ORIGINAL ARTICLE

Investigating the effector suite profile of Australian *Fusarium oxysporum* isolates from agricultural and natural ecosystems

Saidi R. Achari[1,2] (iD) | Jatinder K. Kaur[1] | Ross C. Mann[1] | Tim Sawbridge[1,2] |
Brett A. Summerell[3] | Jacqueline Edwards[1,2]

[1]AgriBio, Agriculture Victoria Research, DJPR, Bundoora, Victoria, Australia

[2]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia

[3]Australian Institute of Botanical Science, Royal Botanic Gardens and Domain Trust, Sydney, NSW, Australia

Correspondence
Saidi R. Achari, AgriBio, Agriculture Victoria Research, DJPR, Bundoora, Victoria, Australia.
Email: saidi.achari@agriculture.vic.gov.au

Abstract

Pathogenic and putatively nonpathogenic isolates of *Fusarium oxysporum* are ubiquitously present in soils. Pathogenic isolates designated as *formae speciales* are very host specific. The genes that determine host-specific pathogenicity may be expected to be similar between strains within a *forma specialis*. Three different pathways were used for mining putative effectors from 85 assembled genome sequences representing isolates from the agricultural and natural ecosystems. A total of 436 putative effectors were identified, of which 115 were present in all the isolates. The presence–absence phylogeny showed some clustering of isolates based on the ecosystems, while some clustered with isolates from different ecosystems. Isolates from the agricultural and natural ecosystems shared 358 putative effectors. Agricultural isolates had 78 unique putative effectors, while isolates from the natural ecosystems had no unique effectors. Isolates from the natural ecosystems had effectors that have previously been confirmed with virulence functions in other filamentous fungi, suggesting that the natural ecosystems may be acting as a reservoir for pathogenic isolates awaiting compatible hosts. They also possessed putative effectors with domains involved in polysaccharide degradation. The putative effectors in the natural ecosystem isolates may be important for both parasitic and ecological fitness. Further analysis of the three characterized ff. sp.—*lycopersici*, *niveum*, and *pisi*—revealed high conservation in the pathogenicity mechanism in *F. oxysporum*. However, each strain had unique putative effectors, which may be responsible for host specificity.

KEYWORDS
agricultural ecosystems, domains, effectors, natural ecosystems, pathogenicity

1 | INTRODUCTION

The *Fusarium oxysporum* species complex (FOSC) comprises both pathogenic, usually defined as *formae speciales* (ff. sp.), and putatively nonpathogenic isolates ubiquitously present in soils. It is not possible to differentiate these strains based on morphological characteristics (Summerell, 2019). Additionally, molecular characterization of most of these strains using conserved genes is not possible, as phylogenetic studies based on conserved genes have revealed considerable diversity between and within isolates of a *forma specialis* (f. sp.),

spreading them across several clades (Achari et al., 2020). This is due to the polyphyletic nature of most *F. oxysporum* (Fo) ff. sp., brought about by their ability to undergo horizontal gene transfer (Fokkens et al., 2018; Ma et al., 2010).

The Fo genome is divided into core and lineage, accessory or adaptive genomes (Ma et al., 2010). Fo f. sp. *lycopersici* (Fol) strain Fol4287 has been studied intensively, and its genome assembled into chromosomes. It has a genome size of 53.9 Mb, with a core genome of 41.8 Mb (Ma et al., 2010). The core genome size was reduced further with the identification of chromosomes 11, 12, and 13 as being more divergent than other core chromosomes and harbouring genes that were differentially expressed during infection (Fokkens et al., 2018). The transfer of these three chromosomes to the adaptive genome increased the adaptive genome to 18.1 Mb and decreased the core genome to 35.7 Mb, resulting in the adaptive genome constituting 33.7% of the total Fol4287 genome. The genes located in the core genome are responsible for housekeeping, such as maintenance of essential cellular functions (Ma et al., 2010). Homologues of these genes with near-identical sequence identity are present in all Fo isolates, rendering them unsuitable as molecular markers (van Dam et al., 2018; Ma et al., 2010). The adaptive genome of Fo, which is between 4 and 19 Mb, is very diverse in size and content (Fokkens et al., 2018; Ma et al., 2010). Although it is a gene-poor region of the genome, the genes responsible for determining host specificity and pathogenicity are located here (Fokkens et al., 2018; Ma et al., 2010). In Fol4287, chromosomes 3, 6, 14, 15, scaffolds 27 and 31 in chromosomes 1 and 2, respectively (Ma et al., 2010), and chromosomes 11, 12, and 13 (Fokkens et al., 2018) form the adaptive genomic region. Chromosome 14 is known as the pathogenicity chromosome, as it houses all the secreted-in-xylem (*SIX*) genes except for *SIX8* (Ma et al., 2010). The transfer of this chromosome from a pathogenic isolate to a nonpathogenic isolate converted the latter to be pathogenic on tomato, but less virulent compared to the donor strain (Ma et al., 2010).

Pathogens secrete effector proteins to counteract the pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) defence responses of the host (Jones & Dangl, 2006). The impact of these effector molecules on the function and structure of the plant's defence network can either facilitate infection or trigger defence, in the case of avirulence genes, if these effectors are recognized by their complementary resistance genes (Jones & Dangl, 2006). Effectors are also used by pathogens to modify the host plant physiology to accommodate their growth and to provide them with nutrients (Lo Presti et al., 2015). While they play a vital role in pathogen virulence, only a few have been characterized to determine the virulence functionality.

Effectors are often responsible for host or cultivar specificity and, as such, form the genetic functional elements within the pathogen–host interaction (van Dam et al., 2016; Ma et al., 2010). Effector gene sequences are very diverse between strains of different ff. sp. but are conserved between different clonal lineages within a f. sp., making them potential markers for host-specific pathogenicity (van Dam et al., 2018). A recent study to differentiate cucurbit-infecting

ff. sp.—*cucumerinum*, *melonis*, *niveum*, *radicis-cucumerinum*, *luffae*, and *lagenariae*—revealed that these ff. sp. share some effector genes, while some are host-specific (van Dam et al., 2018). These host-specific effectors have been used as molecular markers in a PCR to successfully differentiate these isolates (van Dam et al., 2018).

As a group, the secreted effector proteins span a diverse range of functions, with many described as enzymes (van Dam et al., 2016) and avirulence proteins in the form of toxins, elicitors, or virulence factors (Ma et al., 2015). Many of the Fo effectors have been grouped into a functionally diverse set of genes derived from a set of proteins observed to be secreted in xylem (SIX1 to SIX14) (Ma et al., 2015; Schmidt et al., 2013). Most of the SIX protein genes are located downstream of a highly conserved miniature impala (mimps) transposable element (Schmidt et al., 2013). These proteins have conserved sequences but lack conserved domains.

Different forms of effector proteins are secreted depending on the lifestyle of the fungus. Effectors secreted by biotrophic fungi are typically used to impede the plant recognition receptors (PRR) and suppress the plant-triggered immunity and improve susceptibility (Lo Presti et al., 2015). In contrast, the effectors secreted by necrotrophic fungi include cell wall-degrading enzymes (CWDEs) and toxins to kill the host (Lo Presti et al., 2015).

Putative effectors are generally mined from a genome by filtering the predicted proteins based on generalized effector characteristics: predicted secretion, extracellular localization, small size molecular weight (<300 amino acids or < 30 kDa), enriched in cysteine residues (four cysteine residues; Lo Presti et al., 2015), lack of sequence similarity to any known functional protein, located near or within regions of repetitive DNA (Schmidt et al., 2013), as well as being located in the adaptive genome or lineage-specific region of the genome (Ma et al., 2010). However, using these generalized thresholds risks excluding a small number of genuine effectors, as large molecular weight and cysteine-poor proteins have been observed from in planta transcriptomics (Sperschneider et al., 2016).

Due to the evolutionary relationship between structure and function, the protein structure is more conserved than the amino acid sequence (Illergard et al., 2009). The structural conservation of the protein is a powerful tool for characterizing and understanding the effector biology of proteins that lack sequence similarity to known functional proteins (Illergard et al., 2009). Although the effector proteins generally lack significant sequence similarity to other known proteins, common protein structural features have been determined in some effector families. Some of the commonly identified structural features (domains) shared by some effectors are the presence of Hce2, ToxA, Phytotoxin PcF protein, putative necrosis-inducing factor (Jones et al., 2018), and LysM domain (Illergard et al., 2009). Effector prediction necessitates a composite approach to address all these characteristics.

The FOSC displays considerable ecological diversity, occurring as endophytes, saprophytes, as well as pathogens in agricultural ecosystems. Substantial populations can be recovered from soils in natural uncultivated ecosystems (Laurence et al., 2012). Many studies on effectors have been on pathogenic isolates from agricultural

ecosystems. The main objective of this study was to predict the putative effectors from the genomes of Fo isolates from the agricultural and natural ecosystems and create an effector profile suite for these ecosystems. A second objective was to identify any conserved protein domains in the Fo putative effectors.

2 | MATERIALS AND METHODS

Assembled genomes of 83 Fo isolates from Australian natural and agricultural ecosystems were used in this study, together with Fo47 and Fol4287. Thirteen and 71 isolates were from natural and agricultural ecosystems, respectively. Isolates from the agricultural ecosystems were isolated from plants with and without symptoms. The Australian isolates were assembled as part of an earlier study by Achari et al. (2020). All these assemblies are available from the National Center for Biotechnology Information (NCBI; Table S1).

2.1 | Genome compartmentalization

The Fo genome is divided into two compartments: the core and the adaptive genomes. The Fol4287 genome has been assembled to the chromosome level. The sequences of the genes present on chromosomes 4, 5, 7, 8, 9, 10, and non-subtelomeric regions of chromosomes 1 and 2 were concatenated to create a core genome of Fol4287. Using minimap (Li, 2016), all the assembled genomes were mapped to the core genome of Fol4287, and the mapped contigs were designated as the core genome of the individual isolates, while the unmapped reads were designated as the adaptive genome.

2.2 | Identification of putative effectors

Putative effector proteins need to have a secretion signal and be secreted extracellularly. These proteins are small and enriched with cysteine residues. They lack significant sequence similarity to other known proteins and are located close to mimps (Schmidt et al., 2013) on chromosomes present in the adaptive genome (Ma et al., 2010). The domains of some putative effector proteins have conserved structures and functions. Three pathways were created to identify proteins having these effector-like characteristics.

2.2.1 | Pathway 1: prediction of effectors based on protein properties

The genomes were annotated using AUGUSTUS v. 3.2.3 (augustus --species = fusarium) (Stanke et al., 2006) and Maker2 v. 2.31.9 with default parameters (Holt and Yandell, 2011). The annotated protein sequences were submitted through a web-based program SECRETOOL (http://genomics.cicbiogune.es/SECRETOOL/Secretool.php) (Cortázar et al., 2014) [accessed on 11/2019] with default parameters. This software has SignalP, TargetP, PredGPI, WoLFPSORT, and TMHMM to identify proteins with signal peptides with no transmembrane domains. The secretomes or the secretory proteins obtained from SECRETOOL were then submitted to Effector2P (http://effectorp.csiro.au/) (Sperschneider et al., 2018) [accessed on 11/2019] for prediction of effector-like proteins from these secretory proteins. Protein sequences of fewer than 30 amino acids were discarded. The sequences for putative effectors obtained from AUGUSTUS v. 3.2.3 and Maker2 v. 2.31.9 were combined for each isolate. These combined effector sequences were then checked for duplicated sequences (100% identity) and removed by submitting the effector sequence FASTA file for each of the genomes to the CD-HIT suite (http://weizhong-lab.ucsd.edu/cdhit-web-server/cgi-bin/index.cgi?cmd=cd-hit) (Li and Godzik, 2006) [accessed on 11/2019] with default settings except that the 'Sequence identity cut-off' was set to 1.0. These protein sequences were then screened for subcellular localization using DeepLoc-1.0 (http://www.cbs.dtu.dk/services/DeepLoc-1.0/index.php (Almagro Armenteros et al., 2017) [accessed on 11/2019]. Protein sequences with only the extracellular localization were selected and were then used in searches against the NCBI protein sequence database using the DIAMOND BLASTP (Benjamin et al., 2015) for the identification of these putative effectors. Nucleotide sequences of the putative effectors were extracted from the annotation and added to the effector database.

2.2.2 | Pathway 2: prediction of mimp-associated effectors

This pathway was devised by van Dam et al. (2016) and is based on the relationship of the putative effector to the mimps. Their pipeline was used with some modifications. In Fo, mimps are always found within 1,500 bp upstream of the translation start site of *SIX* genes and several other genes expressed during plant infection (Schmidt et al., 2013). This pathway identifies an open reading frame (ORF) 2,500 bp downstream of a mimp terminal inverted repeat (TIR) while AUGUSTUS v. 3.2.3 predicts genes 5,000 bp downstream of the mimp TIR. Proteins identified in both are then passed through SignalP v. 4.0 (Petersen et al., 2011) to select only secreted proteins. These secreted proteins are combined and self-blasted to identify only the unique ones. Duplicated sequences (100% identity) were identified using the CD-HIT suite with default settings except that the 'Sequence identity cut-off' was set to 1.0 and removed. Protein sequences having $30 < x < 800$ amino acids were retained. These protein sequences were then screened for subcellular localization using DeepLoc-1.0.

2.2.3 | Pathway 3: effectors predicted in previous studies

In an earlier study, van Dam et al. (2016) predicted 104 candidate effectors from 59 genomes of Fo. These genomes included isolates of ff. sp. *conglutinans*, *cubense*, *cucumerinum*, *lycopersici*, *melonis*,

niveum, pisi, radicis-cucumerinum, radicis-lycopersici, raphani, vasinfectum, Fo47, and a clinical strain. All the effectors identified in their study were added to the effector database. Chang et al. (2019) had identified 20 putative effectors in close vicinity to the mimps in two strains of Fo f. sp. *cubense*: races 1 and 4. These were also added to the effector database.

2.2.4 | Finalizing the putative effector database

When all the putative effectors identified from the three pathways were added to the effector database, identities and sequences of the effectors were searched for duplicates using the CD-HIT suite to ensure uniqueness. The core genome of Fol4287 was created as described previously. A BLASTN search of the sequences in the effector database was performed against the Fol4287 core genome to identify effectors that were present in the core genome. Because genes and gene sequences present in the core genome are conserved in all Fo isolates, the effectors that mapped to the core genome of Fol4287 would be present in the core genome of the respective Fo isolates. These effectors were removed. For putative effectors found to be present on chromosomes 1 and 2, only those identified via Pathway 2 (close to mimps) are believed to be in the subtelomeres. Hence, they are in the adaptive region and were retained.

Putative effectors per genome were identified by conducting a BLASTN search (e-value 10^{-3}) on each assembled genome. The presence of a candidate effector gene in a genome was defined as having at least one BLAST hit with an e-value 10^{-3} and an identity score (number of identical nucleotides in the correct position in the alignment divided by the query length) of at least 30% as used by van Dam et al. (2016). The putative effectors in the effector database (those added through Pathway 3) that did not map to any of the genomes in this study were removed.

2.3 | Homologues of putative effector proteins in other fungi

Putative effector proteins do not have significant homology to known proteins. The protein sequences of the putative effectors were used to find homologues in other Fo and other fungi using BLASTP in NCBI, and the presence was confirmed if there was more than 40% amino acid identity with more than 50% query coverage, as used by Schmidt et al. (2016).

2.3.1 | Putative effectors found differentially expressed in previous studies

One of the characteristics of effector proteins is that they are differentially expressed during infection. Because this study was based on in silico mining of putative effectors from the assembled genomes, previous Fo gene expression studies were used to identify effectors found to be differentially expressed during infection. Chang et al. (2019) used Fo f. sp. *cubense* races 1 and 4 for their study, while 13 strains of Fo were used by van Dam et al. (2016). Taylor et al. (2016) used real-time reverse transcription (RT) PCR to identify pathogenesis-related genes expressed by Fo f. sp. *cepae* in onion seedlings. Genes reported to be differentially expressed in these studies were mapped to the putative effectors in the current study to confirm the role of the latter in infection.

2.4 | Conserved domain analysis

Although putative effector proteins do not share any significant sequence similarities, some possess similar structures (Illergard et al., 2009; Jones et al., 2018). Protein domains are evolutionarily conserved units of proteins, widely used to infer molecular and cellular protein functions (Illergard et al., 2009). Conserved domains of the proteins were determined using the InterProScan 5 (Jones et al., 2014) with E-value $< 10^{-10}$.

2.5 | Confirmation of a pathogenicity role of the putative effectors on pathogen–host interaction (PHI-base) database

To find homologues of the putative effectors that have been functionally characterized in other organisms, a BLAST search was performed of the 436 putative effector protein sequences on the PHI-base database (www.phi-base.org; accessed on 11/2019; Urban et al., 2019). The parameter used was E-value $< 10^{-5}$.

2.6 | Putative effector suite of isolates from the agricultural and natural ecosystems

The isolates were grouped based on the ecosystem from which they were obtained. Similarly, the putative effectors from these isolates were grouped to form putative effectors from the agricultural and natural ecosystems. A Venn diagram for the ecosystems was created to identify the putative effectors unique to each ecosystem and shared between isolates from these ecosystems. A further analysis was carried out for the pathogenic Fo isolates. In the datasets, we had three characterized ff. sp: *lycopersici* (Fol), *niveum* (Fon), and *pisi* (Fop). We created a Venn diagram for the putative effectors from these three ff. sp. to identify the effectors that are shared by all the ff. sp. and those that are unique.

2.6.1 | Presence and absence of putative effectors in the isolates

A binary data matrix was created containing the presence ("1") or absence ("0") of each candidate effector in each genome. This data

matrix was then analysed using Genedata Expressionist Analyst v. 10.0 (Genedata), and results viewed as a presence–absence phylogeny. The presence–absence phylogeny was created using the hierarchical clustering analysis using the default parameters. The species for the isolates were determined in an earlier study by Achari et al. (2020), and this information was added in the phylogeny tree to determine the relationship between the putative effectors and the putative species.

3 | RESULTS

3.1 | Genome compartmentalization

Eighty-three assembled genomes of isolates from natural and agricultural ecosystems together with Fol4287 and a biological control strain, Fo47, were used in this study. Thirteen and 71 isolates were from the Australian natural and agricultural ecosystems, respectively. The whole genome of the isolates tested ranged from 45 to 63 Mb. The core genomes ranged from 35 to 38 Mb, while the adaptive genomes ranged from 9.9 to 26.7 Mb, constituting 20.9%–42.4% of the whole genomes of the isolates (Table S1).

3.2 | Effectors identified through the three pathways

Three different pathways based on the characteristics of effector-like proteins were used to predict such proteins from the genomes. The reason for using three different pathways was to ensure that we could in silico predict most, if not all, of the effector-like proteins from the genomes for further analysis.

3.2.1 | Pathway 1: prediction of effectors based on protein properties

The number of proteins predicted by AUGUSTUS v. 3.2.3 and Maker2 v. 2.31.9 ranged from 14,294 to 25,037 per isolate. The number of secretory proteins, as predicted by SECRETOOL, ranged from 560 to 967 per isolate. Secretory proteins having the effector-like characteristics ranged from 122 to 173 per isolate, as predicted by Effector2P. Three hundred and twenty-eight putative effector protein sequences were discovered through this pathway (Table S2).

3.2.2 | Pathway 2: prediction of mimp-associated effectors

There were 316 putative effectors identified via this pathway. Seventy-five of these were already identified through Pathway 1. Of the remaining 241 putative effector proteins encoding for these effectors, only 128 were found to have an extracellular secretion, and these were added to the effector database (Table S2).

3.2.3 | Pathway 3: effectors predicted in previous studies

Two previous studies have identified some putative effectors from the genomes of different ff. sp. of Fo using the proximity of effectors to transposons. One hundred and four putative effectors had been identified by van Dam et al. (2016) from 59 Fo genomes. Fifteen of these were identified by Pathways 1 and 2. The remaining 89 effectors were added to the effector database (Table S2). In another study, Chang et al. (2019) identified 20 putative effectors from Fo f. sp. *cubense* races 1 and 4. Fifteen of these effectors were already in the database, identified from Pathways 1 and 2. The remaining five were added to the effector database (Table S2).

3.2.4 | Finalizing the putative effector database

Out of the 550 putative effectors identified via the three pathways, 83 mapped onto the core chromosomes of Fol4287 (Table S2) and were removed, retaining 467. A further 31 effectors added through Pathway 3 were removed from the effector database as they did not map to any genome sequences of the 85 isolates, leaving 436 effectors in the finalized effector database (Table S2).

3.3 | Overview of the putative effectors and their homologues in the NCBI database

All of the putative effectors had a signal peptide at the N-terminus except for two of the putative effectors added through Pathway 3 (Table S2). They all had an extracellular localization, hence would be involved in host interactions. The size of the putative effectors ranged from 58 to 788 amino acids in length, with only 21 of them being longer than 300 amino acids. None of the 436 putative effectors mapped to the core genome of the isolates, and hence are located in the adaptive genomes. One hundred and sixty of the putative effectors did not map to any of the Fol4287 chromosomes in the adaptive genome, hence are specific to the current data set (Table S2). Two hundred and seventy-six putative effectors mapped to chromosomes in the adaptive region of the Fol4287 genome (Table S2).

High levels of expression during infection is another characteristic of effector-like proteins. Previous studies have identified genes that are differentially expressed in different Fo isolates during infection. Ninety of the putative effector genes from the effector database mapped to the differentially expressed genes from previous studies (Table S3). Seventy-two and 29 of the putative effectors in the database mapped to genes found to be up-regulated in the differential gene expression studies by van Dam et al. (2016)

and Chang et al. (2019), respectively. Additionally, nine of the putative effectors mapped to pathogenesis-related genes (*CRX1*, *CRX2*, and *SIX*) that were expressed by Fo f. sp. *cepae* in onion seedlings (Taylor et al., 2016). Only one gene, *SIX9*, was common in all three gene expression studies (Chang et al., 2019; van Dam et al., 2016; Taylor et al., 2016).

Protein sequence homology was used to understand the evolutionary history of these putative effectors (Figure S1). Sixty-six percent of the effector candidates had homologues in Fo and other *Fusarium* species. Ten percent had homologues in only Fo, and 3% had no significant match to anything on the NCBI database. Twenty-one percent of the effectors had homologues in Fo and other fungi. Although 14 *SIX* genes are thought to be specific to Fo, homologues have been identified for some of these *SIX* genes in other fungal genera, namely *SIX2*, *SIX3*, *SIX5*, *SIX6*, *SIX11*, *SIX13*, and *SIX14* (Table S2).

3.4 | Conserved domain analysis of the putative effectors

Pathogen effector proteins evolve rapidly by diversifying the amino acid sequences to avoid detection by the host defence mechanism; however, the protein structure is conserved (Illergard et al., 2009). Only 113 of the 436 putative effector proteins could be mapped to a domain (Tables S2 and S4). These putative effector proteins belonged to 41 different protein families (domains). The most common domains were the protein of unknown function DUF3455, cutinase/acetylxylan esterase, glycoside hydrolase, necrosis-inducing protein (NPP1), and tuberculosis necrotizing toxin (Table S4).

3.5 | Confirmation of a pathogenicity role of the putative effectors on the PHI-base database

Fifty-six putative effector proteins had homologues in other phytopathogens and were functionally characterized with pathogenicity and virulence roles on the PHI-base database (Table S5). Seventeen of these homologues were present in Fo. This suggests conservation in genes involved in pathogenicity in Fo. Twenty-four of the putative effector proteins were homologous to proteins encoded by genes that have been found to play an effector or plant avirulence determinant role in other phytopathogens. Details of these genes and the phytopathogens are present in Table S5. Only 11 of the 24 putative effector proteins have a domain, with necrosis-inducing proteins (NPP1) and LysM being the most common.

Another 25 of the putative effector proteins had homologues to proteins playing a role in virulence. The deletion of these proteins led to a reduction in virulence of these phytopathogens. The genes encoding these proteins and the phytopathogen details are provided in Table S5. Twenty of these putative effectors have a domain, with pectin lyase fold and cap domain being the most common.

Seven of the putative effector proteins were homologous to proteins encoded by genes necessary for pathogenicity on the PHI-base

database. Details of these pathogenicity genes and phytopathogens are present in Table S5. Interestingly, six of the putative effector proteins had cutinase/acetylxylan esterase as their domain, while one had lactonase.

3.6 | Effector suite of isolates from the agricultural and natural ecosystems

It was presumed that the isolates from natural ecosystems would have a saprophytic lifestyle, while most of the isolates obtained from the agricultural ecosystems would be pathogenic, having been isolated from diseased plants. Putative effectors were predicted from the genomes of the isolates from both the agricultural and natural ecosystems. Although we could not find any unique putative effectors present in isolates from the natural ecosystems, these isolates shared 358 putative effectors with the isolates from the agricultural ecosystems (Table S6). Ninety-eight of the putative effector proteins could be mapped to a domain. Thirty-four of these effectors were previously found to be differentially expressed during infection (Chang et al., 2019; van Dam et al., 2016; Taylor et al., 2016), and 39 of these had homologues that were found to have a role in pathogenicity on the PHI-base database (Table S6). In addition to the 358 putative effectors, 78 putative effectors were unique to the agricultural isolates (Figure 1). Thirty-six of these effectors were previously found to be differentially expressed during infection (Chang et al., 2019; van Dam et al., 2016; Taylor et al., 2016). Nineteen of the unique putative effectors were homologous to genes encoding proteins involved in pathogenicity on the PHI-base database (Table S6). Only 15 of the unique putative effector proteins from the agricultural ecosystem could be mapped to a domain.

Within the agricultural ecosystems, there were isolates from three characterized ff. sp., Fol, Fon, and Fop. Analyses of the effector suites
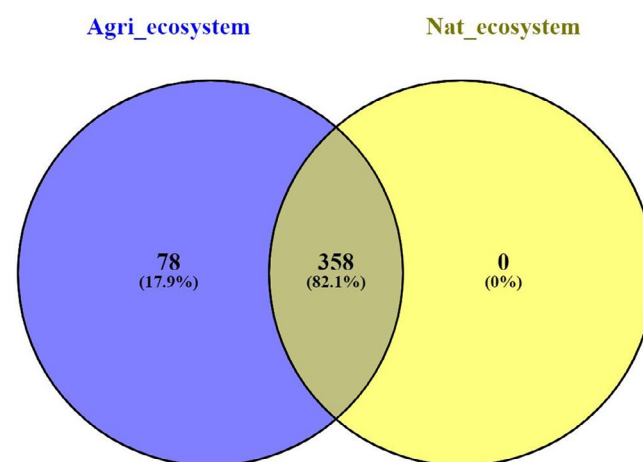

FIGURE 1 Venn diagram showing the relationship between the putative effectors identified in the genome sequences of isolates from agricultural and natural ecosystems [Colour figure can be viewed at wileyonlinelibrary.com]
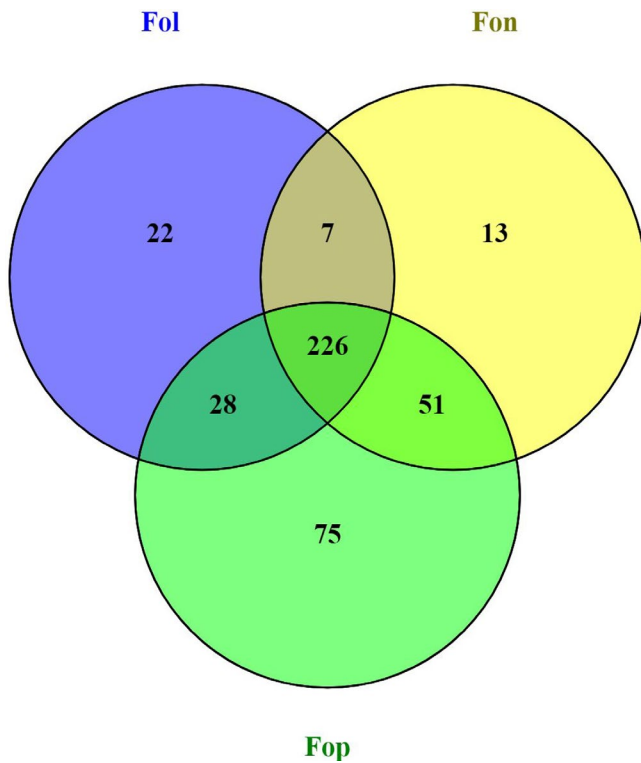
3.6.1 | Presence–absence of putative effectors in all the isolates

From the 85 assembled genome sequences, 436 putative effectors were identified using the three different pathways. These effectors were present in different combinations in the genomes (Table S8). The presence–absence phylogeny (Figure 3) showed that some isolates from the natural ecosystems clustered together. These isolates also clustered with other putatively nonpathogenic isolates from the agricultural ecosystems. These putatively nonpathogenic isolates were isolated from diseased plants but were not identified as the disease-causing agents on those hosts. The characterized isolates of Fon and Fop clustered with other isolates from the same ff. sp. forming the *niveum* and *pisi* clusters, respectively. Fol4287 clustered with VPRI11681, which was isolated from a diseased tomato plant, and further analysis of the *SIX3* gene showed that it was homologous with that found in Fo f. sp. *lycopersici* race 2 isolates such as Fol4287, indicating that it is likely to be this particular pathogen. Two of the isolates, VPRI42198 and VPRI16963, clustered together. These were isolated from diseased potato plants and tubers and were considered to be the disease-causing agents. In a previous study by Achari et al. (2020), it was shown that the FOSC comprises three clades, which were defined as putative species. Interestingly, we can see some clustering based on these groupings in the effector presence–absence phylogeny of the present study.

4 | DISCUSSION

The main aim of this study was to predict the effector suites of Fo isolates from agricultural and natural ecosystems in Australia. Previous studies on Fo effectors have concentrated on pathogenic Fo isolates from agricultural ecosystems, reflecting its economic importance. To our knowledge, this is the first report comparing the effector suites of Fo isolates from agricultural and natural ecosystems. Previous studies on isolates from natural ecosystems identified an intermediate and low abundance of some putative effector genes, such as pisatin demethylase 1 (*PDA1*) and *SIX* genes (Rocha et al., 2016). We created three different pathways to ensure that most, if not all, the effector-like proteins could be predicted from the 83 Australian isolates and two published genomes. None of the pathways was robust enough to predict the putative effectors that could make the isolates cluster based on the host plant; however, the combination of putative effectors predicted from the three pathways produced a phylogeny whereby the isolates clustered based on the host plant. Some putative effectors were predicted from both Pathways 1 and 2; however, most of them were unique to the pathways. Transcriptomics data of these isolates could further optimize the findings from the pathways by identifying putative effectors involved in host–pathogen interactions.

Using these pathways, we were able to predict 436 putative effectors in silico. Pathway 1 was based on the protein characteristics of effector-like proteins. Pathway 2 was based on the relationship
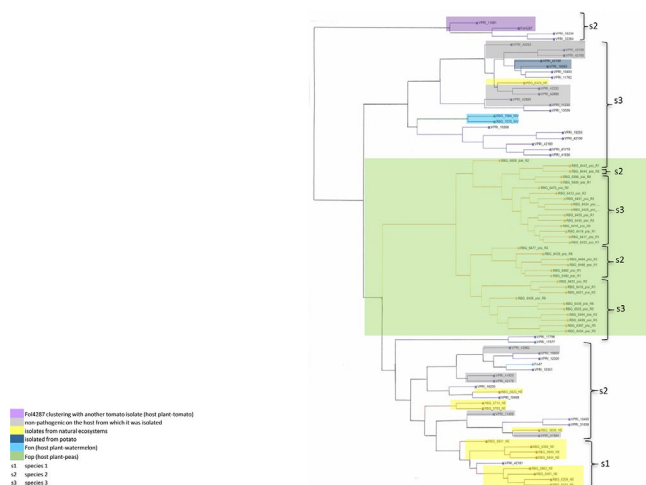

FIGURE 2 Venn diagram showing the relationship between the putative effectors identified in the genome sequences of *Fusarium oxysporum* ff. sp. *lycopersici* (Fol), *niveum* (Fon), and *pisi* (Fop)

FIGURE 3 Putative effector presence–absence phylogeny showing the clustering of isolates based on the host plant. The putative species information is also shown on the phylogeny

from these characterized ff. sp. revealed that they shared 226 putative effectors, while Fol, Fon, and Fop had 22, 13, and 75 unique putative effectors, respectively (Figure 2, Table S7). Forty-five of the 226 putative effector proteins shared by the three ff. sp. had a domain, while only 8, 1, and 13 of the putative effector proteins unique to Fol, Fon, and Fop, respectively, had a domain.

of *SIX* genes with transposons (Schmidt et al., 2013). However, we do not know whether this relationship is valid for other putative effectors apart from *SIX* genes. This pathway was previously used by van Dam et al. (2016). The results from this pathway are very much dependent on how well the genome assemblies are created, and to avoid missing putative effectors due to this, we used two additional pathways. We wanted to replicate the van Dam et al. (2016) pathway; hence we used SignalP v. 4.0 (Petersen et al., 2011) to select the secretory proteins in Pathway 2. In Pathway 1, we used SECRETOOL for predicting secretory proteins because this program uses TargetP and SignalP for predicting secretory proteins; hence it is more stringent. Additionally, it also identifies proteins with the transmembrane domains. Proteins with this domain need to be removed from secretory proteins as they are transmembrane bound. Because we had to screen a large pool of predicted proteins, using a program that combines all these screening tools rather than using individual programs with lower sequence input restrictions seemed like a more efficient approach. This also made the prediction of secretory proteins faster. Pathway 1 predicted 14,294–25,037 proteins per isolate, and SECRETOOL can process all these protein sequences in a single batch submission per isolate, while SignalP v. 4.0 can only process 5,000 sequences per batch submission.

Characteristics that define effector-like proteins are very broad and are often insufficient to capture all the effector-like proteins in silico. Using three pathways provided a composite approach to cater to effector characteristics of being small, secretory, not having significant homology to any known protein, and being located close to a transposon in the adaptive genome. Combining transcriptomics data with in silico mining would have strengthened the effector mining pathways, but was unavailable. Twenty percent of the predicted effector-like proteins were previously expressed in planta (Chang et al., 2019; van Dam et al., 2016; Taylor et al., 2016), showing a role in host interactions, while a further 12.8% of the putative effectors were homologous to genes involved in pathogenicity and virulence.

Effector prediction through sequence identity in Fo is challenging due to high effector diversity lacking sequence conservation, except for the *SIX* genes. Effectors are found predominantly in the most dynamic compartments of the genome (Ma et al., 2010), which provides the opportunity for an accelerated rate of sequence evolution, leading to impressive levels of effector diversity. Domains are distinct structural or functional units in a protein that are evolutionarily conserved and are used to determine the protein function (Illergard et al., 2009). Due to high selective pressure, leading to the rapid diversification of protein sequences, many putative effector proteins lack sequence similarity to proteins of known function. However, there are some commonly identified domains shared by some effector proteins such as Hce2, ToxA, Phytotoxin PcF protein, putative necrosis-inducing factor (Jones et al., 2018) and LysM domain (Illergard et al., 2009). Having conserved domains for putative effectors provides another characteristic when in silico mining for such proteins. Only 25.9% of the putative effector proteins had a domain, suggesting that many of these proteins are under selective pressure and could not be mapped to known domains or proteins

of known functions. There were two Ecp2 proteins with the LysM domain, belonging to the Hce2 superfamily. Additionally, there were another four proteins with the LysM domain and seven proteins with a necrosis-inducing protein (NPP1) domain. Other protein domains were also present, such as the protein of unknown function (DUF3455), cutinase/acetylxylam esterase, glycoside hydrolase, tuberculosis necrotizing toxin, and pectin lyase fold.

Sequence homology searches revealed that most of the putative effector proteins had homologues conserved in *Fusarium* and other fungal genera. The majority of the putative effector proteins had homologues in other *Fusarium* species, suggesting that these may be ancestral proteins within *Fusarium*. It is hypothesized that these putative effector proteins may have evolved independently in Fo and other *Fusarium* species and were subject to different selection pressures, leading to sequence variations. A further 10% of the putative effector proteins had homologues conserved in Fo only. These proteins may have coevolved with the host or horizontally transferred between the Fo strains and then evolved independently in different ff. sp. Twenty-one percent of the putative effector proteins were present in Fo and other fungal genera. This category included *SIX* gene-encoded proteins (3, 5, 6, 11, 13, and 14). We found homologues of *SIX* gene encoded proteins in other Fo isolates as well as in *Colletotrichum*, *Rhynchosporim*, *Verticillium*, and *Exserohilum* genera. Sequence similarity analysis showed that there was a higher sequence similarity within the FOSC in comparison to any other genera. This suggests that these proteins may have evolved independently in Fo and the other fungal genera, or they could have been acquired after speciation through horizontal gene transfer from Fo and then evolved separately. No homologue could be detected for 3% of the putative effector proteins. These may be very specific to the isolates in this study or may have evolved so rapidly that they could not fit into the criteria used while searching for homologues.

Comparing the effector suite of isolates from the agricultural and natural ecosystems showed that effector-like proteins are also present in Fo living in the natural ecosystem (presumed saprophytes). Effectors act to modulate host cell physiology to promote susceptibility to pathogens. In pathogenic isolates, effectors are involved in host interactions. Saprophytes do not interact with living host plants for accessing nutrients, yet they possess effector-like proteins. The effectors shared between isolates from both ecosystems had domains involved in polysaccharide degradation. Having many polysaccharide-degrading proteins would enable the saprophytes to explore more materials as a potential food source. However, there were also proteins with domains involved in host interactions such as necrosis-inducing proteins (NPP1), peroxidases, and Ecp2 proteins. These proteins are present in phytopathogenic fungi on various hosts (Gijzen and Nürnberger, 2006; Mir et al., 2015; Stergiopoulos et al., 2012), suggesting that these are part of the pathogenic core and target broad plant defence mechanisms. Ecp2 was initially described as an effector in the tomato pathogen *Cladosporium fulvum* and has also been detected in saprophytic fungi (Stergiopoulos et al., 2012).

Thirty-nine putative effectors present in the natural ecosystem isolates were homologous to genes involved in pathogenicity

on the PHI-base database. Because the shared putative effectors are important for parasitic fitness, these saprophytes may convert into pathogenic isolates if they have access to a compatible host. Furthermore, success for Fo, a soilborne pathogen, requires it to be able to cause infection on a host while at the same time compete with numerous other microbial species present in its environment and survive as a saprophyte. This suggests that effectors may be playing more than a host–pathogen interaction role in Fo. It can be hypothesized that putative effectors shared between the isolates from the agricultural and natural ecosystems may have dual roles in parasitic fitness and ecological survival. An understanding of the microbial ecology is needed to fully understand the role of these putative effectors in parasitic fitness and ecological survival. Twenty-eight percent of the putative effector proteins present in the saprophytes could be mapped to a domain, compared to 19% of the putative effectors that were unique to the agricultural isolates. This shows a more rapid evolution of the putative effectors unique to the agricultural isolates.

Isolates from the agricultural ecosystems had more putative effectors because pathogenic isolates need a suite of effectors to overcome many immune receptors present in plants that could recognize them. Fo, being a hemibiotroph, has a biotrophic and a necrotrophic infection stage, and at each stage it releases a wave of effectors having different functions (Toruño et al., 2016). Effectors expressed at the biotrophic stage are involved in suppressing cell death and immunity, while effectors expressed at the necrotrophic stage are responsible for stimulating cell death and immunity (Toruño et al., 2016). Additionally, it has been shown that Fo strains have certain effectors that determine host specificity (van Dam et al., 2016).

Comparison of the putative effector suites of the three characterized Fo ff. sp. showed that there were 226 putative effectors shared by all of them. We hypothesize that these putative effectors may be involved in core pathogenicity functions in Fo. These data also suggest that in Fo, the pathogenic mechanism is highly conserved irrespective of the ff. sp. Thirty-four percent of these putative effector proteins could be mapped to a domain. These proteins had CAP, cutinase/acetylxylan esterase, LysM, necrosis-inducing (NPP1), pectin lyase, rapid alkalinization factor, tuberculosis necrotizing toxin, and protein of unknown function DUF3455 as their domains. These proteins are not as rapidly evolving as the putative effectors that are unique to each f. sp.. Fol, Fon, and Fop had 36%, 7%, and 17% of the unique putative effectors mapped to a domain, respectively, showing a more rapid evolution of the putative effectors involved in host specificity. Host specialization is a well-documented phenomenon in many fungal pathogen–plant host systems (Sanchez-Vallet et al., 2018). In Fo, effectors are the genetic factors determining host specificity (van Dam et al., 2016). Host specificity in pathogenic isolates of Fo may be due to the convergent evolution of the pathogen with the host. This theory is further supported by the effector presence–absence phylogeny, which shows isolates clustering based on the host plant species. Convergent evolution of the putative effectors with the host has led to differences in effectors between the ff.sp. The coevolution of the pathogen with the host is often described as an evolutionary arms race, which is more apparent in agroecosystems where large monocultures pose very specific selection pressures on the pathogen population, leading to effector suites specialized to the hosts (Depotter et al., 2020). There was also some clustering of the isolates based on the putative species identified by Achari et al. (2020). This indicates that the effectors in Fo had initially coevolved with the host and then diverged into species, therefore clustering based on the host rather than with species.

This study showed that putative effectors are also present in Fo from the natural ecosystem and may be playing a role in both parasitic and ecological fitness. Although there is high conservation in the pathogenicity mechanism of Fo, there are also putative effectors that determine host specificity. These putative effectors should be explored for the molecular characterization of pathogenic isolates. The putative effectors from isolates recovered from agricultural ecosystems and, in particular, those that are responsible for host specificity, are under more diversifying selection pressure.

DATA AVAILABILITY STATEMENT
The gene and protein sequence data that support the findings of this study are available in Table S2. Other data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Saidi R. Achari https://orcid.org/0000-0002-4086-6250

REFERENCES

Achari, S.R., Kaur, J., Dinh, Q., Mann, R., Sawbridge, T., Summerell, B.A. et al. (2020) Phylogenetic relationship between Australian *Fusarium oxysporum* isolates and resolving the species complex using the multispecies coalescent model. *BMC Genomics*, 21, 248.

Almagro Armenteros, J.J., Sonderby, C.K., Sonderby, S.K., Nielsen, H. & Winther, O. (2017) DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics*, 33, 3387–3395.

Benjamin, B., Chao, X. & Daniel, H.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12, 59–60.

Chang, W., Li, H., Chen, H., Qiao, F. & Zeng, H. (2019) Identification of mimp-associated effector genes in *Fusarium oxysporum* f. sp. *cubense* race 1 and race 4 and virulence confirmation of a candidate effector gene. *Microbiological Research*, 232, 126375.

Cortázar, A.R., Aransay, A.M., Alfarom, M., Oguiza, J.A. & Lavín, J.L. (2014) SECRETOOL: integrated secretome analysis tool for fungi. *Amino Acids*, 46, 471–473.

Depotter, J.R.L. & Doehlemann, G. (2020) Target the core: durable plant resistance against filamentous plant pathogens through effector recognition. *Pest Management Science*, 76, 426–431.

Fokkens, L., Shahi, S., Connolly, L.R., Stam, R., Schmidt, S.M., Smith, K.M. et al. 2018. The multi-speed genome of *Fusarium oxysporum* reveals association of histone modifications with sequence divergence and footprints of past horizontal chromosome transfer events. bioRxiv. https://doi.org/10.1101/465070 [preprint]

Gijzen, M. & Nürnberger, T. (2006) Nep1-like proteins from plant pathogens: recruitment and diversification of the NPP1 domain across taxa. *Phytochemistry*, 67, 1800–1807.

Holt, C. & Yandell, M. (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, 12, 491.

Illergard, K., Ardell, D.H. & Elofsson, A. (2009) Structure is three to ten times more conserved than sequence—A study of structural response in protein cores. *Proteins: Structure. Function, and Bioinformatics*, 77, 499–508.

Jones, D.A.B., Bertazzoni, S., Turo, C.J., Syme, R.A. & Hane, J.K. (2018) Bioinformatic prediction of plant–pathogenicity effector proteins of fungi. *Current Opinion in Microbiology*, 46, 43–49.

Jones, J.D. & Dangl, J.L. (2006) The plant immune system. *Nature*, 444, 323–329.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C. et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236–1240.

Laurence, M.H., Burgess, L.W., Summerell, B.A. & Liew, E.C.Y. (2012) High levels of diversity in *Fusarium oxysporum* from non-cultivated ecosystems in Australia. *Fungal Biology*, 116, 289–297.

Li, H. (2016) Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, 32, 2103–2110.

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

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M. et al. (2015) Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, 66, 513–545.

Ma, L., Houterman, P.M., Gawehns, F., Cao, L., Sillo, F., Richter, H. et al. (2015) The *AVR2–SIX5* gene pair is required to activate *I-2*-mediated immunity in tomato. *New Phytologist*, 208, 507–518.

Ma, L., van der Does, H., Borkovich, K., Coleman, J., Daboussi, M., Di Pietro, A. et al. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, 464, 367–373.

Mir, A.A., Park, S.-Y., Abu Sadat, M., Kim, S., Choi, J., Jeon, J. et al. (2015) Systematic characterization of the peroxidase gene family provides new insights into fungal pathogenicity in *Magnaporthe oryzae*. *Scientific Reports*, 5, 11831.

Petersen, T.N., Brunak, S., von Heijne, G. & Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8, 1701.

Rocha, L.O., Laurence, M.H., Ludowici, V.A., Puno, V.I., Lim, C.C., Tesoriero, L.A. et al. (2016) Putative effector genes detected in *Fusarium oxysporum* from natural ecosystems of Australia. *Plant Pathology*, 65, 914–929.

Sanchez-Vallet, A., Fouche, S., Fudal, I., Hartmann, F.E., Soyer, J.L., Tellier, A. et al. (2018) The genome biology of effector gene evolution in filamentous plant pathogens. *Annual Review of Phytopathology*, 56, 21–40.

Schmidt, S.M., Houterman, P.M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B. et al. (2013) MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics*, 14, 119.

Schmidt, S.M., Lukasiewicz, J., Farrer, R., van Dam, P., Bertoldo, C. & Rep, M. (2016) Comparative genomics of *Fusarium oxysporum* f. sp. *melonis* reveals the secreted protein recognized by the *Fom-2* resistance gene in melon. *New Phytologist*, 209, 307–318.

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

Sperschneider, J., Gardiner, D.M., Dodds, P.N., Tini, F., Covarelli, L., Singh, K.B. et al. (2016) EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytologist*, 210, 743–761.

Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*, 7, 62.

Stergiopoulos, I., Kourmpetis, Y.A., Slot, J.C., Bakker, F.T., De Wit, P.J. & Rokas, A. (2012) In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. *Molecular Biology and Evolution*, 29, 3371–3384.

Summerell, B. (2019) Resolving *Fusarium*: current status of the genus. *Annual Review of Phytopathology*, 57, 323–339.

Taylor, A., Vágány, V., Jackson, A.C., Harrison, R.J., Rainoni, A. & Clarkson, J.P. (2016) Identification of pathogenicity-related genes in *Fusarium oxysporum* f. sp. *cepae*. *Molecular Plant Pathology*, 17, 1032–1047.

Toruño, T.Y., Stergiopoulos, I. & Coaker, G. (2016) Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. *Annual Review of Phytopathology*, 54, 419–441.

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S.Y. et al. (2019) PHI-base: the pathogen–host interactions database. *Nucleic Acids Research*, 48, D613–D620.

van Dam, P., de Sain, M., Ter Horst, A., van der Gragt, M. & Rep, M. (2018) Comparative genomics-based markers: discrimination of host-specificity in *Fusarium oxysporum*. *Applied and Environmental Microbiology*, 84, e01868–e1917.

van Dam, P., Fokkens, L., Schmidt, S., Linmans, J., Kistler, H., Ma, L.-J. et al. (2016) Effector profiles distinguish formae speciales of *Fusarium oxysporum*. *Environmental Microbiology*, 18, 4087–4102.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.