Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

# Bioinformatic prediction of plant–pathogenicity effector proteins of fungi

Darcy AB Jones[1], Stefania Bertazzoni[1], Chala J Turo[1], Robert A Syme[1] and James K Hane[1,2]

Effector proteins are important virulence factors of fungal plant pathogens and their prediction largely relies on bioinformatic methods. In this review we outline the current methods for the prediction of fungal plant pathogenicity effector proteins. Some fungal effectors have been characterised and are represented by conserved motifs or in sequence repositories, however most fungal effectors do not generally exhibit high conservation of amino acid sequence. Therefore various predictive methods have been developed around: general properties, structure, position in the genomic landscape, and detection of mutations including repeat-induced point mutations and positive selection. A combinatorial approach incorporating several of these methods is often employed and candidates can be prioritised by either ranked scores or hierarchical clustering.

**Addresses**
[1] Centre for Crop and Disease Management, Curtin University, Kent Street, Bentley, WA 6102, Australia
[2] Curtin Institute for Computation, Curtin University, Kent Street, Bentley, WA 6102, Australia

Corresponding author: Hane, James K (james.hane@curtin.edu.au)

**Current Opinion in Microbiology** 2018, **46**:43–49

This review comes from a themed issue on **Host–microbe interactions: fungi**

Edited by **Ralph Dean** and **Thierry Rouxel**

https://doi.org/10.1016/j.mib.2018.01.017



## Introduction

Fungal and oomycete plant pathogens are a major food security problem, with as few as five major species destroying stocks capable of feeding >600 million people [1]. Many plant pathogens possess a battery of 'effector' molecules, usually proteins, which initiate disease and circumvent host defences by either masking the presence of the pathogen or killing the host cell directly [2–4]. Effector identification is critical to developing crop resistance [5] and their prediction largely relies on bioinformatics [6]. Wide adoption and affordability of genome sequencing has enabled multiple pathogen genome studies predicting numerous effector candidates with limited capacity for experimental validation. This is indicative of an ongoing community need for improved knowledge around definitive effector properties and bioinformatic prediction methods to prioritise validation of candidate effectors. This review provides an overview of current and emerging methods for proteinaceous effector prediction in fungi.

The term 'effector' is used to describe multiple loosely conserved families of proteins that are cytotoxic or otherwise compromise cells of a host organism. In plant pathogenic fungal species, these may share basic properties including: low molecular weight, externalisation from the pathogen cell, and cysteine-richness [2–4], however there are several exceptions [7•,8]. Identification of conserved sequence motifs that correlate to pathogenicity-related functions has had mixed success. Two publicly available repositories of proteins with confirmed roles in pathogenicity exist (PHIbase [8] and DFVF [9]). PHIbase aggregates experimental reports validating virulence activities, predominantly of fungal–plant interactions. DFVF groups confirmed pathogenicity factors according to host range. General conserved domain databases also contain small but growing sets of plant–pathogenic functional domains (Table 1A).

A handful of conserved amino-acid motifs have been identified in plant pathogen effector proteins (Table 1B). These are primarily observed in oomycetes and tend to be commonly enriched in the secretomes of species from the same genus [4,10–13]. The crinkler motif is broader, and observed across many sequenced oomycetes [4]. For conserved effector families such as these (Table 1), it is possible to generate profile hidden Markov models (HMMs) to represent the class. Based on current examples from the oomycetes [4], it would appear that motif enrichment analysis within predicted secretomes may be sufficient to predict pathogenicity motifs in novel oomycete genomes.

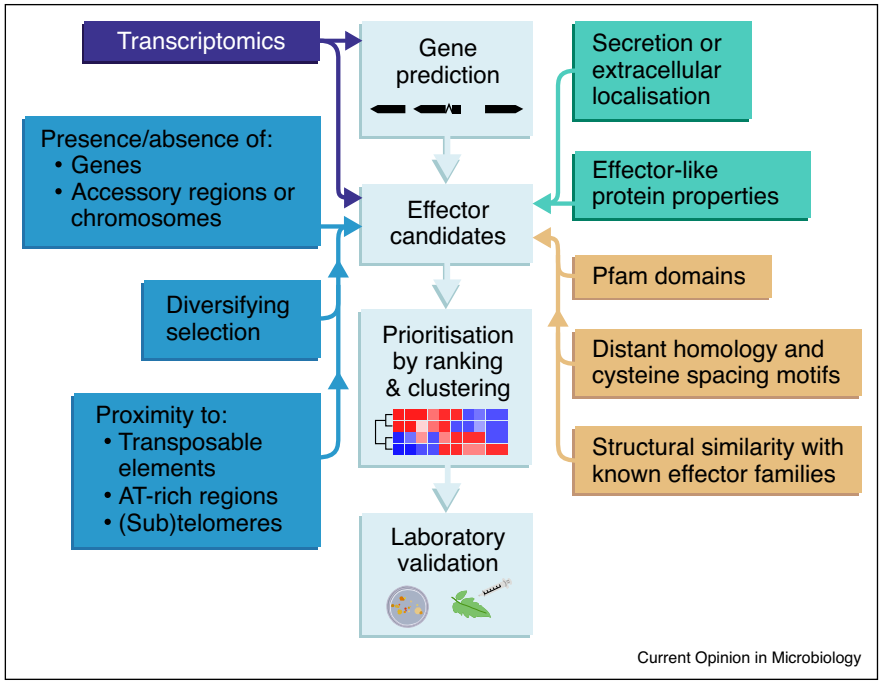
Prediction methods in fungi (Figure 1, Table 2) are more complex as their known effectors tend to lack sequence conservation, likely due to genome-wide mutagenesis processes that make most fungal genomes inherently plastic [14–16]. For example, while some fungi possess known 'RXLR-like' effectors with proposed similar membrane interaction functions [17] a simple pattern-based representation of known fungal RXLR-like motifs

**Table 1**

**Summary of conserved domains (A) and conserved amino-acid motifs (B) observed in plant–pathogen effector proteins.**

| **(A) Plant–pathogenic conserved domains** | Pfam ID |
|---|---|
| ToxA | Toxin_ToxA (PF11584) |
| Phytotoxin PcF protein | PcF (PF09461) |
| Putative necrosis-inducing factor | Hce2 (PF14856) |
| RXLR phytopathogen effector protein | RXLR (PF16810) |
| Avirulence protein ATR13, RxLR effector | ATR13 (PF16829) |
| Elicitin | Elicitin (PF00964) |
| **(B) Plant–pathogen effector conserved AA motifs** | Primary taxa |
| RxLR . . . dEER | Phytophthora spp. |
| Crinklers: LxLFLAK . . . (DWL)n...HVLVxxP | Oomycetes, for example, *Phytophthora* spp., *H. arabidopsidis*, *B. lactucae*, *Pythium* spp. |
| CxHC | *Albugo laibachii* |
| YxSL[RK] | *Phythium* spp. |
| [YFW]xC | *Blumeria graminis* |
| ETVIC and HRxxH | *Blumeria graminis* |

**Figure 1**




Suggested bioinformatic workflow for generating and prioritising fungal effector candidates.

yields a high false discovery rate in fungal genomes. Similarly, reports of fungal genes possessing the crinkler motif are rare and diverse in sequence [18,19]. Although fungal effectors tend to lack conservation, an exception are the conserved 'Secreted In Xylem' (SIX) genes of *Fusarium oxysporum formae speciales* (ff. spp.) [20], some of which are downstream of a highly conserved miniature impala transposable element that can be used as a predictive marker [21]. However, in most cases the overall lack of usable sequence conservation for effector prediction necessitates a composite approach using various other properties that have been observed for known effectors (Figure 1, Table 2).

Although effector sequences generally lack sequence similarity, common protein structural features have been recently observed within a handful of effector 'families' [3], including ToxA-like [22,23•], MAX [24•], AvrLm6-like [25], RXLR-like with WY-domains [13,26] and RALPH [27•]. Interestingly, both the ToxA-like and MAX families possess a β sandwich tertiary structure formed from 7 and 6 β sheets respectively. Structural homology searches based on position specific scoring matrices (PSSMs) or profile-HMMs (Table 1) have been used to predict new candidates based on matches to existing families [13,25] or for assumption-free prediction [28]. Outside of plant pathology other