



Computational Prediction of Effector Proteins in Fungi: Opportunities and Challenges

Humira Sonah[†], Rupesh K. Deshmukh[†] and Richard R. Bélanger^{*}

Département de Phytologie, Faculté des Sciences de l'Agriculture et de l'Alimentation, Centre de Recherche en Horticulture, Université Laval, Québec, QC, Canada

OPEN ACCESS

Edited by:

Mark Findlay Belmonte,
University of Manitoba, Canada

Reviewed by:

Yusuke Saijo,
Nara Institute of Science
and Technology, Japan
Mark James Barfield,
John Innes Centre, UK

*Correspondence:

Richard R. Bélanger
richard.belanger@fsaa.ulaval.ca

[†]These authors have contributed
equally to this work.

Specialty section:

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

Received: 27 October 2015

Accepted: 23 January 2016

Published: 12 February 2016

Citation:

Sonah H, Deshmukh RK
and Bélanger RR (2016)
Computational Prediction of Effector
Proteins in Fungi: Opportunities
and Challenges.
Front. Plant Sci. 7:126.
doi: 10.3389/fpls.2016.00126

Effector proteins are mostly secretory proteins that stimulate plant infection by manipulating the host response. Identifying fungal effector proteins and understanding their function is of great importance in efforts to curb losses to plant diseases. Recent advances in high-throughput sequencing technologies have facilitated the availability of several fungal genomes and 1000s of transcriptomes. As a result, the growing amount of genomic information has provided great opportunities to identify putative effector proteins in different fungal species. There is little consensus over the annotation and functionality of effector proteins, and mostly small secretory proteins are considered as effector proteins, a concept that tends to overestimate the number of proteins involved in a plant–pathogen interaction. With the characterization of *Avr* genes, criteria for computational prediction of effector proteins are becoming more efficient. There are 100s of tools available for the identification of conserved motifs, signature sequences and structural features in the proteins. Many pipelines and online servers, which combine several tools, are made available to perform genome-wide identification of effector proteins. In this review, available tools and pipelines, their strength and limitations for effective identification of fungal effector proteins are discussed. We also present an exhaustive list of classically secreted proteins along with their key conserved motifs found in 12 common plant pathogens (11 fungi and one oomycete) through an analytical pipeline.

Keywords: computational tool and servers, classification and prediction, effector proteins, fungal secretome, host–pathogen interaction

INTRODUCTION

The importance of fungi as plant pathogens has spurred scientists to study their biology. Fungal pathogens cause enormous yield losses in agricultural crops and post-harvest products (Dean et al., 2012). Generally, the losses caused by pests and diseases are considered to be 20–40% of the total production, and the resulting consequences on human health, the world economy, environmental and ecological losses are significant factors to be considered (Savary et al., 2012; Balint-Kurti and Holland, 2015). To prevent such losses, the use of resistance genes and the application of fungicides are the two major options available for the farmers (Delourme et al., 2006; Dean et al., 2012; Sonah et al., 2012). In the latter case, fungal pathogens are known to quickly develop resistance to most chemicals and the use of fungicides is generally perceived as negative for human health and the

environment (Van de Wouw et al., 2014; Balint-Kurti and Holland, 2015). For this reason, genetic approaches are considered safer and more durable, and considerable efforts are deployed toward the identification and introgression of resistance genes into plant material (Channamallikarjuna et al., 2010; Raman et al., 2012; Singh et al., 2012; Saha et al., 2014). However, the use of a single source of resistance also brings tremendous selection pressure on the pathogen, and the resistance often breaks down quite rapidly (Kutcher et al., 2013; Van de Wouw et al., 2014). For instance, resistance breakdown to the blackleg disease in canola crops has been reported recently in Australia (Van de Wouw et al., 2014). To achieve more durable resistance against a wide range of fungal pathogen races, a thorough understanding of the virulence factors released by the pathogen and the resulting plant immune responses is a prerequisite.

Fungi have adopted diverse strategies to interact with host plants and to overcome a complex network of plant defense mechanisms. The first line of defense involves recognition of the pathogen based on conserved molecular features generally known as pathogen-associated molecular patterns (PAMPs; Silva-Gomesa et al., 2014). The PAMPs, like chitin or glucan residues of fungi, are recognized by plant receptors known as pattern recognition receptors (PRRs). PRRs recognize PAMPs and induce PAMP triggered immunity (PTI) through the secretion of antifungal compounds, production of reactive oxygen species (ROS), phytoalexins, protease inhibitors, chitinases and glucanases. In turn, to overcome PRR responses, pathogens secrete effector molecules, which can lead to plant effector-triggered immunity (ETI; Giraldo and Valent, 2013). The functional and structural alterations in plants caused by effector molecules either facilitate infection by the pathogen through release of virulence factors and toxins, or trigger defense responses based on recognition of avirulence factors and elicitors, or both (Jones and Dangl, 2006; Kamoun, 2006; Morgan and Kamoun, 2007). The effectors are recognized by the specific resistance gene(s) mostly coding for proteins having interactive domains, such as the NB-LRR protein that induces the ETI in plants. Natural selection of pathogens against the resistance pressure applied by ETI involves diversifying unrecognizable effectors (Jones and Dangl, 2006). Such co-evolution of genes involved in plant-pathogen interactions has been previously described by Jones and Dangl (2006) in the form of the simplified and understandable “Zigzag model.” The zigzag model can be summarized with four stages: in the first stage, PRRs recognize PAMPs; in the second stage, to overcome PRR responses, pathogens secrete effectors to interfere with PTI; in the third stage, NB-LRRs recognize effectors; and finally in the fourth stage, diversification and loss or gain of effectors lead to co-evolution.

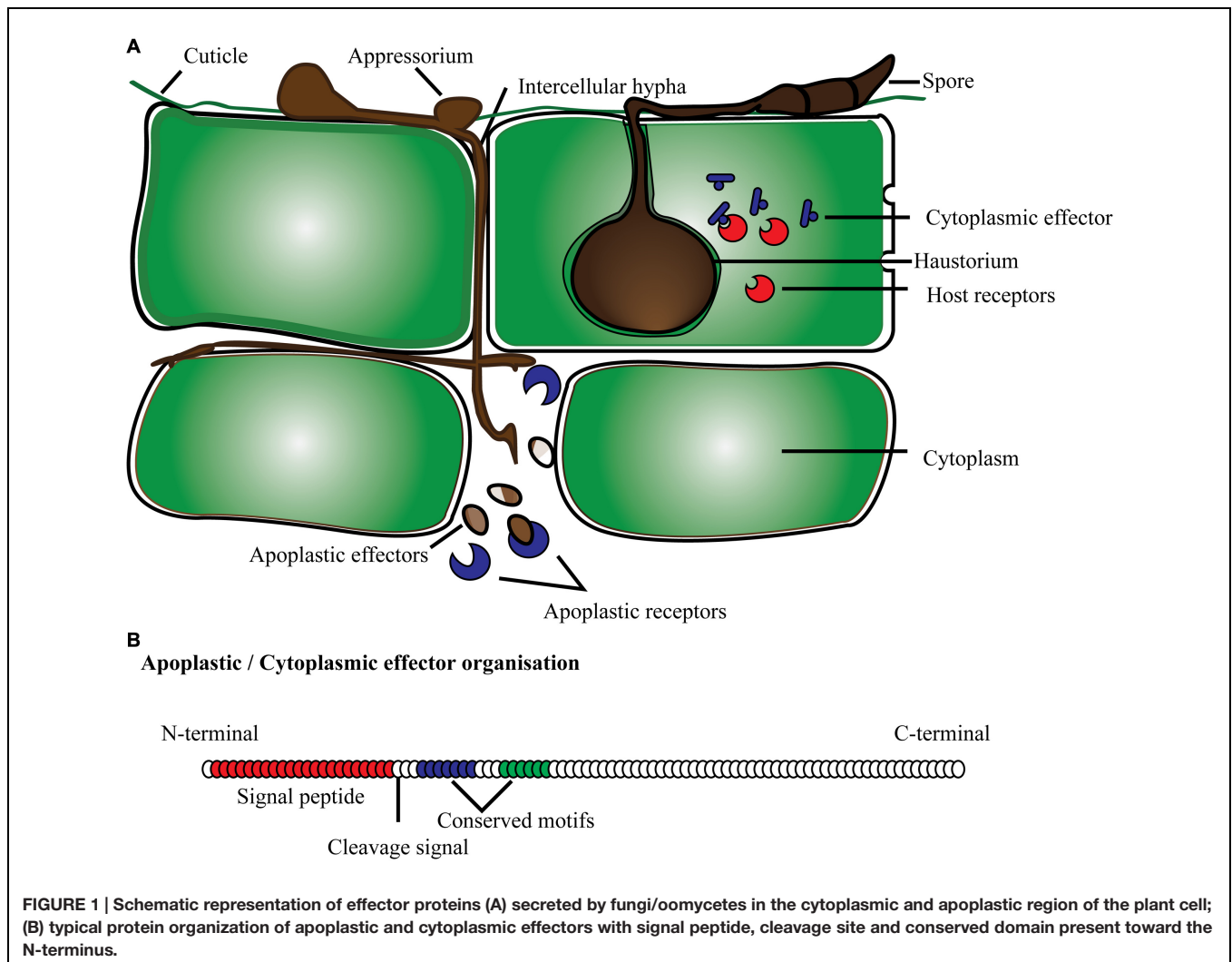
The genes coding for effectors are mostly known as *Avr* genes and the complementary trigger-coded responses by the host are denoted as *R* genes. The ETI involves the hypersensitive response (HR) that restricts pathogen growth. Evolutionary changes in effector (*Avr*) genes make them unrecognizable by the host *R* genes resulting in a compatible interaction, or disease. Since *Avr* genes evolve quickly, they can overcome the plant defense

mechanisms within a short period of time. Therefore, effectors are important targets to consider in attempts to enhance plant immunity against pathogens.

CHARACTERISTICS OF EFFECTOR PROTEINS

The definition of effector is constantly evolving with the increased understanding of the molecular mechanisms involved in pathogenicity. At times, plant pathologists will use the term effector in a broader sense including all molecules, like proteins, carbohydrates, and secondary metabolites, potentially involved in the infection process. Based on a broader definition, PAMPs can also be referred to as effectors (Kamoun, 2006; Nemri et al., 2014).

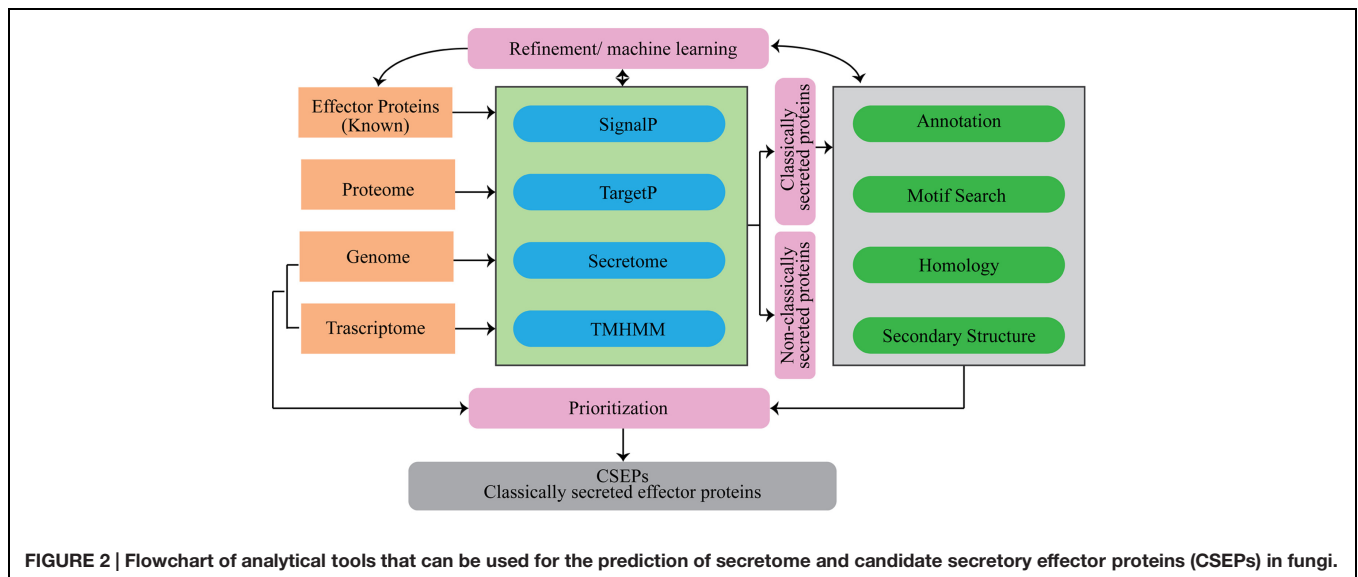
Effector proteins are mostly secretory proteins that alter host cells to suppress host defense mechanisms, and facilitate infection by the pathogen so it can derive nutrients from the host. Effectors may also activate defense strategies in resistant plant genotypes. Criteria to fit the definition of candidate secreted effector proteins (CSEPs) include: fungal proteins with a signal peptide for secretion, no trans-membrane domains, no similarity with other obvious protein domains, fairly small size and mostly species-specific (Jones and Dangl, 2006; Stergiopoulos and de Wit, 2009; Djamei et al., 2011; Lo Presti et al., 2015). In general, effector proteins are modular proteins. Expression of effector proteins follows contact with the host tissue and it is very specific with different stages of disease development. Fungal pathogens have evolved the capacity to deliver effector proteins inside the host cell through diverse mechanisms (Figure 1). They can secrete effector proteins inside the host cytoplasm as well as in the extracellular space, and are subsequently classified as cytoplasmic and apoplastic effectors, respectively. The standard protein organization of apoplastic effectors contains a signal peptide within the initial 60 amino acids (AA) at the N terminus followed by multiple domains toward the C terminus. These types of effectors are comparatively small, and rich in cysteine residues like most of the serine or cysteine protease inhibitor proteins. For instance, known effectors of the tomato fungal pathogen *Cladosporium fulvum* such as *Avr2*, *Avr9*, *Avr4*, and *ECP2*, are small cysteine-rich proteins that are thought to function exclusively in the apoplast (Thomma et al., 2005). The apoplastic effectors of *C. fulvum*, and other fungal and oomycete pathogens have the ability to inhibit and protect against plant hydrolytic enzymes, such as proteases, glucanases, and chitinases (reviewed by Misas-Villamil and van der Hoorn, 2008). Another example is effector protein *SnTox1* identified in the fungal pathogen *Stagonospora nodorum*, which consists of 117 amino acids with the first 17 predicted as a signal peptide and 16 of the remaining 100 amino acids being cysteine residues (Liu et al., 2012). Similarly, cytoplasmic effectors have a secretion signal at the N terminus, and multi-domain toward the C terminus. In addition, conserved amino acid motifs specific to effectors have been reported, namely in oomycetes (Morgan and Kamoun, 2007; Jiang et al., 2008; Ye et al., 2015). The most common



motif, RxLR (arginine, any AA, leucine, arginine), has been identified in over 700 CSEPs predicted in two *Phytophthora* species, *P. sojae* and *P. ramorum* (Jiang et al., 2008). The majority of RxLR carrying effectors also possess a second conserved motif termed dEER (aspartate, glutamate, glutamate, arginine), which is present toward the C-terminus. Similarly, with the increased number of predicted CSEPs, more conserved features may be discovered. A comparative analysis of *Phytophthora* CSEPs has identified three more conserved motifs denoted as W, Y and L toward the C-terminus (Jiang et al., 2008; Win et al., 2012; Wirthmueller et al., 2013). These domains form an alpha-helical fold termed WY fold that is supposed to provide a structure flexibility leading toward the surface diversification of RxLR effectors (Win et al., 2012; Wirthmueller et al., 2013).

The effector protein family encompassing the RxLR motif is found to be the largest among oomycete CSEPs. Even with such a common conserved motif, this CSEP family is very diverse mostly because of high positive selection pressure. Recently, secondary structure analyses of the RxLR effectors have identified abundant

short alpha-helices at the C-terminus in the majority of proteins (Ye et al., 2015). Similarly, de Guillen et al. (2015) have observed common 3-dimensional structures despite a lack of sequence similarity among the AVR1-CO39 and AVR-Pia effectors of *Magnaporthe oryzae*. Structural similarity searches have also succeeded to identify two more effectors, one each from *M. oryzae* (AvrPiz-t), and *Pyrenophora tritici-repentis* (ToxB; de Guillen et al., 2015). The identification of similar secondary or tertiary structures may represent another promising approach to identify functional effectors. The abundant short alpha-helices have also been confirmed in the previously characterized RxLR effectors including PcAvr3a4, PcAvr3a11, PsAvh5, PexRD2, HaATR1, and HaATR13, and also observed in effectors lacking RxLR (Boutemy et al., 2011; Chou et al., 2011; Yaeno et al., 2011; Sun et al., 2013; Ye et al., 2015). The RxLR motif is found to be more common in oomycetes particularly in *Phytophthora* species but is also found, albeit in reduced numbers, in other oomycetes and even in fungal species (Morgan and Kamoun, 2007; Jiang et al., 2008; Ye et al., 2015). This suggests that fungi might contain other functionally important motifs like RxLR, but with a relatively



lower frequency, which makes it difficult to identify based on the degree of conservation. For instance, a highly conserved pattern of seven amino acids “RSIDELD” at the C-terminus (named DELD) has been identified in 25 CSEPs of *Piriformospora indica* (root endophyte; Zuccaro et al., 2011). A total of 107, 178, and 57 CSEPs have been identified in powdery mildew of barley, stem rust, and leaf rust of wheat, respectively, with a conserved motif of three AA in which the first AA is aromatic like tyrosine, phenylalanine or tryptophan, and the last is always a cysteine (Y/F/WxC; Godfrey et al., 2010; Pedersen et al., 2012). This finding suggests that the Y/F/WxC motif containing CSEPs constitutes a new class of effectors that could denote specificity to haustoria-producing pathogenic fungi.

COMPUTATIONAL TOOLS AND PIPELINES AVAILABLE FOR PREDICTION OF CANDIDATE SECRETORY EFFECTOR PROTEINS

Many studies employing computational prediction of CSEPs followed by identification of conserved motifs lack experimental validation of the results (Godfrey et al., 2010; Zuccaro et al., 2011; Ye et al., 2015). Nevertheless, computational prediction serves as an excellent starting point to screen CSEPs for functional analysis and also helps to understand the evolution, distribution and characterisation of effectors.

Several computational tools and web servers are available for the characterization of proteins using the AA sequence as an input. In the case of CSEP prediction, computational tools have been used to systematically sort the list based on some basic pre-established criteria (Figure 2).

Signal Peptide

Commonly, the first step of the CSEP prediction is to look for the extracellular secretion signals. Eukaryotic as well as

prokaryotic proteins usually contain a signal peptide that guides their translocation across the membranes. As a general rule, signal peptides are 20–30 AA in length and they have a positively charged N-terminus, followed by a hydrophobic region, and cleavage site at the C-terminus. In spite of these unique properties, there is limited sequence homology or similarity among signal peptides. Therefore, routine BLAST search alone is not useful for signal peptide prediction and it requires complex analytical algorithms like neural networks, machine learning systems, and Hidden Markov model (HMM). There are several computational tools available that use a combination of different sophisticated algorithms and generally have a very high sensitivity and accuracy for predicting signal peptides (Table 1).

Transmembrane Domains

Distinguishing a secretory protein from a transmembrane (TM) protein is difficult since both have hydrophobic segments. In the case of TM proteins, the hydrophobic segment is usually longer than in the secretory proteins. Therefore, to avoid false positive prediction of secretory proteins, it is always necessary to identify TM domains in candidate proteins. As with signal prediction tools, TM domain prediction tools also use complex algorithms. There are several online tools and web-servers available for the purpose of predicting TM-domains (Table 2). To make prediction of secretory proteins, more sophisticated tools like ProtComp, Phobius, and SPOCTOPUS hosts combine algorithms for TM-domain and signal peptide prediction. Proteins having signal peptides for secretion are not systematically secreted, since some of them may be anchored in the endoplasmic reticulum due to the hydrophobic signal at the C terminus, or the presence of one or more TM domains. Similarly, proteins with glycosylphosphatidylinositol (GPI) anchors stay inserted in the membrane since they have glycolipids attached to the C-terminus (Petersen et al., 2011). Therefore, during secretome analysis, it is always better to predict features like signal-anchors, GPI-anchors, and transit peptides of plastids

TABLE 1 | Features of important tools available for the identification of secretory proteins in fungi and other eukaryotes.

Tool (specification)	Important features	Reference/website
SignalP server	Predicts signal peptide cleavage sites in proteins from different organisms including eukaryotes. Prediction is based on a combination of several artificial neural networks	Petersen et al., 2011 www.cbs.dtu.dk/services/SignalP
Philius server	Predicts topology and signal peptides using dynamic Bayesian networks	Reynolds et al., 2008 http://www.yeastrc.org/philius
SPOCTOPUS server	Combines prediction of signal peptides and membrane protein topology, suitable for genome-scale studies	Viklund et al., 2008 http://octopus.cbr.su.se/
SPSCAN	Predicts secretory signal peptides in protein sequences. For each sequence, provides secretory signal peptides sorted according to score	http://www.csd.hku.hk/bruhk/gcgdoc/spscan.html
Phobius web server	Combines transmembrane (TM) topology and signal peptide predictions and provides optimal choice between TM segments and signal peptides	Käll et al., 2007 http://phobius.binf.ku.dk
ProtComp	Uses neural networks-based prediction; direct comparison with known homologous proteins; comparisons of pentamer distributions calculated for query and DB sequences; prediction of signal peptides, signal-anchors, GPI-anchors, transit peptides of mitochondria, and TM segments	http://linux1.softberry.com
PrediSi (PREDIction of Signal peptides)	Uses position weight matrix approach that is improved by a frequency correction that considers amino acid bias present in proteins. The tool is also trained using a large number of sequences from SwissProt database	Hiller et al., 2004 http://www.predisi.de/home.html
Signal-CF	Uses automated methods that first predict secretory or non-secretory proteins and further identify the cleavage site of the signal peptide	Chou and Shen, 2007 http://chou.med.harvard.edu/bioinf/Signal-CF/
Signal-3L	Consists of three prediction engines for identification of secretory or non-secretory proteins by OET-KNN and PseAA; predicts signal peptide cleavage sites by a subsite-coupled discrimination algorithm; and determines the final cleavage site by fusing the global sequence alignment outcome	Shen and Chou, 2007 http://chou.med.harvard.edu/bioinf/Signal-3L/
WoLF PSORT	Converts protein amino acid sequences into numerical localization features; based on sorting signals, amino acid composition and functional motifs to predict protein subcellular location	Horton et al., 2007 www.wolfpsort.org

along with signal peptides and TM-domains for an effective characterisation of CSEPs.

Secretome

The entire secretome is expectedly not confined to disease-related proteins, and therefore, it needs to be sorted using features that are more specific to CSEPs. To apply different CSEP-specific criteria, several tools need to be applied in a systematic manner. The sequential use of different computation tools to obtain the desired outcome is known as an analytical pipeline. The literature offers a number of analytical pipelines for the identification of CSEPs. Notably, a pipeline based on HMM analyses followed by unsupervised protein clustering has been developed and implemented for the identification of 2830 CSEPs in the cereal pathogen *Fusarium graminearum* (Sperschneider et al., 2013). This pipeline has successfully identified CSEPs, conserved patterns and fungal motifs related to pathogenesis. Similarly, a pipeline developed by Saunders et al. (2012) proposes general basic features expected for the effective identification of CSEPs in rust fungi. The pipeline incorporates six major steps including secretome prediction, grouping of secreted and non-secreted proteins based on Markov clustering, functional annotation based on homology searches, searches for conserved motifs, effector features annotation, and finally hierarchical tribe clustering to rank and classify CSEPs (Saunders et al., 2012). The final ranking based on the fulfillment of different criteria is very helpful for the prioritization of candidates for functional

characterization. In addition, understanding of the secondary and tertiary structure organization of effectors and their counterpart R genes will definitely improve the efficiency of computational tools to identify effectors more precisely (de Guillen et al., 2015; Maqbool et al., 2015; Ye et al., 2015).

DIFFERENT CONSERVED MOTIFS IDENTIFIED IN FUNGAL GENOMES WITH COMPUTATIONAL MINING

Amino acid sequences of functionally important motifs in CSEPs appear to be conserved across the fungal/omycete species. Therefore to understand the function of a given protein, analysis of such conserved motifs is required. Several reports have identified conserved motifs in effectors, namely in oomycetes, and validated their functionality (Morgan and Kamoun, 2007; Jiang et al., 2008; Godfrey et al., 2010; Zuccaro et al., 2011). The conserved motifs are found to play an important role in delivering effector proteins more efficiently during pathogenesis (Kale and Tyler, 2011; Petre and Kamoun, 2014). Natural variants of motif sequences, or variants created using mutagenesis, have been routinely evaluated with different approaches to confirm the functional role of the motifs. Plant transient-expression systems, in which candidate effectors are expressed in the plant and the translated protein observed for its secretion and re-entry into the plant cell, are commonly used to demonstrate the functional role

TABLE 2 | Features of the most common computational tools available for the prediction of trans-membrane (TM) domains.

Tools	Important features	Website/reference
TMHMM Server v. 2.0	Predicts trans-membrane (TM) domains; a common mistake by the program consists in reversing the direction of proteins with one TM segment	www.cbs.dtu.dk/services/TMHMM/
TMpred	Predicts TM regions and orientation The TMpred program makes a prediction of membrane-spanning regions and their orientation	www.ch.embnet.org Hofman, 1993
TMbase	Offers a good database of TM proteins and their helical membrane-spanning domains; TMbase was originally meant as a tool for analyzing the properties of TM proteins	www.ch.embnet.org Hofman, 1993
HMMTOP	Serves as automatic server for predicting TM helices and topology of proteins	Tusnady and Simon, 2001 www.enzim.hu/hmmtop/
PredictProtein	Predicts secondary structures	Rost et al., 2004 www.predictprotein.org/
SOSUI	Classifies and predicts secondary structures of membrane proteins	http://harrier.nagahama-i-bio.ac.jp/sosui/
TopPred 1.10	Predicts topology of membrane proteins	http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::toppred
DAS-TMfilter server	Filters false positive TM protein predictions	http://mendel.imp.ac.at/sat/DAS/DAS.html
CCTOP (Consensus Constrained TOPology prediction)	Performs TM topology predictions	http://cctop.enzim.ttk.mta.hu/
MetaTM	Predicts TM topologies through a consensus method	Klammer et al., 2009 http://metatm.sbc.su.se/
MINNOU	Predicts membrane proteins with and without explicit use of hydropathy profiles and alignments	http://minnou.cchmc.org/
PHDhtm	Predicts the location of helical TM segments in integral membrane proteins through a neural network system	https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_htm.html
Phobius	Combines TM topology and signal peptide predictions	Käll et al., 2007 http://phobius.sbc.su.se/poly.html
PRED-TMR	Refines a standard hydrophobicity with a detection of potential edges	Pasquier et al., 1999 http://athina.biol.uoa.gr/
SCAMPI	Uses position-specific amino acid contributions to the free energy of membrane insertion and also uses the current best statistics-based topology predictors	Bernsel et al., 2008 http://scampi.cbr.su.se/
SOMRuler	Offers an interpretable TM helices predictor	Yu et al., 2011 www.csbio.sjtu.edu.cn/bioinf/SOMRuler/
ConPred	Predicts TM topology based on a consensus approach by combining the results of several methods	Arai et al., 2004 http://bioinfo.si.hirosaki-u.ac.jp/~ConPred2/
TMBB-DB	Compiles the predictions made by the Freeman–Wimley algorithm	http://beta-barrel.tulane.edu/
TMalphaDB	Quantifies the structural distortion induced by a sequence motif in alpha TM segments	http://lmc.uab.cat/TMalphaDB/
TMexpo	Predicts rotational preferences of TM helices to facilitate structural modeling. TMexpo calculates rotational angles of TMHs based on the predicted relative accessible surface area	http://bio-cluster.iis.sinica.edu.tw/TMexpo/
TMMOD	Uses an improved hidden Markov model for the identification and topology prediction of TM proteins	http://liao.cis.udel.edu/website/servers/TMMOD/scripts/frame.php?p=submit
Asymmetric Ez	Assesses the energy and positions of protein sequences or structures in and on the membrane through a knowledge-based potential	http://ez.degradolab.org/ez/

of a motif and/or an effector (Kale and Tyler, 2011). Another approach consists in the application of purified effector proteins to leaf or root segments, where the entry of proteins into the cell is observed with the help of fluorescent peptide tags or by the use of antibodies (Kale and Tyler, 2011; Tanaka et al., 2015).

Several conserved motifs observed in oomycetes have also been found in different fungal genomes (Table 3). A systematic similarity search performed in secretomes of 11 fungi and one oomycete species, representing some of the most devastating plant pathogens, has shown the presence of different conserved

TABLE 3 | Number of proteins, classically and small secreted proteins, and proteins bearing known conserved motifs identified in the genomes of 11 fungal and oomycete pathogens of crop plants.

S.No.	Species [†]	Total proteins	Classically secreted proteins [#]	Small secreted proteins [*]	Known conserved motif [§]					
					RxLR	DEER	RxLX [EDQ]	[KRHQSA] [DENQJEL]	[Y/W]xC	RSIVEQD
1	<i>Leptosphaeria maculans</i>	12469	552	263	12	3	16	8	22	14
2	<i>Magnaporthe oryzae</i>	12755	983	528	43	2	22	16	43	27
3	<i>Ustilago maydis</i>	6522	358	142	21	0	13	13	5	10
4	<i>Puccinia graminis f. sp. tritici</i>	15979	886	612	17	2	13	14	43	14
5	<i>Cladosporium fulvum</i>	14127	711	296	17	3	17	17	20	19
6	<i>Fusarium oxysporum</i>	17696	824	361	17	3	19	17	44	30
7	<i>Mycosphaerella graminicola</i>	10933	518	235	14	1	6	8	22	20
8	<i>Colletotrichum graminicola</i>	12020	829	352	19	2	20	20	39	20
9	<i>Blumeria graminis</i>	6470	247	143	5	0	3	6	7	2
10	<i>Alternaria brassicicola</i>	10688	508	228	11	3	12	19	23	7
11	<i>Pyrenophora tritici-repentis</i>	12169	669	328	12	6	10	18	33	17
12	<i>Phytophthora infestans</i> (Oomycete)	18140	671	343	74	22	16	10	18	11

[†]Details of protein sequences retrieved from different databases are provided in Supplementary Table S1; [#]The analytical pipeline (Supplementary Figure S1) was used for the identification of classically secreted proteins (Cortázar et al., 2014); ^{*}Classically secreted proteins having less than 300 amino acid length; [§]Previously known Motif was used as query to perform similarity based motif-search using FIMO software tool implemented in MEME suite (Grant et al., 2011). To claim a significant match, an E-value cut-off at 0.0001 was used.

motifs (Table 3, Supplementary Table S1, Supplementary Figure S1). Most of the conserved motifs identified to date, such as RxLR and DEER are small in length. Consequently, there are more chances to identify false positives of such motifs when using a similarity-based search. For example when we performed a similarity search using the FIMO software tool with an E-value cut-off at 0.001 (Grant et al., 2011), we found four times more CSEPs with a RxLR motif in *Magnaporthe grisea* than we did by using a more stringent cut-off at 0.0001 (Table 3, Supplementary Table S2). By using similar stringent conditions, we still observed the presence of the RxLR motif in all fungal secretomes studied, although with a considerably lower number than in *Phytophthora infestans*. The presence of a functional RxLR motif in a fungal genome has been debated since it is not as abundant as in the oomycetes. However, effector re-entry assays performed with Avr2 (*Fusarium oxysporum*) and AvrLm6 (*Leptosphaeria maculans*) have shown loss of functionality when mutations were made in RxLR-like motifs (Kale et al., 2010; Kale and Tyler, 2011). This suggests that RxLR-like motifs, in spite of their low occurrence, have a functional role in fungal effectors, and similar findings are expected for other motifs like DEER, [KRHQSA][DENQ]EL, [Y/W]xC, and RSIVEQD. Interestingly, unlike RxLR, we found that the motif RxLx[EDQ] occurred with a similar frequency in both fungal and oomycete secretomes (Table 3).

SECRETORY PROTEINS AND CANDIDATE SECRETORY EFFECTOR PROTEIN (CSEP) DATABASES

Numerous accessible online databases have been developed to provide a catalog of well-characterized predicted secretory proteins and publically available CSEPs (Table 4). For instance,

the Fungal Secretome Database (FSD) comprises predicted secretory proteins from 158 fungal/oomycete genomes. FSD relies on nine different prediction programs to build its inventory, namely SignalP 3.0, SigCleave, SigPred, RPSP, TMHMM 2.0c, TargetP 1.1b, PSort II, SecretomeP 1.0, and predictNLS (Choi et al., 2010). This secretome resource is very useful to identify and characterize species-specific conserved motifs. For instance, 734 putative RxLR effectors have been identified from three *Phytophthora* species, data that are well-correlated with those previously reported by Jiang et al. (2008) in the same species. Interestingly, the RxLR motif was observed with a very low frequency (0.04%) in the other 153 fungal genomes (Choi et al., 2010). This finding is surprising since many more fungal genomes have been observed to have a much higher number of RxLR and RxLR-like effectors (Table 3). While there is no doubt that the RxLR motif is more abundant and conserved in oomycetes, and more particularly in *Phytophthora* species, these observations raise interesting questions about the evolution, transfer specificity and functionality of RxLR effectors.

Another useful database for CSEPs is FunSecKB, which hosts fungal secretomes identified using six different prediction tools (Lum and Min, 2011). The improved version of FunSecKB comprises about two million proteins covering over 200 fungal species (Meinken et al., 2014). This massive data has enabled to answer several questions regarding the frequency and distribution of secretory proteins in fungi. For instance, Meinken et al. (2014) have observed that fungi with a biphasic lifestyle, such as the hemibiotroph *M. grisea*, have a larger proportion of secreted proteins compared to strict biotrophs or facultative parasites. In general, the size of the secretome is highly correlated with the total size of the proteome.

The accuracy of computation prediction always depends upon functionally validated data used for the training of prediction tools. The mere use of a larger number of tools is not

sufficient to achieve higher sensitivity and accuracy. In this context, manual curation and the continuous use of the growing number of experimentally validated protein database should lead to more accurate predictions. In an effort to develop a library of fungal stress response database (FSRD), about 2000 publications, sorted systematically from the PubMed entries, have been used to obtain and define over 2000 stress-related proteins in fungi (Karányi et al., 2013). For the FSRD, care has been taken to avoid including proteins labeled as putative (identified based strictly with computational tools) and to include only genuine proteins characterized experimentally. In spite of this screening procedure, a homology-based search led to the identification of over 29,000 orthologs in 28 fungal/oomycete species (Karányi et al., 2013). Similarly, *in silico* identification of small secretory proteins with several tools, followed by manual curation and homology-based search has identified 1184 and 1066 CSEPs respectively in *Melampsora larici-populina* and *Puccinia graminis* (Duplessis et al., 2011). Considering that, in well-studied fungi such as *Ustilago maydis*, functional studies through gene knockout have identified less than 100 CSEPs (Kämper et al., 2006), it appears that the strategy of identification of homologs using manually verified list of CSEPs, where over 1000 CSEPs per species are predicted, greatly overestimates the number of *bona fide* CSEPs. Therefore, to avoid the identification of false positives, more computational filters should be applied. In this context, a pathogen–host interaction database (PHI-base) has been developed based on functionally characterized proteins involved in disease and initiation of host responses (Winnenburg et al., 2008). The PHI-base initially comprised 405 experimentally verified proteins related to pathogenicity, virulence, and effectors belonging to 54 fungal and oomycete pathogens (Winnenburg et al., 2008). The current version of PHI-base (v 3.6) now

comprises about 3000 genes from 4000 interactions, and 160 species including 103 plant pathogens, along with information extracted from 1243 high quality publications (Urban et al., 2014). Such manual curation process and use of experimental studies should be considered along with computational tools to improve the prediction of functional effector proteins.

GENOME-WIDE IDENTIFICATION OF CANDIDATE SECRETORY PROTEINS (CSEPS)

Recent advances in computational tools have made it easier to perform genome-wide identification of CSEPs. However, this approach can often be overlooked considering that several databases hosting predicted secretomes in 100s of fungal and oomycete species are now easily accessible. An obvious drawback to relying on this information is that most of the databases only offer a listing of the secreted proteins with no further characterization of their function or possible role as CSEPs (Table 4). Moreover, genome-wide studies provide a better understanding of the distribution and organization of CSEPs within a given species. The characterization of CSEPs in *U. maydis* represents a very good example of the importance of genome-wide analysis. Following whole genome sequencing of *U. maydis*, 426 secretory proteins were identified, 70% of which were annotated with unknown function (based on homology search; Kämper et al., 2006). Of particular importance, most of the *U. maydis* secreted proteins were found to be present in clusters with 3–26 genes per cluster. Knockout of specific genes or clusters allowed a precise identification of about 50 secreted proteins that were involved in pathogenesis (Kämper et al., 2006).

TABLE 4 | Features of databases available for effectors, secreted proteins and virulence factors identified in fungal genomes.

Database (specificity)	Important features	Reference/website
DFVF (Database of fungal virulence factors)	Comprises host fungal virulence factors, including 2058 pathogenic genes produced by 228 fungal strains from 85 genera	http://sysbio.unl.edu/DFVF/ Lu et al., 2012
FSD (Integrated platform for annotation of fungal secretomes)	Comprises putative secretory proteins in 158 fungal/oomycete genomes (208,883 proteins) identified using a three-layer hierarchical identification rule based on nine prediction programs	http://fsd.snu.ac.kr/ Choi et al., 2010
FSRD (Fungal stress response database)	Incorporates 1985 fungal stress response proteins with verified physiological function(s) and their orthologs identified and annotated in 28 species including human and plant pathogens, as well as important industrial fungi	http://internal.med.unideb.hu/fsrd Karányi et al., 2013
PHI-base (Pathogen–host interaction database)	Provides catalogs of experimentally verified pathogenicity, virulence, and effector genes from bacterial, fungal, and oomycete pathogens, which infect human, animal, plant, insect, fish, and fungal hosts	http://www.phi-base.org/ Winnenburg et al., 2008
FunSecKB (The Fungal Secretome KnowledgeBase)	Provides secretory proteins identified from all available fungal protein data in the NCBI RefSeq database The secreted proteins were identified using several computational tools	Lum and Min, 2011 http://proteomics.ysu.edu/secretomes/fungi.php
FunSecKB2 (Fungal protein subcellular location knowledgebase)	Provides an improved and updated version of the fungal secretome and subcellular proteome, i.e., protein subcellular location, knowledgebase	Meinken et al., 2014 http://proteomics.ysu.edu
Secretool	Secretool comprises a group of web tools that enable secretome predictions	Cortázar et al., 2014 http://genomics.cicbiogune.es/SECRETTOOL

In a comparative analysis with other pathogenic Ustilaginales and *Pseudozyma flocculosa*, a non-pathogenic Ustilaginale with biocontrol properties, whole-genome-sequencing revealed a higher conservation of virulent secreted proteins in the three pathogens and a near complete loss in *P. flocculosa* (Lefebvre et al., 2013). In depth analysis of *P. flocculosa* genome revealed that predicted secreted proteins were nearly the same in both *P. flocculosa* and *U. maydis* genome and that the total number of clusters and gene organization of secreted proteins were also quite similar. This approach was thus extremely useful in not only corroborating the secreted proteins involved in virulence in *U. maydis* but also in identifying potential factors involved in the biocontrol properties of *P. flocculosa*. For instance, the presence of two NPP1-containing proteins in the secretome of *P. flocculosa*, absent in all pathogenic Ustilaginales, offers good targets to understand its elusive mode of action. Other striking features, such as introns per gene, have been observed to vary considerably between the two groups (Lefebvre et al., 2013). The role of intron frequency in the structural and functional attributes of genomes has already been suggested in several fungal and plant genomes (Torriani et al., 2011; Deshmukh et al., 2015). Similarly, in addition to the presence of effectors, many other genomic features like GC content, codon bias, gene gain-loss, and in-depth analysis of gene families can be addressed with genome-wide analyses.

OVERVIEW OF CANDIDATE SECRETORY EFFECTOR PROTEINS IN BIOTROPHS AND HEMIBIOTROPHS

The biotrophic fungus *U. maydis* is arguably one of the best model pathogens for the study of host–pathogen interactions and molecular mechanisms involved in pathogenesis (Kämper et al., 2006). Its well-annotated genome, and advanced tools for transformation and genome manipulation make it suitable for functional characterization of putative effectors (Kämper et al., 2006; Schuster et al., 2015). In fact, the effector Pep1 is one of the best studied virulence-related proteins for its role in the *U. maydis*-maize interaction. Pep1 inhibits plant peroxidases and suppresses the primary immune response by preventing the oxidative burst. The initial colonization of biotrophs requires a suppression of the immune response in order to interface with its host and acquire nutrients. It has been observed, with confocal microscopy, global expression profiling and metabolic profiling, that *U. maydis* will initially up-regulate defense-response related genes, but, after penetration, will down-regulate the early response genes and also induce genes associated with suppression of cell death (Doehlemann et al., 2008). In mutant *U. maydis* strains with *pep1* gene deletion, no down-regulation of the early response genes was observed (Doehlemann et al., 2009). *U. maydis* was also found to induce genes involved in the synthesis of jasmonic acid but to repress salicylic acid synthesis, a typical response generally observed with biotrophs. Such response was not observed in *U. maydis* Pep1 deletion strain (Doehlemann et al., 2009). Recently, Hemetsberger et al. (2015)

identified Pep1 orthologs in genomes of related smut species and performed functional characterization of orthologs by heterologous expression in *U. hordei* and *Hordeum vulgare*. Heterologous expression of Pep1 in *U. hordei* conferred a higher virulence to the mutant strain compared to the wild type. Conversely, heterologous expression of Pep1 in *H. vulgare* was found to increase its susceptibility against the powdery mildew fungus *Blumeria graminis* f. sp. *hordei*, a completely different pathosystem than the maize-*U. maydis*. This suggests the functional conservation of the Pep1 effector across and against different monocots. The high level of sequence conservation suggests the pivotal role of Pep1-like effectors in the pathogenicity of biotrophic fungi. The functional redundancy of Pep1-like effectors has also been observed in pathogens of diverse hosts, both monocots and dicots (Hemetsberger et al., 2015).

Because of their combined biotrophic and necrotrophic lifestyles, hemibiotrophs also produce effectors to suppress early defense responses and maintain their host alive by preventing cell death. At later stages of infection, hemibiotrophs are reported to produce necrotrophic effectors that kill the host. For instance, *P. infestans* secretes AVR3a from its haustoria during the early biotrophic infection stages that suppress cell-death (Whisson et al., 2007). Later in the necrotrophic stages, AVR3a is found to be down-regulated, while INF1 and Nep1-like effectors are secreted, which helps the pathogen to switch from a biotrophic to a necrotrophic stage (Kanneganti et al., 2006).

EFFECTORS IN BACTERIA, NEMATODES, AND INSECTS

Compared to fungi and oomycetes, bacteria have received considerably more attention with respect to understanding the role of effectors in pathogenicity. Progress has been achieved mostly with the characterization of effectors in gram-negative bacteria that deliver effectors into the host cell by type III (T3SS) or type IV secretion systems (Angot et al., 2007). The whole genome sequencing of 1000s of bacterial isolates and identification of effectors have been used to develop effective computational tools for their prediction (Table 5). As a matter of fact, the tools for bacterial effector identification seem more accurate compared to those for fungal effectors. Recently, Teper et al. (2015) used a machine learning algorithm based on 79 features differentiating effector proteins from non-effector proteins to identify novel effectors. The features used for the development of the machine learning approach include several characteristics such as genomic proximity to other effectors, GC content, differential conservation among phytopathogens that do or do not encode a T3S system, amino acid composition at the N-terminus and in the entire protein, T3S-dependent regulation, homology to known T3S effectors of animal- and plant-pathogenic bacteria and similarity to host proteins. After validation of candidate effectors identified in the first round of machine learning, new information is incorporated for the second round of analysis (Teper et al., 2015). Such self-evolving computational approach would also be helpful to identify CSEPs

TABLE 5 | Features of databases and tools available for the effectors, secreted proteins, and virulence factors identified in bacterial genomes.

Database (specificity)	Important features	Reference/website
AtlasT4SS [Database for type IV secretion system (T4SS) related proteins]	Describes a large number of proteins related to the type IV secretion system reported in Gram-negative, Gram-positive bacteria, and Archaea	Souza et al., 2012 http://www.t4ss.incc.br/
DBSecSys (Database of <i>Burkholderia mallei</i> Secretion Systems)	Provides a database for computationally predicted bacterial secretion system proteins and their host factors	Memi et al., 2014
LAB-Secretome [Database for the extracellular and surface-associated proteins of Lactic Acid Bacteria (LAB)]	Provides a database comparison of the secretomes from 26 sequenced LAB genomes	Zhou et al., 2010 www.cmbi.ru.nl/lab_secretome/
Mycosec (A database for signal peptide bearing genes of <i>Mycobacterium</i>)	Provides computationally predicted signal peptides in the 21 different strains of <i>Mycobacterium</i> genome	Sen, 2011 www.bicbu.in/mycosec/
SecReT4	Provides a web-based bacterial type IV secretion system resource. T4SSs are versatile assemblages that promote genetic exchange and/or effector translocation with consequent impacts on pathogenesis and genome plasticity	Bi et al., 2012 http://db-mm1.sjtu.edu.cn/SecReT4/
BEAN (Bacterial effector analyzer)	Provides an integrated web resource to predict, analyze, and store type III secreted effectors (T3SEs)	Dong et al., 2013 http://protein.cau.edu.cn:8080/bean/ .
BPBAac (Bacterial type III secreted protein identifier)	Provides a tool that was developed and trained using Support Vector Machine (SVM) based on the Aac feature extracted using a Bi-profile Bayes model. BPBAac classifies T3S and non-T3S proteins very effectively	Wang et al., 2011 http://biocomputer.bio.cuhk.edu.hk/software/BPBAac .
SIEVE Server (Web tool for prediction of type III secreted effectors)	Provides computational prediction of type III and IV secreted effectors in gram-negative bacteria	Samudrala et al., 2009 http://www.sysbep.org/sieve/
T3SEdb [Database for effectors of Type III secretion system (T3SS)]	Provides keyword as well as sequence searches. More than 171 clusters of T3SEs have been detected based on sequence identity comparisons	Tay et al., 2010 http://effectors.bic.nus.edu.sg/T3SEdb/index.php
T3SS effector prediction tool	Presents a signal prediction method together with comprehensive survey of potential type III secretion system effectors extracted from 918 published bacterial genomes	Lower and Schneider, 2009 http://www.modlab.org/

in fungal genomes leading to the identification of more realistic and manageable numbers.

Plant pathogenic nematodes are mostly obligate parasites and depend on living host cells for nutrition. The plant response to nematode presence is genetically similar to the one observed with fungal and bacterial pathogens. Gene for gene evolution is well-documented in the case of nematode resistance and several *Avr* genes and corresponding *R* genes are known (Woo et al., 2014; Kadam et al., 2015; Vuong et al., 2015). Nematodes release degrading enzymes and peptides that mimic plant hormones into the apoplast to make feeding sites by modifying the host cells. The nematode proteins are secreted from specific glands and those are key for the pathogenesis process, in a manner very similar to that observed with the bacterial and fungal effector systems (Mitchum et al., 2013). As a matter of fact, nematode effectors may have evolved after horizontal transfer from bacteria and fungi (Haegeman et al., 2011). Presently very little is known about the specific characteristics of nematode effectors, and as a result, reliable computational tools are more limited for CSEP prediction.

Plant–insect interactions are also being investigated in view of the current understanding of effectors in bacterial and

fungal organisms (Stuart, 2015). There are several *Avr* and *R* genes known to dictate plant–insect interactions, and most of these fit well in the gene for gene concept. This suggests the likelihood of molecular mechanisms similar to those found in fungal/bacterial effectors. As with nematodes, horizontal gene transfer from bacteria and fungi has been observed in insects, thereby suggesting a similar process of effector acquisition (Husnik et al., 2013). Plants recognize insects by herbivore-associated molecular patterns (HAMP), similar to PAMPs, which induce an immune response. Insect elicitors are secreted through the saliva at the host–insect interface and induce JA, ethylene and SA biosynthesis, as well as the reactive oxygen burst (Wu and Baldwin, 2010). Such insect recognition and plant response has been observed in *Arabidopsis* in response to proteins present in the green-peach aphid saliva (De Vos and Jander, 2009).

CONCLUDING REMARKS

The rapidly increasing availability of fungal genomes and functionally validated effectors has provided opportunities to improve CSEP identification in many fungal pathogens.

In turn, this has led to the development of a large number of computational tools and pipelines to study CSEPs. Given that each tool or pipeline has its own advantages and limitations, the analytical path proposed in this review (Figure 2) offers a good balance between computational prediction and effector functionality.

Our review also highlights the need to increase the prediction efficiency of functional secreted proteins by continuously fine-tuning tools with every newly characterized effector. In this context, approaches based on machine learning that can integrate all the information generated through phenotypic and genomic data in a very systematic manner will be helpful in improving identification of effectors. In addition, considering that effectors evolve rapidly through gene-for-gene interactions, comparative genome sequencing data analysis can provide useful insights with respect to CSEP identification, origin, functionality, and important structural features. For instance, secondary and tertiary structure information, gene expression data, and information about gene and genomic organization are likely to increase the accuracy with which effectors are identified in fungi and other organisms. Most of the available pipelines and automated servers do not currently integrate such data. Combining available pipelines with the ever increasing structural, genomic and transcriptomic data will lead to a better prioritization strategy where the most promising effectors

can be rapidly targeted for future analyses aimed at a better understanding of pathogenesis processes in plant–pathogen interactions.

AUTHOR CONTRIBUTIONS

HS, RD, RB compiled the data, draw the conclusions and wrote the Manuscript. RB designed and supervised the research.

ACKNOWLEDGMENTS

The authors thank the Agri-Innovation program Growing Forward 2, SaskCanola and Agriculture and Agri-Food Canada and the Canada Research Chair Program for financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00126>

FIGURE S1 | Flowchart of analytical tools implemented in the Secretool (<http://genomics.cicbiogune.es/SECRETOOL/Secretool.php>) that was used for the identification of candidate secretory proteins in different plant fungi/oomycetes in Table 2.

REFERENCES

- Angot, A., Vergunst, A., Genin, S., and Peeters, N. (2007). Exploitation of eukaryotic ubiquitin signaling pathways by effectors translocated by bacterial type III and type IV secretion systems. *PLoS Pathog.* 3:e3. doi: 10.1371/journal.ppat.0030003
- Arai, M., Mitsuke, H., Ikeda, M., Xia, J.-X., Kikuchi, T., Satake, M., et al. (2004). ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability. *Nucleic Acids Res.* 32, W390–W393. doi: 10.1093/nar/gkh380
- Balint-Kurti, P. J., and Holland, J. B. (2015). New insight into a complex plant-fungal pathogen interaction. *Nat. Genet.* 47, 101–103. doi: 10.1038/ng.3203
- Bernsel, A., Viklund, H., Falk, J., Lindahl, E., Von Heijne, G., and Elofsson, A. (2008). Prediction of membrane-protein topology from first principles. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7177–7181. doi: 10.1073/pnas.0711151105
- Bi, D., Liu, L., Tai, C., Deng, Z., Rajakumar, K., and Ou, H.-Y. (2012). SecReT4: a web-based bacterial type IV secretion system resource. *Nucleic Acids Res.* 41, D660–D665. doi: 10.1093/nar/gks1248
- Boutemy, L. S., King, S. R., Win, J., Hughes, R. K., Clarke, T. A., Blumenschein, T. M., et al. (2011). Structures of *Phytophthora* RXLR effector proteins a conserved but adaptable fold underpins functional diversity. *J. Biol. Chem.* 286, 35834–35842. doi: 10.1074/jbc.M111.262303
- Channamallikarjuna, V., Sonah, H., Prasad, M., Rao, G. J., Chand, S., Upreti, H., et al. (2010). Identification of major quantitative trait loci qSBR11-1 for sheath blight resistance in rice. *Mol. Breed.* 25, 155–166. doi: 10.1007/s11032-009-9316-5
- Choi, J., Park, J., Kim, D., Jung, K., Kang, S., and Lee, Y.-H. (2010). Fungal secretome database: integrated platform for annotation of fungal secretomes. *BMC Genomics* 11:105. doi: 10.1186/1471-2164-11-105
- Chou, K.-C., and Shen, H.-B. (2007). Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides. *Biochem. Biophys. Res. Commun.* 357, 633–640. doi: 10.1016/j.bbrc.2007.03.162
- Chou, S., Krasileva, K. V., Holton, J. M., Steinbrenner, A. D., Alber, T., and Staskawicz, B. J. (2011). Hyaloperonospora *arabidopsidis* ATR1 effector is a repeat protein with distributed recognition surfaces. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13323–13328. doi: 10.1073/pnas.1109791108
- Cortázar, A. R., Aransay, A. M., Alfaro, M., Oguiza, J. A., and Lavín, J. L. (2014). SECRETOOL: integrated secretome analysis tool for fungi. *Amino Acids* 46, 471–473. doi: 10.1007/s00726-013-1649-z
- de Guillen, K., Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., and Padilla, A. (2015). Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. *PLoS Pathog.* 11:e1005228. doi: 10.1371/journal.ppat.1005228
- 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
- Delourme, R., Chevre, A., Brun, H., Rouxel, T., Balesdent, M., Dias, J., et al. (2006). Major gene and polygenic resistance to *Leptosphaeria maculans* in oilseed rape (*Brassica napus*). *Euro. J. Plant Pathol.* 114, 41–52. doi: 10.1007/s10658-005-2108-9
- Deshmukh, R. K., Sonah, H., and Singh, N. K. (2015). Intron gain, a dominant evolutionary process supporting high levels of gene expression in rice. *J. Plant Biochem. Biotechnol.* 1–5. doi: 10.1007/s13562-015-0319-5
- De Vos, M., and Jander, G. (2009). Myzus persicae (green peach aphid) salivary components induce defence responses in *Arabidopsis thaliana*. *Plant Cell Environ.* 32, 1548–1560. doi: 10.1111/j.1365-3040.2009.02019.x
- Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J., et al. (2011). Metabolic priming by a secreted fungal effector. *Nature* 478, 395–398. doi: 10.1038/nature10454
- Doehlemann, G., Van Der Linde, K., Aßmann, D., Schwambach, D., Hof, A., Mohanty, A., et al. (2009). Pep1, a secreted effector protein of *Ustilago maydis*, is required for successful invasion of plant cells. *PLoS Pathog.* 5:e1000290. doi: 10.1371/journal.ppat.1000290
- Doehlemann, G., Wahl, R., Horst, R. J., Voll, L. M., Usadel, B., Poree, F., et al. (2008). Reprogramming a maize plant: transcriptional and metabolic changes induced by the fungal biotroph *Ustilago maydis*. *Plant J.* 56, 181–195. doi: 10.1111/j.1365-313X.2008.03590.x

- Dong, X., Zhang, Y.-J., and Zhang, Z. (2013). Using weakly conserved motifs hidden in secretion signals to identify type-III effectors from bacterial pathogen genomes. *PLoS ONE* 8:e56632. doi: 10.1371/journal.pone.0056632
- 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
- Giraldo, M. C., and Valent, B. (2013). Filamentous plant pathogen effectors in action. *Nat. Rev. Microbiol.* 11, 800–814. doi: 10.1038/nrmicro3119
- Godfrey, D., Böhlenius, 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
- Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. doi: 10.1093/bioinformatics/btr064
- Haegeman, A., Jones, J. T., and Danchin, E. G. (2011). Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Mol. Plant Microbe Interact.* 24, 879–887. doi: 10.1094/MPMI-03-11-0055
- Hemetsberger, C., Mueller, A. N., Matei, A., Herrberger, C., Hensel, G., Kumlehn, J., et al. (2015). The fungal core effector Pep1 is conserved across smuts of dicots and monocots. *New Phytol.* 206, 1116–1126. doi: 10.1111/nph.13304
- Hiller, K., Grote, A., Scheer, M., Münch, R., and Jahn, D. (2004). PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res.* 32, W375–W379. doi: 10.1093/nar/gkh378
- Hofman, K. (1993). TMbase-A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* 374:166.
- Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259
- Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R. P., Fujie, M., et al. (2013). Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* 153, 1567–1578. doi: 10.1016/j.cell.2013.05.040
- Jiang, R. H., Tripathy, S., Govers, F., and Tyler, B. M. (2008). RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4874–4879. doi: 10.1073/pnas.0709303105
- Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286
- Kadam, S., Vuong, T. D., Qiu, D., Meinhardt, C. G., Song, L., Deshmukh, R., et al. (2015). Genomic-assisted phylogenetic analysis and marker development for next generation soybean cyst nematode resistance breeding. *Plant Sci.* 242, 342–350. doi: 10.1016/j.plantsci.2015.08.015
- Kale, S. D., Gu, B., Capelluto, D. G., Dou, D., Feldman, E., Rumore, A., et al. (2010). External lipid PI3P mediates entry of eukaryotic pathogen effectors into plant and animal host cells. *Cell* 142, 284–295. doi: 10.1016/j.cell.2010.06.008
- Kale, S. D., and Tyler, B. M. (2011). Entry of oomycete and fungal effectors into plant and animal host cells. *Cell Microbiol.* 13, 1839–1848. doi: 10.1111/j.1462-5822.2011.01659.x
- Käll, L., Krogh, A., and Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* 35, W429–W432. doi: 10.1093/nar/gkm256
- Kamoun, S. (2006). A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu. Rev. Phytopathol.* 44, 41–60. doi: 10.1146/annurev.phyto.44.070505.143436
- Kämper, J., Kahmann, R., Bölker, 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
- Kanneganti, T.-D., Huitema, E., Cakir, C., and Kamoun, S. (2006). Synergistic interactions of the plant cell death pathways induced by *Phytophthora infestans* Nep1-like protein PiNPP1.1 and INF1 elicitor. *Mol. Plant Microbe Interact.* 19, 854–863. doi: 10.1094/MPMI-19-0854
- Karányi, Z., Holb, I., Hornok, L., Pocs, I., and Miskei, M. (2013). FSRD: fungal stress response database. *Database* 2013:bat037. doi: 10.1093/database/bat037
- Klammer, M., Messina, D. N., Schmitt, T., and Sonnhammer, E. L. (2009). MetaTM—a consensus method for transmembrane protein topology prediction. *BMC Bioinformatics* 10:314. doi: 10.1186/1471-2105-10-314
- Kretschmer, M. (2012). “Emergence of multi-drug resistance in fungal pathogens: a potential threat to fungicide performance in agriculture,” in *Fungicide Resistance in Crop Protection: Risk and Management*, ed. T. S. Thind (Oxfordshire: CABI), 251–267.
- Kutcher, H. R., Brandt, S. A., Smith, E. G., Ulrich, D. J., Malhi, S. S., and Johnston, A. M. (2013). Blackleg disease of canola mitigated by resistant cultivars and four-year crop rotations in western Canada. *Can. J. Plant Pathol.* 35, 209–221. doi: 10.1080/07060661.2013.775600
- Lefebvre, F., Joly, D. L., Labbé, C., Teichmann, B., Linning, R., Belzile, F., et al. (2013). The transition from a phytopathogenic smut ancestor to an anamorphic biocontrol agent deciphered by comparative whole-genome analysis. *Plant Cell* 25, 1946–1959. doi: 10.1105/tpc.113.113969
- Leonelli, L., Pelton, J., Schoeffler, A., Dahlbeck, D., Berger, J., Wemmer, D. E., et al. (2011). Structural elucidation and functional characterization of the *Hyaloperonospora arabidopsidis* effector protein ATR13. *PLoS Pathog.* 7:e1002428. doi: 10.1371/journal.ppat.1002428
- Liu, Z., Zhang, Z., Farris, J. D., Oliver, R. P., Syme, R., McDonald, M. C., et al. (2012). The cysteine rich necrotrophic effector SnTox1 produced by *Stagonospora nodorum* triggers susceptibility of wheat lines harboring Snn1. *PLoS Pathog.* 8:e1002467. doi: 10.1371/journal.ppat.1002467
- 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
- Lower, M., and Schneider, G. (2009). Prediction of type III secretion signals in genomes of gram-negative bacteria. *PLoS ONE* 4:e5917. doi: 10.1371/journal.pone.0005917
- Lu, T., Yao, B., and Zhang, C. (2012). DFVF: database of fungal virulence factors. *Database* 2012:bas032. doi: 10.1093/database/bas032
- Lum, G., and Min, X. J. (2011). FunSecKB: the fungal secretome knowledgebase. *Database* 2011:bar001. doi: 10.1093/database/bar001
- Maqbool, A., Saitoh, H., Franceschetti, M., Stevenson, C. E. M., Uemura, A., Kanzaki, H., et al. (2015). Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife* 4:e08709. doi: 10.7554/eLife.08709
- Meinken, J., Asch, D. K., Neizer-Ashun, K. A., Chang, G.-H., Cooper, C. R. Jr., and Min, X. J. (2014). FunSecKB2: a fungal protein subcellular location knowledgebase. *Comput. Mol. Biol.* 4, 1–17. doi: 10.5376/cmb.2014.04.0007
- Memi, V., Kumar, K., Cheng, L., Zavaljevski, N., Deshazer, D., Wallqvist, A., et al. (2014). DBSecSys: a database of *Burkholderia mallei* secretion systems. *BMC Bioinformatics* 15:244. doi: 10.1186/1471-2105-15-244
- Misas-Villamil, J. C., and van der Hoorn, R. A. (2008). Enzyme-inhibitor interactions at the plant-pathogen interface. *Curr. Opin. Plant Biol.* 11, 380–388. doi: 10.1016/j.pbi.2008.04.007
- Mitchum, M. G., Hussey, R. S., Baum, T. J., Wang, X., Elling, A. A., Wubben, M., et al. (2013). Nematode effector proteins: an emerging paradigm of parasitism. *New Phytol.* 199, 879–894. doi: 10.1111/nph.12323
- Morgan, W., and Kamoun, S. (2007). RXLR effectors of plant pathogenic oomycetes. *Curr. Opin. Microbiol.* 10, 332–338. doi: 10.1016/j.mib.2007.04.005
- Nemri, A., Saunders, D. G., Anderson, C., Upadhyaya, N. M., 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
- Pasquier, C., Promponas, V., Palaios, G., Hamodrakas, J., and Hamodrakas, S. (1999). A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: the PRED-TMR algorithm. *Protein Eng.* 12, 381–385. doi: 10.1093/protein/12.5.381
- Pedersen, C., van Themaat, E. V. L., McGuffin, L. J., Abbott, J. C., Burgis, T. A., Barton, G., et al. (2012). Structure and evolution of barley powdery mildew effector candidates. *BMC Genomics* 13:694. doi: 10.1186/1471-2164-13-694
- 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

- Raman, R., Taylor, B., Marcroft, S., Stiller, J., Eckermann, P., Coombes, N., et al. (2012). Molecular mapping of qualitative and quantitative loci for resistance to *Leptosphaeria maculans* causing blackleg disease in canola (*Brassica napus* L.). *Theor. Appl. Genet.* 125, 405–418. doi: 10.1007/s00122-012-1842-6
- Reynolds, S. M., Käll, L., Riffle, M. E., Bilmes, J. A., and Noble, W. S. (2008). Transmembrane topology and signal peptide prediction using dynamic bayesian networks. *PLoS Computat. Biol.* 4:e1000213. doi: 10.1371/journal.pcbi.1000213
- Rost, B., Yachdav, G., and Liu, J. (2004). The predictprotein server. *Nucleic Acids Res.* 32, W321–W326. doi: 10.1093/nar/gkh377
- Saha, P., Kalia, P., Sonah, H., and Sharma, T. R. (2014). Molecular mapping of black rot resistance locus Xca1bo on chromosome 3 in Indian cauliflower (*Brassica oleracea* var. botrytis L.). *Plant Breed.* 133, 268–274. doi: 10.1111/pbr.12152
- Samudrala, R., Heffron, F., and Mcdermott, J. E. (2009). Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for type III secretion systems. *PLoS Pathog.* 5:e1000375. doi: 10.1371/journal.ppat.1000375
- 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
- Savary, S., Ficke, A., Aubertot, J.-N., and Hollier, C. (2012). Crop losses due to diseases and their implications for global food production losses and food security. *Food Sec.* 4, 519–537. doi: 10.1007/s12571-012-0200-5
- Schuster, M., Schweizer, G., Reissmann, S., and Kahmann, R. (2015). Genome editing in *Ustilago maydis* using the CRISPR-Cas system. *Fungal Genet. Biol.* doi: 10.1016/j.fgb.2015.09.001 [Epub ahead of print].
- Sen, A. (2011). MycoSec-A database for signal peptide bearing genes of mycobacterium. *J. Nat. Sci. Biol. Med.* 2, 86–87.
- Shen, H.-B., and Chou, K.-C. (2007). Signal-3L: a 3-layer approach for predicting signal peptides. *Biochem. Biophys. Res. Commun.* 363, 297–303. doi: 10.1016/j.bbrc.2007.08.140
- Silva-Gomesa, S., Decouta, A., and Nigoua, J. (2014). “Pathogen-associated molecular patterns (PAMPs),” in *Encyclopedia of Inflammatory Diseases*, ed. M. Parnham (Basel: Springer), 1–16. doi: 10.1007/978-3-0348-0620-6_35-1
- Singh, S., Sharma, S., Kalia, P., Deshmukh, R., Kumar, V., Sharma, P., et al. (2012). Molecular mapping of the downy mildew resistance gene Ppa3 in cauliflower (*Brassica oleracea* var. botrytis L.). *J. Hortic. Sci. Biotechnol.* 87, 137–143.
- Sonah, H., Deshmukh, R., and Kotasthane, A. (2012). Fungicidal interference during infection related developmental stages in *Magnaporthe grisea*. *eSci. J. Plant Pathol.* 1, 49–55.
- Souza, R. C., Saji, G. D. R. Q., Costa, M. O., Netto, D. S., Lima, N. C., Klein, C. C., et al. (2012). AtlasT4SS: a curated database for type IV secretion systems. *BMC Microbiol.* 12:172. doi: 10.1186/1471-2180-12-172
- Sperschneider, J., Gardiner, D. M., Taylor, J. M., Hane, J. K., Singh, K. B., and Manners, J. M. (2013). A comparative hidden Markov model analysis pipeline identifies proteins characteristic of cereal-infecting fungi. *BMC Genomics* 14:807. doi: 10.1186/1471-2164-14-807
- Stergiopoulos, I., and de Wit, P. J. (2009). Fungal effector proteins. *Annu. Rev. Phytopathol.* 47, 233–263. doi: 10.1146/annurev.phyto.112408.132637
- Stuart, J. (2015). Insect effectors and gene-for-gene interactions with host plants. *Curr. Opin. Insect Sci.* 9, 56–61. doi: 10.1016/j.cois.2015.02.010
- Sun, F., Kale, S. D., Azurmendi, H. F., Li, D., Tyler, B. M., and Capelluto, D. G. (2013). Structural basis for interactions of the *Phytophthora sojae* RxLR effector Avh5 with phosphatidylinositol 3-phosphate and for host cell entry. *Mol. Plant Microbe Interact.* 26, 330–344. doi: 10.1094/MPMI-07-12-0184-R
- Tanaka, S., Djamei, A., Presti, L. L., Schipper, K., Winterberg, S., Amati, S., et al. (2015). Experimental approaches to investigate effector translocation into host cells in the *Ustilago maydis*/maize pathosystem. *Eur. J. Cell Biol.* 94, 349–358. doi: 10.1016/j.ejcb.2015.06.007
- Tay, D. M., Govindarajan, K. R., Khan, A. M., Ong, T. Y., Samad, H. M., Soh, W. W., et al. (2010). T3SEdb: data warehousing of virulence effectors secreted by the bacterial Type III secretion system. *BMC Bioinformatics* 11:S4. doi: 10.1186/1471-2105-11-S7-S4
- Teper, D., Burstein, D., Salomon, D., Gershovitz, M., Pupko, T., and Sessa, G. (2015). Identification of novel *Xanthomonas euvesicatoria* type III effector proteins by a machine-learning approach. *Mol. Plant Pathol.* doi: 10.1111/mpp.12288 [Epub ahead of print].
- Thomma, B. P., Van Esse, H. P., Crous, P. W., and de Wit, P. J. (2005). *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic *Mycosphaerellaceae*. *Mol. Plant Pathol.* 6, 379–393. doi: 10.1111/j.1364-3703.2005.00292.x
- Torriani, S. F., Stukenbrock, E. H., Brunner, P. C., McDonald, B. A., and Croll, D. (2011). Evidence for extensive recent intron transposition in closely related fungi. *Curr. Biol.* 21, 2017–2022. doi: 10.1016/j.cub.2011.10.041
- Tusnady, G. E., and Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17, 849–850. doi: 10.1093/bioinformatics/17.9.849
- Urban, M., Pant, R., Raghunath, A., Irvine, A. G., Pedro, H., and Hammond-Kosack, K. E. (2014). The pathogen-host interactions database (PHI-base): additions and future developments. *Nucleic Acids Res.* 43, D645–D655. doi: 10.1093/nar/gku1165
- Van de Wouw, A. P., Marcroft, S. J., Ware, A., Lindbeck, K., Khangura, R., and Howlett, B. J. (2014). Breakdown of resistance to the fungal disease, blackleg, is averted in commercial canola (*Brassica napus*) crops in Australia. *Field Crops Res.* 166, 144–151. doi: 10.1016/j.fcr.2014.06.023
- Viklund, H., Bernsel, A., Skwark, M., and Elofsson, A. (2008). SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics* 24, 2928–2929. doi: 10.1093/bioinformatics/btn550
- Vuong, T. D., Sonah, H., Meinhard, C. G., Deshmukh, R., Kadam, S., Nelson, R. L., et al. (2015). Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics* 2015:593. doi: 10.1186/s12864-015-1811-y
- Wang, Y., Zhang, Q., Sun, M.-A., and Guo, D. (2011). High-accuracy prediction of bacterial type III secreted effectors based on position-specific amino acid composition profiles. *Bioinformatics* 27, 777–784. doi: 10.1093/bioinformatics/btr021
- Whisson, S. C., Boevink, P. C., Moleleki, L., Avrova, A. O., Morales, J. G., Gilroy, E. M., et al. (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* 450, 115–118. doi: 10.1038/nature06203
- 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. doi: 10.1371/journal.ppat.1002400
- Winnenburg, R., Urban, M., Beacham, A., Baldwin, T. K., Holland, S., Lindeberg, M., et al. (2008). PHI-base update: additions to the pathogen-host interaction database. *Nucleic Acids Res.* 36, D572–D576. doi: 10.1093/nar/gkm858
- Wirthmueller, L., Maqbool, A., and Banfield, M. J. (2013). On the front line: structural insights into plant-pathogen interactions. *Nat. Rev. Microbiol.* 11, 761–776. doi: 10.1038/nrmicro3118
- Woo, M. O., Beard, H., MacDonald, M. H., Brewer, E. P., Youssef, R. M., Kim, H., et al. (2014). Manipulation of two α -endo- β -1, 4-glucanase genes, AtCel6 and GmCel7, reduces susceptibility to *Heterodera glycines* in soybean roots. *Mol. Plant Pathol.* 15, 927–939. doi: 10.1111/mpp.12157
- Wu, J., and Baldwin, I. T. (2010). New insights into plant responses to the attack from insect herbivores. *Annu. Rev. Genet.* 44, 1–24. doi: 10.1146/annurev-genet-102209-163500
- Yaeno, T., Li, H., Chaparro-Garcia, A., Schornack, S., Koshiba, S., Watanabe, S., et al. (2011). Phosphatidylinositol monophosphate-binding interface in the oomycete RXLR effector AVR3a is required for its stability in host cells to modulate plant immunity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14682–14687. doi: 10.1073/pnas.1106002108
- Ye, W., Wang, Y., and Wang, Y. (2015). Bioinformatics analysis reveals abundant short alpha-helices as a common structural feature of oomycete RxLR effector proteins. *PLoS ONE* 10:e0135240. doi: 10.1371/journal.pone.0135240
- Yu, D., Shen, H., and Yang, J. (2011). SOMRuler: a novel interpretable transmembrane helices predictor. *IEEE Trans. NanoBiosci.* 10, 121–129. doi: 10.1109/TNB.2011.2160730

- Zhou, M., Theunissen, D., Wels, M., and Siezen, R. J. (2010). LAB-Secretome: a genome-scale comparative analysis of the predicted extracellular and surface-associated proteins of Lactic Acid Bacteria. *BMC Genomics* 11:651. doi: 10.1186/1471-2164-11-651
- Zuccaro, A., Lahrmann, U., Güldener, U., Langen, G., Pfiffi, S., Biedenkopf, D., et al. (2011). Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont *Piriformospora indica*. *PLoS Pathog.* 7:e1002290. doi: 10.1371/journal.ppat.1002290

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 © 2016 Sonah, Deshmukh and Bélanger. 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.