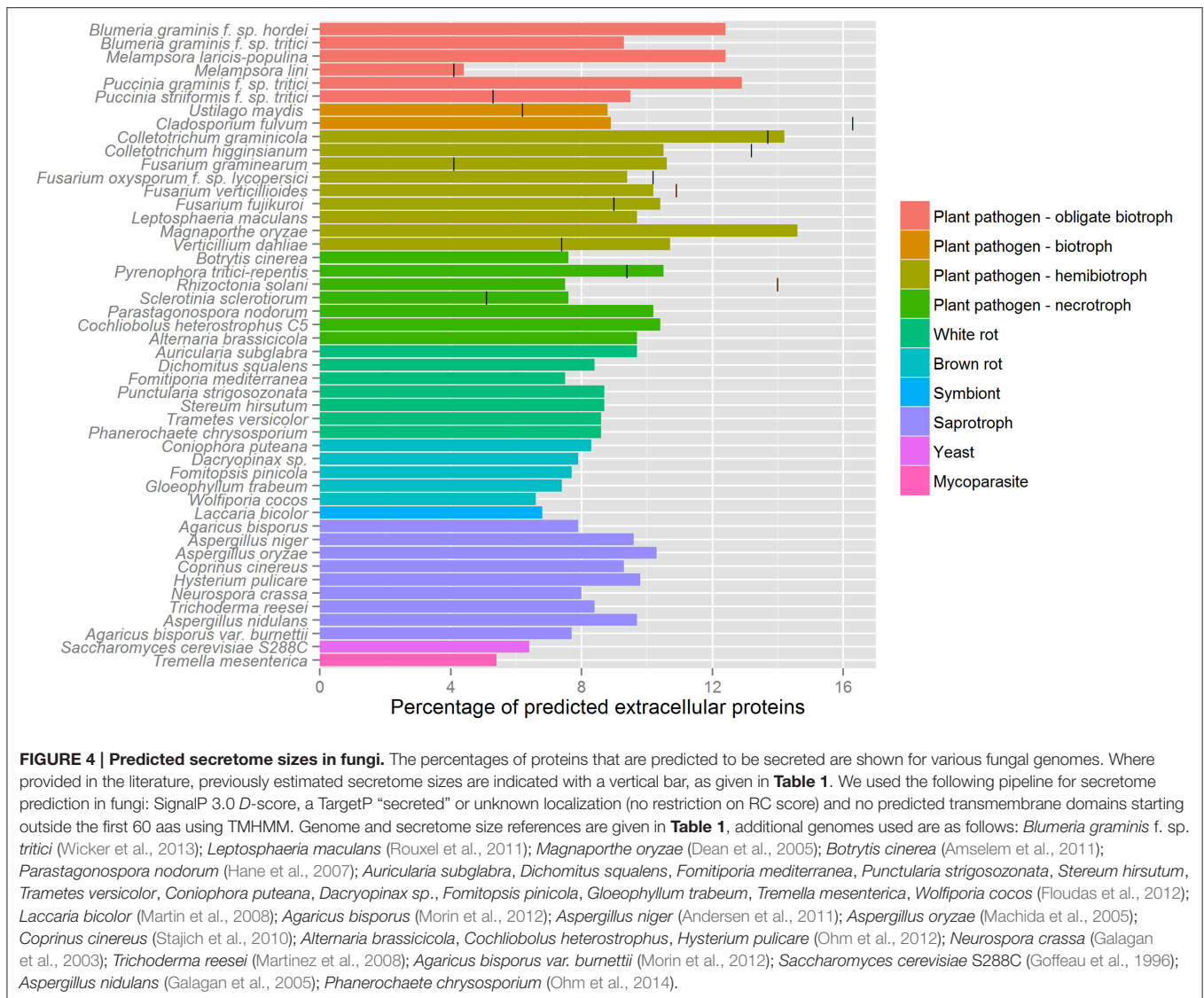


SignalP 2 and 3 perform best for predicting the signal peptide of oomycete effectors and automated pipelines such as PexFinder retain their value (Torto et al., 2003). However, the secretome includes many proteins unrelated to pathogenicity, and a number of additional conditions must be subsequently assessed in order to arrive at a subset that represents a potential set of effectors. In oomycetes, this can be achieved using motif enrichment analysis based on RXLR or Crinkler effector families (Petre and Kamoun, 2014), whereas in fungi this process is not feasible and alternative criteria such as small size, an enrichment in cysteines, genomic location, or signatures of diversifying selection can be used (Sperschneider et al., 2015).

While the reliability of secretion prediction is highly relevant to effector prediction, one must not overlook the potential for errors to arise from prior steps involved in the generation of sequence resources. The annotation of gene structure in effector genes can be particularly error-prone for various reasons, stemming from idiosyncrasies related to their genomic context for example their tendency to be associated with repetitive regions of the genome (Raffaele and Kamoun, 2012). There is potential for errors to occur in the assembled genome sequence, especially for those assembled from short-read data only, and the subsequent use of automated annotation pipelines can contribute to inaccurate or fragmented gene predictions. If this occurs in the 5' region it can lead to misprediction of N-terminal signal peptides. We also note that due to high gene density in fungi that transcript UTRs of adjacent gene loci frequently overlap (Guida et al., 2011; Wang et al., 2014), potentially resulting in gene annotations that are merged products of two or more adjacent loci. Therefore, the use of RNA-seq-based annotation methods specifically designed for fungi (Reid et al., 2014; Testa et al., 2015)

can be beneficial to arrive at an optimal set of gene annotations for subsequent secretion prediction.

Our results showed that one of the areas that is currently suffering from poor accuracy is the prediction of subcellular localization for effector proteins that are first secreted from the fungus and then targeted to a host organelle. In particular, we recommend that the requirement of extracellular localization as predicted by WoLF PSORT should not be used for effector mining in secretomes. Re-training subcellular localization tools with updated data sets including experimentally validated effectors might help to improve accuracy. There are few well-studied fungal effectors with confirmed host-localization, one being the SP7 effector of the arbuscular mycorrhizal *Glomus intraradices* (Kloppholz et al., 2011). SP7 is initially secreted to the apoplast, then imported into the host cell, and then into its nucleus. This localization is determined by multiple motifs, including a signal peptide, nuclear localization domain and an array of imperfect tandem hydrophilic repeats possibly involved in membrane integration. Both TargetP and WoLF PSORT predicted that the complete version of SP7 was secreted, however after removal of the signal peptide based on SignalP analysis, the TargetP prediction changed to "other" and WoLF PSORT (plant mode) predicted nuclear localization. Intriguingly, this suggests that subcellular localization prediction has the potential to become a powerful tool for providing insight into potential modes of action for candidate effectors based on their organelle targets. Additionally, there are currently no tools designed to predict proteins secreted in a non-classical manner that have been specifically trained on either fungi or oomycetes sequences due to a lack of training data. Although tools like SecretomeP are able to predict some cases (Liu et al., 2014), in the future refined tools for



non-classical secretion prediction could be a source of significant improvements in effector prediction.

In summary, whilst existing methods for signal peptide prediction achieve high accuracy, the main areas for improving eukaryotic effector secretion prediction will come from advances in subcellular localization prediction tools as well as from investigations of non-classical secretion pathways and improved gene prediction tools for pathogen genomes.

AUTHOR CONTRIBUTIONS

JS conceived the study and all authors contributed to the design of the study. JS, AW, and JH acquired, analyzed and interpreted the data. All authors drafted the manuscript and approved the final version.

FUNDING

JS was partially supported by the Australian Grains Research and Development Corporation.

ACKNOWLEDGMENTS

We thank Dr. Louise Thatcher and Dr. Ian Dry for their constructive feedback on this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.01168>

REFERENCES

- Amselem, J., Cuomo, C. A., Van Kan, J. A., Viaud, M., Benito, E. P., Couloux, A., et al. (2011). Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet.* 7:e1002230. doi: 10.1371/journal.pgen.1002230
- Andersen, M. R., Salazar, M. P., Schaap, P. J., van de Vondervoort, P. J., Culley, D., Thykaer, J., et al. (2011). Comparative genomics of citric-acid-producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88. *Genome Res.* 21, 885–897. doi: 10.1101/gr.112169.110
- Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., and Brunak, S. (2004a). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* 17, 349–356. doi: 10.1093/protein/gzh037
- Bendtsen, J. D., Nielsen, H., Von Heijne, G., and Brunak, S. (2004b). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795. doi: 10.1016/j.jmb.2004.05.028
- Bringans, S., Hane, J. K., Casey, T., Tan, K. C., Lipscombe, R., Solomon, P. S., et al. (2009). Deep proteogenomics; high throughput gene validation by multidimensional liquid chromatography and mass spectrometry of proteins from the fungal wheat pathogen *Stagonospora nodorum*. *BMC Bioinformatics* 10:301. doi: 10.1186/1471-2105-10-301
- Brown, N. A., Antoniw, J., and Hammond-Kosack, K. E. (2012). The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis. *PLoS ONE* 7:e33731. doi: 10.1371/journal.pone.0033731
- Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K. K., et al. (2011). Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. tritici, the causal agent of wheat stripe rust. *PLoS ONE* 6:e24230. doi: 10.1371/journal.pone.0024230
- Choo, K. H., Tan, T. W., and Ranganathan, S. (2009). A comprehensive assessment of N-terminal signal peptides prediction methods. *BMC Bioinformatics* 10(Suppl 15):S2. doi: 10.1186/1471-2105-10-S15-S2
- Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., et al. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* 13, 414–430. doi: 10.1111/j.1364-3703.2011.00783.x
- Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434, 980–986. doi: 10.1038/nature03449
- de Wit, P. J., van Der Burgt, A., Ökmen, B., Stergiopoulos, I., Abd-Elsalam, K. A., Aerts, A. L., et al. (2012). The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. *PLoS Genet.* 8:e1003088. doi: 10.1371/journal.pgen.1003088
- Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* 11, 539–548. doi: 10.1038/nrg2812
- Duplessis, S., Cuomo, C. A., Lin, Y. C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., et al. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9166–9171. doi: 10.1073/pnas.1019315108
- Emanuelsson, O. (2002). Predicting protein subcellular localisation from amino acid sequence information. *Brief. Bioinformatics* 3, 361–376. doi: 10.1093/bib/3.4.361
- Emanuelsson, O., Brunak, S., Von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2, 953–971. doi: 10.1038/nprot.2007.131
- Emanuelsson, O., Nielsen, H., Brunak, S., and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903
- Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R. A., Henriessat, B., et al. (2012). The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* 336, 1715–1719. doi: 10.1126/science.1221748
- Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., et al. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422, 859–868. doi: 10.1038/nature01554
- Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S., et al. (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438, 1105–1115. doi: 10.1038/nature04341
- Godfrey, D., Bohlenius, H., Pedersen, C., Zhang, Z., Emmersen, J., and Thordal-Christensen, H. (2010). Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *BMC Genomics* 11:317. doi: 10.1186/1471-2164-11-317
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996). Life with 6000 genes. *Science* 274:546, 563–547. doi: 10.1126/science.274.5287.546
- Guida, A., Lindstädt, C., Maguire, S. L., Ding, C., Higgins, D. G., Corton, N. J., et al. (2011). Using RNA-seq to determine the transcriptional landscape and the hypoxic response of the pathogenic yeast *Candida parapsilosis*. *BMC Genomics* 12:628. doi: 10.1186/1471-2164-12-628
- Guyon, K., Balagué, C., Roby, D., and Raffaele, S. (2014). Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. *BMC Genomics* 15:336. doi: 10.1186/1471-2164-15-336
- Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H., Handsaker, R. E., Cano, L. M., et al. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461, 393–398. doi: 10.1038/nature08358
- Hane, J. K., Anderson, J. P., Williams, A. H., Sperschneider, J., and Singh, K. B. (2014). Genome sequencing and comparative genomics of the broad host-range pathogen *Rhizoctonia solani* AG8. *PLoS Genet.* 10:e1004281. doi: 10.1371/journal.pgen.1004281
- Hane, J. K., Lowe, R. G., Solomon, P. S., Tan, K. C., Schoch, C. L., Spatafora, J. W., et al. (2007). Dothideomycete plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. *Plant Cell* 19, 3347–3368. doi: 10.1105/tpc.107.052829
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259
- Käll, L., Krogh, A., and Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* 338, 1027–1036. doi: 10.1016/j.jmb.2004.03.016
- Kamoun, S., Furzer, O., Jones, J. D., Judelson, H. S., Ali, G. S., Dalio, R. J., et al. (2014). The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* 16, 413–434. doi: 10.1111/mpp.12190
- Kämper, J., Kahmann, R., Bolker, M., Ma, L. J., Brefort, T., Saville, B. J., et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444, 97–101. doi: 10.1038/nature05248
- Kemen, E., Gardiner, A., Schultz-Larsen, T., Kemen, A. C., Balmuth, A. L., Robert-Seilaniantz, A., et al. (2011). Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biol.* 9:e1001094. doi: 10.1371/journal.pbio.1001094
- Klee, E. W., and Ellis, L. B. (2005). Evaluating eukaryotic secreted protein prediction. *BMC Bioinformatics* 6:256. doi: 10.1186/1471-2105-6-256
- Kloppholz, S., Kuhn, H., and Requena, N. (2011). A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr. Biol.* 21, 1204–1209. doi: 10.1016/j.cub.2011.06.044
- Klosterman, S. J., Subbarao, K. V., Kang, S., Veronese, P., Gold, S. E., Thomma, B. P., et al. (2011). Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog.* 7:e1002137. doi: 10.1371/journal.ppat.1002137
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Lévesque, C. A., Brouwer, H., Cano, L., Hamilton, J. P., Holt, C., Huitema, E., et al. (2010). Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol.* 11, R73. doi: 10.1186/gb-2010-11-7-r73
- Links, M. G., Holub, E., Jiang, R. H., Sharpe, A. G., Hegedus, D., Beynon, E., et al. (2011). *De novo* sequence assembly of *Albugo candida* reveals a small genome relative to other biotrophic oomycetes. *BMC Genomics* 12:503. doi: 10.1186/1471-2164-12-503
- Liu, T., Song, T., Zhang, X., Yuan, H., Su, L., Li, W., et al. (2014). Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis. *Nat. Commun.* 5, 4686. doi: 10.1038/ncomms5686
- Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., et al. (2015). Fungal effectors and plant susceptibility. *Annu. Rev. Plant Biol.* 66, 513–545. doi: 10.1146/annurev-arplant-043014-114623

- Lowe, R. G., and Howlett, B. J. (2012). Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi. *PLoS Pathog.* 8:e1002515. doi: 10.1371/journal.ppat.1002515
- Ma, L. J., Van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M. J., Di Pietro, A., et al. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464, 367–373. doi: 10.1038/nature08850
- Machida, M., Asai, K., Sano, M., Tanaka, T., Kumagai, T., Terai, G., et al. (2005). Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* 438, 1157–1161. doi: 10.1038/nature04300
- Manning, V. A., Pandelova, I., Dhillon, B., Wilhelm, L. J., Goodwin, S. B., Berlin, A. M., et al. (2013). Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. *G3 (Bethesda)*, 3, 41–63. doi: 10.1534/g3.112.004044
- Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E. G., Duchaussoy, F., et al. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452, 88–92. doi: 10.1038/nature06556
- Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., et al. (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat. Biotechnol.* 26, 553–560. doi: 10.1038/nbt1403
- Meijer, H. J., Mancuso, F. M., Espadas, G., Seidl, M. F., Chiva, C., Govers, F., et al. (2014). Profiling the secretome and extracellular proteome of the potato late blight pathogen *Phytophthora infestans*. *Mol. Cell. Proteomics* 13, 2101–2113. doi: 10.1074/mcp.M113.035873
- Melhem, H., Min, X. J., and Butler, G. (2013). “The impact of SignalP 4.0 on the prediction of secreted proteins,” in *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2013 IEEE Symposium on: IEEE* (Singapore), 16–22.
- Menne, K. M., Hermjakob, H., and Apweiler, R. (2000). A comparison of signal sequence prediction methods using a test set of signal peptides. *Bioinformatics* 16, 741–742. doi: 10.1093/bioinformatics/16.8.741
- Min, X. J. (2010). Evaluation of computational methods for secreted protein prediction in different eukaryotes. *J. Proteomics Bioinform.* 3, 143–147. doi: 10.4172/jpb.1000133
- Morin, E., Kohler, A., Baker, A. R., Foulongne-Oriol, M., Lombard, V., Nagy, L. G., et al. (2012). Genome sequence of the button mushroom *Agaricus bisporus* reveals mechanisms governing adaptation to a humic-rich ecological niche. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17501–17506. doi: 10.1073/pnas.1206847109
- Nemri, A., Saunders, D. G. O., Anderson, C., Upadhyaya, N., Win, J., Lawrence, G. J., et al. (2014). The genome sequence and effector complement of the flax rust pathogen *Melampsora lini*. *Front. Plant Sci.* 5:98. doi: 10.3389/fpls.2014.00098
- Nielsen, H., Brunak, S., and von Heijne, G. (1999). Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* 12, 3–9. doi: 10.1093/protein/12.1.3
- Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1–6. doi: 10.1093/protein/10.1.1
- Nielsen, H., and Krogh, A. (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6, 122–130.
- O’Connell, R. J., Thon, M. R., Hacquard, S., Amyotte, S. G., Kleemann, J., Torres, M. F., et al. (2012). Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat. Genet.* 44, 1060–1065. doi: 10.1038/ng.2372
- Ohm, R. A., Feu, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., et al. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. *PLoS Pathog.* 8:e1003037. doi: 10.1371/journal.ppat.1003037
- Ohm, R. A., Riley, R., Salamov, A., Min, B., Choi, I. G., and Grigoriev, I. V. (2014). Genomics of wood-degrading fungi. *Fungal Genet. Biol.* 72, 82–90. doi: 10.1016/j.fgb.2014.05.001
- Paper, J. M., Scott-Craig, J. S., Adhikari, N. D., Cuomo, C. A., and Walton, J. D. (2007). Comparative proteomics of extracellular proteins *in vitro* and in planta from the pathogenic fungus *Fusarium graminearum*. *Proteomics* 7, 3171–3183. doi: 10.1002/pmic.200700184
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Petre, B., and Kamoun, S. (2014). How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biol.* 12:e1001801. doi: 10.1371/journal.pbio.1001801
- Poppe, S., Dorsheimer, L., Happel, P., and Stukenbrock, E. H. (2015). Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). *PLoS Pathog.* 11:e1005055. doi: 10.1371/journal.ppat.1005055
- Raffaele, S., and Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* 10, 417–430. doi: 10.1038/nrmicro2790
- Raffaele, S., Win, J., Cano, L. M., and Kamoun, S. (2010). Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. *BMC Genomics* 11:637. doi: 10.1186/1471-2164-11-637
- Reid, I., O’toole, N., Zabaneh, O., Nourzadeh, R., Dahdouli, M., Abdellatef, M., et al. (2014). SnowyOwl: accurate prediction of fungal genes by using RNA-Seq and homology information to select among ab initio models. *BMC Bioinformatics* 15:229. doi: 10.1186/1471-2105-15-229
- Ridout, C. J., Skamnioti, P., Porritt, O., Sacristan, S., Jones, J. D., and Brown, J. K. (2006). Multiple avirulence paralogues in cereal powdery mildew fungi may contribute to parasite fitness and defeat of plant resistance. *Plant Cell* 18, 2402–2414. doi: 10.1105/tpc.106.043307
- Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., et al. (2011). Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat. Commun.* 2, 202. doi: 10.1038/ncomms1189
- Saunders, D. G., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S., and Raffaele, S. (2012). Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS ONE* 7:e29847. doi: 10.1371/journal.pone.0029847
- Schornack, S., Van Damme, M., Bozkurt, T. O., Cano, L. M., Smoker, M., Thines, M., et al. (2010). Ancient class of translocated oomycete effectors targets the host nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17421–17426. doi: 10.1073/pnas.1008491107
- Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stuber, K., et al. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* 330, 1543–1546. doi: 10.1126/science.1194573
- Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., Manners, J. M., and Taylor, J. M. (2015). Advances and challenges in computational prediction of effectors from plant pathogenic fungi. *PLoS Pathog.* 11:e1004806. doi: 10.1371/journal.ppat.1004806
- Stajich, J. E., Wilke, S. K., Ahrén, D., Au, C. H., Birren, B. W., Borodovsky, M., et al. (2010). Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc. Natl. Acad. Sci. U.S.A.* 107, 11889–11894. doi: 10.1073/pnas.1003391107
- Testa, A. C., Hane, J. K., Ellwood, S. R., and Oliver, R. P. (2015). CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* 16:170. doi: 10.1186/s12864-015-1344-4
- Torto, T. A., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N. A., et al. (2003). EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Res.* 13, 1675–1685. doi: 10.1101/gr.910003
- Tyler, B. M., Kale, S. D., Wang, Q., Tao, K., Clark, H. R., Drews, K., et al. (2013). Microbe-independent entry of oomycete RxLR effectors and fungal RxLR-like effectors into plant and animal cells is specific and reproducible. *Mol. Plant Microbe Interact.* 26, 611–616. doi: 10.1094/MPMI-02-13-0051-1A
- Urban, M., Pant, R., Raghunath, A., Irvine, A. G., Pedro, H., and Hammond-Kosack, K. E. (2015). The Pathogen-Host Interactions database (PHI-base): additions and future developments. *Nucleic Acids Res.* 43, D645–D655. doi: 10.1093/nar/gku1165
- Von Heijne, G. (1990). The signal peptide. *J. Membr. Biol.* 115, 195–201. doi: 10.1007/BF01868635
- Wang, L., Jiang, N., Wang, L., Fang, O., Leach, L. J., Hu, X., et al. (2014). 3’ Untranslated regions mediate transcriptional interference between convergent

- genes both locally and ectopically in *Saccharomyces cerevisiae*. *PLoS Genet.* 10:e1004021. doi: 10.1371/journal.pgen.1004021
- Wawra, S., Djamei, A., Albert, I., Nürnberger, T., Kahmann, R., and Van West, P. (2013). *In vitro* translocation experiments with RxLR-reporter fusion proteins of Avr1b from *Phytophthora sojae* and AVR3a from *Phytophthora infestans* fail to demonstrate specific autonomous uptake in plant and animal cells. *Mol. Plant Microbe Interact.* 26, 528–536. doi: 10.1094/MPMI-08-12-0200-R
- Wicker, T., Oberhaensli, S., Parlange, F., Buchmann, J. P., Shatalina, M., Roffler, S., et al. (2013). The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. *Nat. Genet.* 45, 1092–1096. doi: 10.1038/ng.2704
- Wiemann, P., Sieber, C. M., von Bargen, K. W., Studt, L., Niehaus, E. M., Espino, J. J., et al. (2013). Deciphering the cryptic genome: genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *PLoS Pathog.* 9:e1003475. doi: 10.1371/journal.ppat.1003475
- Yang, F., Jensen, J. D., Svensson, B., Jorgensen, H. J., Collinge, D. B., and Finnie, C. (2012). Secretomics identifies *Fusarium graminearum* proteins involved in the interaction with barley and wheat. *Mol. Plant Pathol.* 13, 445–453. doi: 10.1111/j.1364-3703.2011.00759.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Sperschneider, Williams, Hane, Singh and Taylor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.