

RESEARCH ARTICLE

Open Access



The *Venturia inaequalis* effector repertoire is dominated by expanded families with predicted structural similarity, but unrelated sequence, to avirulence proteins from other plant-pathogenic fungi

Mercedes Rocafort¹ , Joanna K. Bowen² , Berit Hassing¹ , Murray P. Cox³ , Brogan McGreal² ,
Silvia de la Rosa¹ , Kim M. Plummer⁴ , Rosie E. Bradshaw³ and Carl H. Mesarich^{1*}

Abstract

Background: Scab, caused by the biotrophic fungus *Venturia inaequalis*, is the most economically important disease of apples worldwide. During infection, *V. inaequalis* occupies the subcuticular environment, where it secretes virulence factors, termed effectors, to promote host colonization. Consistent with other plant-pathogenic fungi, many of these effectors are expected to be non-enzymatic proteins, some of which can be recognized by corresponding host resistance proteins to activate plant defences, thus acting as avirulence determinants. To develop durable control strategies against scab, a better understanding of the roles that these effector proteins play in promoting subcuticular growth by *V. inaequalis*, as well as in activating, suppressing, or circumventing resistance protein-mediated defences in apple, is required.

Results: We generated the first comprehensive RNA-seq transcriptome of *V. inaequalis* during colonization of apple. Analysis of this transcriptome revealed five temporal waves of gene expression that peaked during early, mid, or mid-late infection. While the number of genes encoding secreted, non-enzymatic proteinaceous effector candidates (ECs) varied in each wave, most belonged to waves that peaked in expression during mid-late infection. Spectral clustering based on sequence similarity determined that the majority of ECs belonged to expanded protein families. To gain insights into function, the tertiary structures of ECs were predicted using AlphaFold2. Strikingly, despite an absence of sequence similarity, many ECs were predicted to have structural similarity to avirulence proteins from other plant-pathogenic fungi, including members of the MAX, LARS, ToxA and FOLD effector families. In addition, several other ECs, including an EC family with sequence similarity to the AvrLm6 avirulence effector from *Leptosphaeria maculans*, were predicted to adopt a KP6-like fold. Thus, proteins with a KP6-like fold represent another structural family of effectors shared among plant-pathogenic fungi.

*Correspondence: c.mesarich@massey.ac.nz

¹ Laboratory of Molecular Plant Pathology/Bioprotection Aotearoa, School of Agriculture and Environment, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions: Our study reveals the transcriptomic profile underpinning subcuticular growth by *V. inaequalis* and provides an enriched list of ECs that can be investigated for roles in virulence and avirulence. Furthermore, our study supports the idea that numerous sequence-unrelated effectors across plant-pathogenic fungi share common structural folds. In doing so, our study gives weight to the hypothesis that many fungal effectors evolved from ancestral genes through duplication, followed by sequence diversification, to produce sequence-unrelated but structurally similar proteins.

Keywords: *Venturia inaequalis*, Apple scab fungus, Biotrophic subcuticular pathogen, Effectors and effector families, Virulence and avirulence, RNA-seq transcriptome, AlphaFold2 protein tertiary structure predictions

Background

Fungal pathogens are responsible for some of the most devastating diseases of crop plants worldwide, causing large economic losses and threatening both food production and security [1]. Resistance to these pathogens is largely governed by the plant immune system and is based on the recognition of invasion patterns (IPs) by plant immune receptors [2]. At the plant cell surface, many of these immune receptors are pattern recognition receptors (PRRs) of the receptor-like protein (RLP) or receptor-like kinase (RLK) classes, which recognize conserved IPs known as microbe-associated molecular patterns (MAMPs) [3]. Following the recognition of these MAMPs, plant defence responses that slow or halt growth of the fungal pathogen are initiated [3]. To circumvent or suppress these defence responses, plant-pathogenic fungi must secrete a collection of virulence factors, termed effectors, to the plant–pathogen interface, a subset of which may be taken up intracellularly. These effectors are predominantly non-enzymatic proteins, but can also be enzymes, secondary metabolites and small RNAs [4–7]. In some cases, however, effectors can be recognized as IPs by extracellular PRRs or intracellular nucleotide-binding leucine-rich repeat (NLR) immune receptors, collectively referred to as qualitative resistance (R) proteins, to activate plant defences [3, 8]. Such effectors are termed avirulence (Avr) effectors because their recognition typically renders the fungal pathogen unable to cause disease.

Scab (or black spot), caused by *Venturia inaequalis*, is the most economically important disease of apple (*Malus x domestica*) worldwide [9, 10]. Under favourable conditions, this disease can render the fruit unmarketable and cause a yield reduction of up to 70% [10]. During biotrophic host colonization, *V. inaequalis* exclusively colonizes the subcuticular environment without penetrating the underlying plant epidermal cells [9, 11, 12]. It is here that the fungus develops specialized infection structures, known as stromata and runner hyphae [9, 11]. Stromata give rise to asexual conidia but are also likely required for nutrient acquisition and effector secretion [9, 11, 13]. Runner hyphae,

on the other hand, enable the fungus to radiate out from the initial site of host penetration, acting as a base from which additional stromata can be differentiated [11]. At the end of the season, in autumn, *V. inaequalis* switches to saprobic growth inside fallen leaves, where it undergoes sexual reproduction [9].

Scab disease is largely controlled through fungicides, which greatly accelerate the development of fungicide resistance in *V. inaequalis* [14, 15]. Scab-resistant apple cultivars, developed through the incorporation of *R* genes, represent a more sustainable disease control option [16]. However, races of *V. inaequalis* that can overcome resistance in apple mediated by single *R* genes have been identified [16, 17]. Therefore, it is likely that multiple ‘unbroken’ *R* genes, perhaps together with quantitative trait loci (QTL), will need to be stacked in each of these cultivars to provide durable disease resistance [16, 17]. For this to be successful, prior knowledge of the molecular mechanisms used by the fungus to overcome *R* gene-mediated resistance, including an understanding of how these mechanisms impact effector function and pathogen fitness, will be required. So far, however, there have been no publications describing the cloning of *Avr* effector genes from *V. inaequalis*. Crucially, the genomes of multiple *V. inaequalis* isolates have been sequenced [18–23] and bioinformatic studies have identified a large catalogue of secreted, non-enzymatic proteinaceous effector candidates (ECs) from which candidate *Avr* effectors can be identified [19]. Most of these ECs belong to expanded (sequence-related) families [19]; however, the driving force behind the expansion of these families is not yet known. Nevertheless, many *EC* genes are located next to repetitive sequences in the genome of *V. inaequalis* and, therefore, it is possible that transposition is one of the mechanisms driving this expansion [19]. Examples where *EC* genes of *V. inaequalis* are located next to repetitive sequences include members of the *AvrLm6-like* family, which encode proteins with sequence similarity to *AvrLm6* [13], an *Avr* effector from the fungal pathogen *Leptosphaeria maculans* (blackleg of canola) [24], as well as members of the *Ave1-like* family [19], which

encode proteins with sequence similarity to Ave1, an antimicrobial Avr effector from the fungal pathogen *Verticillium dahliae* (Verticillium wilt disease) [25, 26].

To develop durable control strategies against scab disease, a better understanding of the roles that effectors play in promoting subcuticular growth by *V. inaequalis* is also required. To date, subcuticular growth has been largely understudied, even though it is exhibited by many plant-pathogenic fungi, including other crop-infecting members of the *Venturia* genus [11, 27–29], as well as, for example, *Rhynchosporium* (scald disease of graminaceous plants) [30, 31] and *Diplocarpon* (e.g. rose black spot) [32, 33]. In recent years, host colonization by plant-pathogenic fungi has been studied by transcriptomic analysis [34–36]. However, comprehensive transcriptomic studies focusing on the subcuticular parasitic strategy are not yet available. Indeed, while previous expression data from interactions between *V. inaequalis* and susceptible apple have been published [19, 37], these data are only based on a limited number of infection time points with no biological replicates.

In this study, we provide the first comprehensive transcriptomic analysis of *V. inaequalis* during colonization of susceptible apple and identify infection-related temporal expression waves associated with EC genes of this fungus. Using recent advances in de novo protein folding algorithms, we also show that the EC repertoire of *V. inaequalis* is dominated by expanded families with predicted structural similarity to Avr proteins from other plant-pathogenic fungi. Collectively, this study furthers our understanding of subcuticular growth by *V. inaequalis* and provides an enriched list of ECs that can be investigated for potential roles in virulence and avirulence.

Results

The different stages of host infection by *V. inaequalis* observed by bright-field microscopy display distinct gene expression profiles

To investigate changes in *V. inaequalis* gene expression during host colonization and relative to growth in culture, we set up an infection time course involving detached leaves of susceptible apple cultivar *M. x domestica* ‘Royal Gala’ and compared it to growth of the fungus on cellophane membranes overlying potato dextrose agar (PDA). Here, six in planta time points (12 and 24 hours post-inoculation [hpi], as well as 2, 3, 5 and 7 days post-inoculation [dpi]) and one in culture time point (7 dpi) were used.

Analysis of leaf material from the infection time course by bright-field microscopy revealed that, at 12 hpi, conidia of *V. inaequalis* had germinated and formed appressoria on the leaf surface (Fig. 1A). At 24 hpi, primary hyphae had developed, indicating that colonization

of the subcuticular environment was underway. Then, by 2 and 3 dpi, stromata had differentiated from primary hyphae and, in many cases, these stromata had undergone a rapid expansion in size through non-polar division (Fig. 1A). Subcuticular runner hyphae had also started to radiate out from stromata (Fig. 1A). By 5 dpi, fungal biomass had accumulated extensively in the subcuticular environment and additional stromata had started to develop from runner hyphae (Fig. 1A). Often, these stromata had formed conidiophores, from which conidia had developed (Fig. 1A). Finally, at 7 dpi, conidia of *V. inaequalis* had started to rupture through the plant cuticle (Fig. 1A) and the first macroscopic olive-brown lesions, indicative of heavy sporulation, were apparent. This represented the end of the biotrophic infection stage, as detached apple leaves had started to decay after this time point.

Inspection of the RNA sequencing (RNA-seq) data revealed that the percentage of reads mapping to the *V. inaequalis* genome [19] increased as the infection time course progressed (Fig. 1B, Additional file 1: Table S1). Furthermore, the biological replicates clustered robustly within time points and a clear distinction between the early and mid-late infection stages, as well as between the in planta and in culture growth conditions, was observed (Fig. 1C).

Genes of *V. inaequalis* are expressed in temporal waves during infection of apple leaves

We set out to identify which genes of *V. inaequalis* are upregulated during infection of apple, when compared to growth in culture, as these genes are most likely to be required for promoting host colonization. For this purpose, we updated the current gene catalogue for isolate MNH120 [19] to increase the total number of annotated genes, including those that encode ECs, which are notoriously difficult to predict in fungi (Additional file 2: Fig. S1). In total, 24,502 genes, excluding splice variants, were predicted and, of these, 3563 were upregulated and 1462 were downregulated at one or more in planta time points (p value of 0.01 and $\log_2$ -fold change of 1.5) (Additional file 3). It must be pointed out here that our approach was to predict as many genes as possible and, consequently, it is expected that some spurious genes were included in the annotation. However, as many of these spurious genes were anticipated to show a negligible level of expression, most would not have featured in our list of differentially expressed genes, which formed the central focus of our study. For example, of the 24,502 predicted genes, 9284 had a maximum DESeq2-normalized gene expression count across growth conditions of <1, indicating that they were neither expressed under the conditions tested nor upregulated in planta.

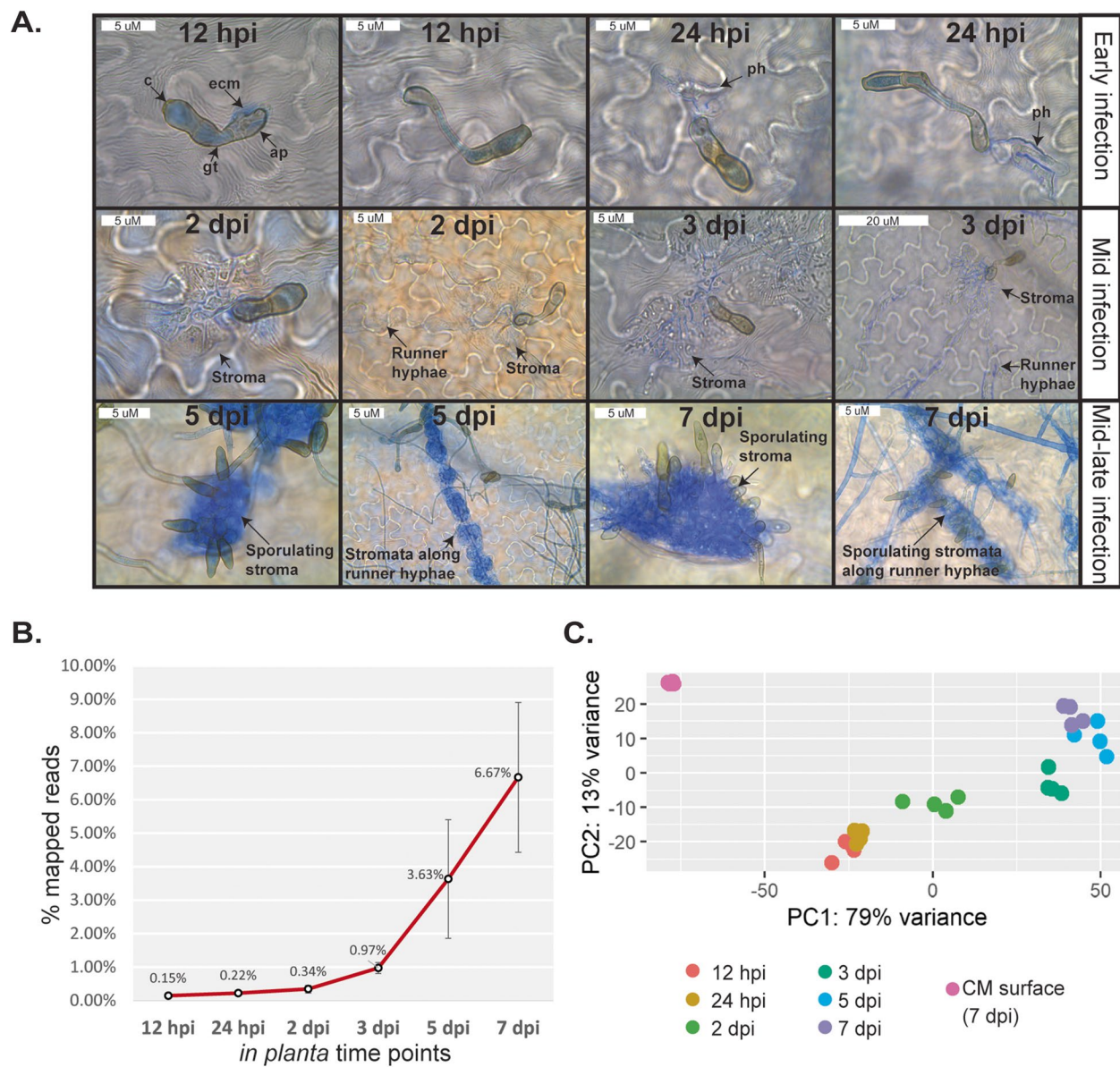


Fig. 1 Microscopic and transcriptomic profile of *Venturia inaequalis* during infection of detached leaves from susceptible apple cultivar 'Royal Gala'. **A** Microscopic evaluation of *V. inaequalis* during colonization of apple leaves. Infection structures observed by bright-field microscopy were stained with aniline blue. Infected leaves are representative of material used in the RNA-seq transcriptome sequencing experiment. c: conidium; gt: germ tube; ap: appressorium; ecm: extracellular matrix; ph: primary hyphae. **B** Percentage (%) of paired RNA-seq reads mapped to the *V. inaequalis* genome relative to the total RNA-seq reads sequenced. Error bars represent standard deviation across four biological replicates. **C** Principal component analysis (PCA) of RNA-seq data from *V. inaequalis* during colonization of apple leaves and in culture on the surface of cellophane membranes (CMs) overlaying potato dextrose agar. Four biological replicates per time point are shown. hpi: hours post-inoculation; dpi: days post-inoculation

The total set of in planta upregulated genes was used to identify temporal host infection-specific gene expression clusters, henceforth referred to as waves. Here, all expression data were scaled across all samples (Z-score) to visualize the gene expression deviation from the overall mean. For hierarchical clustering, the parameters

were set to specifically identify the minimum number of waves for which a distinct gene expression profile could be observed (Fig. 2A). In total, five distinct gene expression waves (Fig. 2A), representing three separate infection stages (Fig. 2B), were identified. More specifically, genes of waves 1 and 2 peaked in expression during early

infection at 12 hpi, with expression largely plateauing (wave 1) or trending downwards (wave 2) throughout the remaining infection time points (Fig. 2A, B). Wave 3 contained genes that peaked in expression during mid infection at 2 dpi (Fig. 2A, B). Genes of wave 4 displayed their lowest level of expression at 12 and 24 hpi, with expression strongly increasing through 2 and 3 dpi, peaking at 5 dpi during mid-late infection (Fig. 2A, B). Finally, for genes of wave 5, a similar profile was observed to genes of wave 4, but with expression strongly increasing from 3 dpi and peaking during mid-late infection at 7 dpi (Fig. 2A, B).

To determine which biological processes are over-represented in the five temporal gene expression waves, gene ontology (GO) (Fig. 2B) and protein family (Pfam) enrichment (Additional file 4) analyses were performed. Genes from waves 1 and 2 were mostly characterized by GO terms associated with high metabolic activity, responses to oxidative stress and cutinase activity. Here, cutinases of carbohydrate esterase family 5 (CE5) were abundant (Additional file 2: Fig. S2). In contrast, genes of wave 3 were mostly characterized by GO terms associated with transmembrane transport (Fig. 2B). Finally, genes of waves 4 and 5 were mostly characterized by GO terms associated with carbohydrate metabolism and transcription (Fig. 2B). In wave 4, for example, a GO enrichment for polygalacturonase activity was observed. This, together with the more general enrichment for carbohydrate metabolism, was supported by the high number of plant cell wall-degrading enzyme (PCWDE)-encoding genes in wave 4, most of which were predicted to encode polygalacturonase enzymes of glycoside hydrolase family 28 (GH28) (Additional file 2: Fig. S2).

Genes encoding non-enzymatic proteinaceous effector candidates of *V. inaequalis* predominantly demonstrate peak expression during the mid-late infection stage of apple

From the 24,502 predicted genes of *V. inaequalis*, 1955 genes were predicted to encode a secreted protein without a transmembrane domain and, of these, 1369 were predicted to encode a non-enzymatic effector candidate (EC). Here, only small, secreted proteins of ≤ 400 amino acid residues in length, as well as larger secreted

proteins with an effector prediction by EffectorP v3.0 [38], were considered to be ECs (Additional file 2: Fig. S3A). Of the 1369 EC proteins, 518 were identical to proteins predicted by Deng et al. [19] (Additional file 2: Fig. S3B). Sequence similarities within our predicted set of ECs were investigated using BLASTp, with the ECs then grouped into families using spectral clustering. Based on this analysis, 759 of the ECs were grouped into 118 families ranging in size from two to 75 members. Of these, 32 families were expanded with five or more members. Due to the observed similarity in sequence between family members, and the fact that expanded families had previously been predicted by Deng et al. [19], we do not expect the gene models encoding these ECs to be spurious. In contrast to the 759 ECs mentioned above, 610 of the ECs were predicted to be singletons that did not belong to any family (Additional file 5). Interestingly, ~22% of the ECs that belonged to families were encoded by genes that either clustered together or were located in close proximity to each other (i.e. within 10 genes) in the genome of isolate MNH120 (Additional file 5).

Based on the differential gene expression analysis presented above, 686 of the 1369 predicted EC genes from *V. inaequalis* were upregulated in planta. To determine when these genes peaked in expression, their expression profile across the five temporal waves during host colonization was investigated. In total, ~20% of the upregulated genes peaked in expression during early infection, ~9% during mid infection and ~71% during mid-late infection (Fig. 3). In all cases, families made up the majority of ECs in each wave (Fig. 3). Most of the EC genes that peaked in expression during early infection (121 genes) encoded proteins that lacked predicted functional domains. Exceptions included the following: (1) a family that encoded proteins with an 'Egh16-like virulence factor' domain (PF11327) and had sequence similarity to the appressorium-specific Gas1 effector from the rice blast fungus *Magnaporthe oryzae* [39], hereafter named the Gas1-like family, (2) a family that encoded proteins with a 'stress-up regulated Nod19' domain (PF07712), hereafter named the Nod19 family, (3) a family that encoded proteins with a 'hydrophobic surface-binding protein A' (HsbA) domain (PF12296), hereafter named the HsbA

(See figure on next page.)

Fig. 2 Genes of *Venturia inaequalis* upregulated during infection of susceptible apple cultivar 'Royal Gala' display five distinct temporal waves of expression. **A** Expression profile of the five distinct temporal expression waves at 12 and 24 h post-inoculation (hpi), as well as 2, 3, 5 and 7 days post-inoculation (dpi), relative to growth in culture on the surface of cellophane membranes overlaying potato dextrose agar at 7 dpi. **B** Heatmap of all *V. inaequalis* genes upregulated in planta when compared with growth in culture. Gene expression data are scaled rlog-normalized counts across all samples (Z-score), averaged from four biological replicates. Genes upregulated in planta were clustered using hclust according to the Ward.D2 and Euclidean distance methods. Early (12 and 24 hpi), mid (2 and 3 dpi) and mid-late infection (5 and 7 dpi) refer to the stage of infection where peak gene expression was observed. Coloured block labels on the left indicate gene expression waves. Numbers in brackets indicate number of genes per wave. The black asterisk indicates the wave significantly enriched for genes encoding effector candidates (ECs) (p value: $3.644e-14$). Grey boxes indicate enriched gene ontology (GO) terms. p : p value

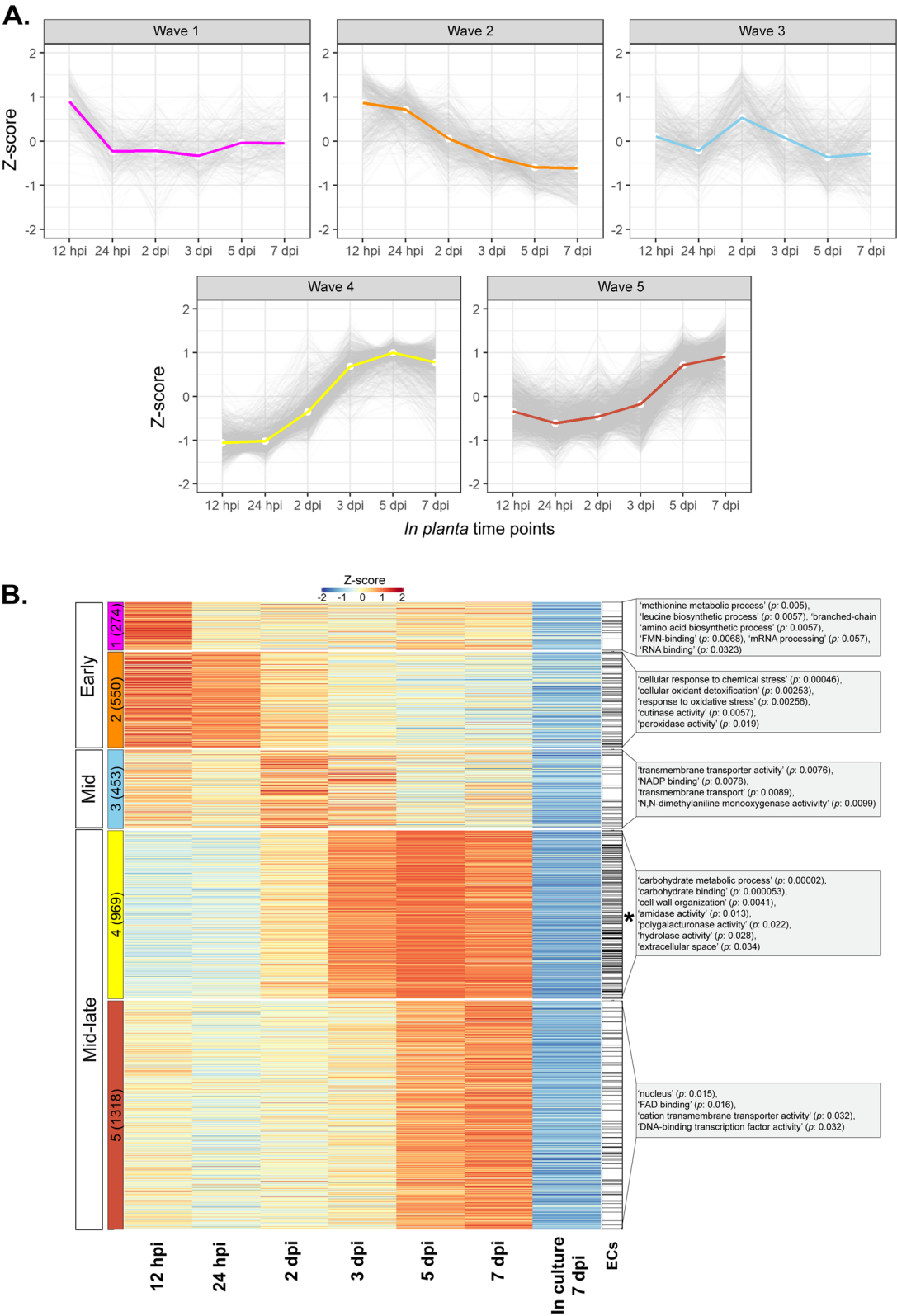
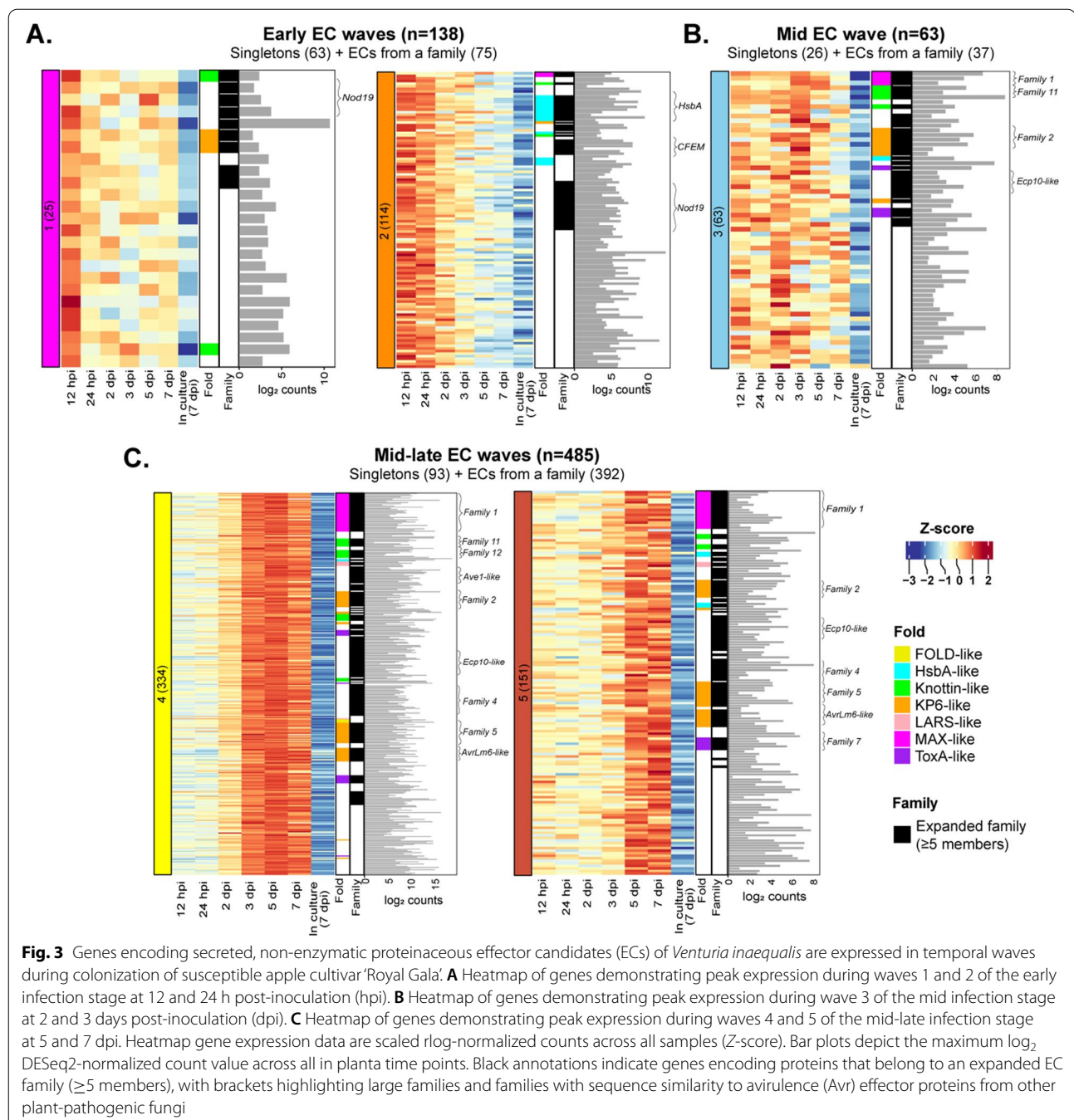


Fig. 2 (See legend on previous page.)



family, and (4) a family that encoded proteins with a ‘common fold in several fungal extracellular membrane proteins’ (CFEM) domain (PF05730) [40], hereafter named the CFEM family.

While the *HsbA* and *CFEM* families possessed genes that mostly peaked in expression during early host colonization (Fig. 3), each of these families also had a smaller number of genes that peaked in expression during waves

4 and 5 of mid-late infection (Additional file 2: Fig. S4). Given that *HsbA* and *cutinase* genes have been shown to be regulated by the same transcription factor in *Aspergillus nidulans* [41, 42] and that most *cutinase* and *HsbA* genes of *V. inaequalis* peaked in expression during wave 2 of the early infection stage (Fig. 3, Additional file 2: Fig. S2), we set out to determine whether the *cutinase* and *HsbA* genes of *V. inaequalis* were co-expressed. Based on

the Pearson correlation coefficient, which was calculated between the *cutinases* and *HsbA* gene expression profiles during the early infection stage, the *cutinase* and *HsbA* genes were indeed found to be co-expressed ($R > 0.8$, $p < 0.01$). Another family with members exhibiting different expression profiles during host colonization was the *Cin1* family (Additional file 2: Fig. S4), which is specific to the *Venturia* genus [43]. This family contains the *Cin1* gene (g8385), which encodes a cysteine-rich protein with eight repeats, and two *Cin1-like* genes, *Cin1-like 1* (g10529) and *Cin1-like 2* (g13013), which encode smaller proteins with only one repeat. *Cin1* peaked in expression during wave 4 and was the most highly expressed gene during mid-late host colonization. In contrast, the *Cin1-like* genes peaked in expression during wave 2 of early infection (Additional file 2: Fig. S4).

The mid and mid-late waves were characterized by 513 genes that mostly encoded EC proteins without a functional domain. The main exception was for members of the Ave1-like family [19], which had an 'RlpA-like domain superfamily annotation' (IPR036908). Notably, most of the expanded EC families that encoded proteins with sequence similarity to effectors or Avr effectors from other plant-pathogenic fungi, such as the *AvrLm6-like* [13] and *Ave1-like* [19] families, displayed a peak level of expression during waves 4 and 5 (Fig. 3, Additional file 1: Table S2). Other examples included the *Ecp10-like* family, which encoded proteins with sequence similarity to the Ecp10-1 Avr candidate from the tomato leaf mold fungus *Fulvia fulva* (formerly *Cladosporium fulvum*), as well as the *Ecp39-like* family, which encoded proteins with sequence similarity to the *F. fulva* EC Ecp39 [44]. An *Ecp6-like* gene (singleton), which encoded a protein with sequence similarity to the Ecp6 effector from *F. fulva* [19, 45], also peaked in expression during wave 4. Interestingly, most expanded EC families that peaked in expression during waves 4 and 5 encoded proteins that lacked sequence similarity to other proteins. Examples included the most expanded EC family in *V. inaequalis*, family 1, as

well members from family 2 (Fig. 4). In most cases, only one or a few family members were very highly expressed during host colonization (Fig. 4).

Finally, as it is known that some fungal effectors are cyclic ribosomally synthesized and post-translationally modified peptides (RiPPs) called dikaritin [46] and that dikaritin precursor peptides are often mistaken for standard secretory proteins, we set out to determine whether any of our 1369 ECs were in fact dikaritin precursor peptides. A hallmark of dikaritin precursor peptides is an N-terminal signal peptide followed by multiple perfect or imperfect tandem repeats [47]. As the genes encoding these precursor peptides form part of a biosynthetic gene cluster that also includes a gene encoding a DUF3328 protein, we screened the *V. inaequalis* genome for DUF3328 genes (and thus, *dikaritin* gene clusters) to determine which ECs could be dikaritins. In total, nine *dikaritin* gene clusters were identified, with each cluster containing one or more *dikaritin* precursor genes. Based on this analysis, ten of the ECs were identified as putative dikaritin precursor peptides and, of these, four were encoded by genes that peaked in expression during mid-late infection (waves 4 and 5) (Additional file 2: Fig S5). The most highly expressed dikaritin precursor gene (g7830, from the dikaritin-2 cluster) corresponded to the previously identified gene *Cin3*, which was formerly considered to encode a repeat-containing EC protein [11, 19].

Several expanded effector candidate families of *V. inaequalis* have predicted structural similarity to Avr effector proteins from other plant-pathogenic fungi

To gain insights into the putative function of ECs, we predicted their tertiary structures using AlphaFold2 [48], and then investigated these structures for similarity to proteins of characterized tertiary structure (and in some cases, function) using the Dali server [49]. This analysis was specifically performed on the most highly expressed member from each EC family (referred to as the representative family member), as well as each singleton,

(See figure on next page.)

Fig. 4 Predicted tertiary structures of secreted non-enzymatic proteinaceous effector candidates (ECs) from *Venturia inaequalis*. **A** Representative EC family members with structural similarity to avirulence (Avr) effector proteins from other plant-pathogenic fungi. MAX-like: representative predicted family 1 protein structure (green; g13386) aligned to an uncharacterized MAX effector from *Magnaporthe oryzae* (6R5J) (purple). ToxA-like: representative predicted family 7 (green; g4781) protein structure aligned to ToxA from *Pyrenophora tritici-repentis* (1ZLE) (purple). LARS-like: representative predicted family 15 (green; g11097) protein structure aligned to AvrLm4-7 from *Leptosphaeria maculans* (7FPR) (purple). FOLD-like: representative predicted family 49 (green; g3787) protein structure aligned to Avr1/Six4 from *Fusarium oxysporum* (7T6A) (purple). **B** Predicted EC family members with a KP6-like fold. KP6-like: representative predicted AvrLm6-like family (green; g20030) protein structure aligned to EC Zt-KP6-1 from *Zymoseptoria tritici* (6QPK). Expression data of upregulated EC genes during colonization of susceptible apple cultivar 'Royal Gala' are DESeq2-normalized counts, averaged from four biological replicates, with error bars representing standard deviation (hpi: hours post-inoculation; dpi: days post-inoculation). Protein structures predicted by AlphaFold2 represent the most highly expressed member of each EC family from *V. inaequalis*. Disulfide bonds coloured in yellow. N: amino (N) terminus; C: carboxyl (C) terminus. pLDDT: predicted Local Distance Difference Test score (0–100). A pLDDT score of 70–100 is indicative of medium to high confidence. A Dali Z-score above 2 indicates 'significant similarities' between proteins. RMSD: root-mean-square deviation. The protein structure of Avr1/Six4 was not present in the Dali database at the time of writing this manuscript and therefore no Dali Z-score is shown

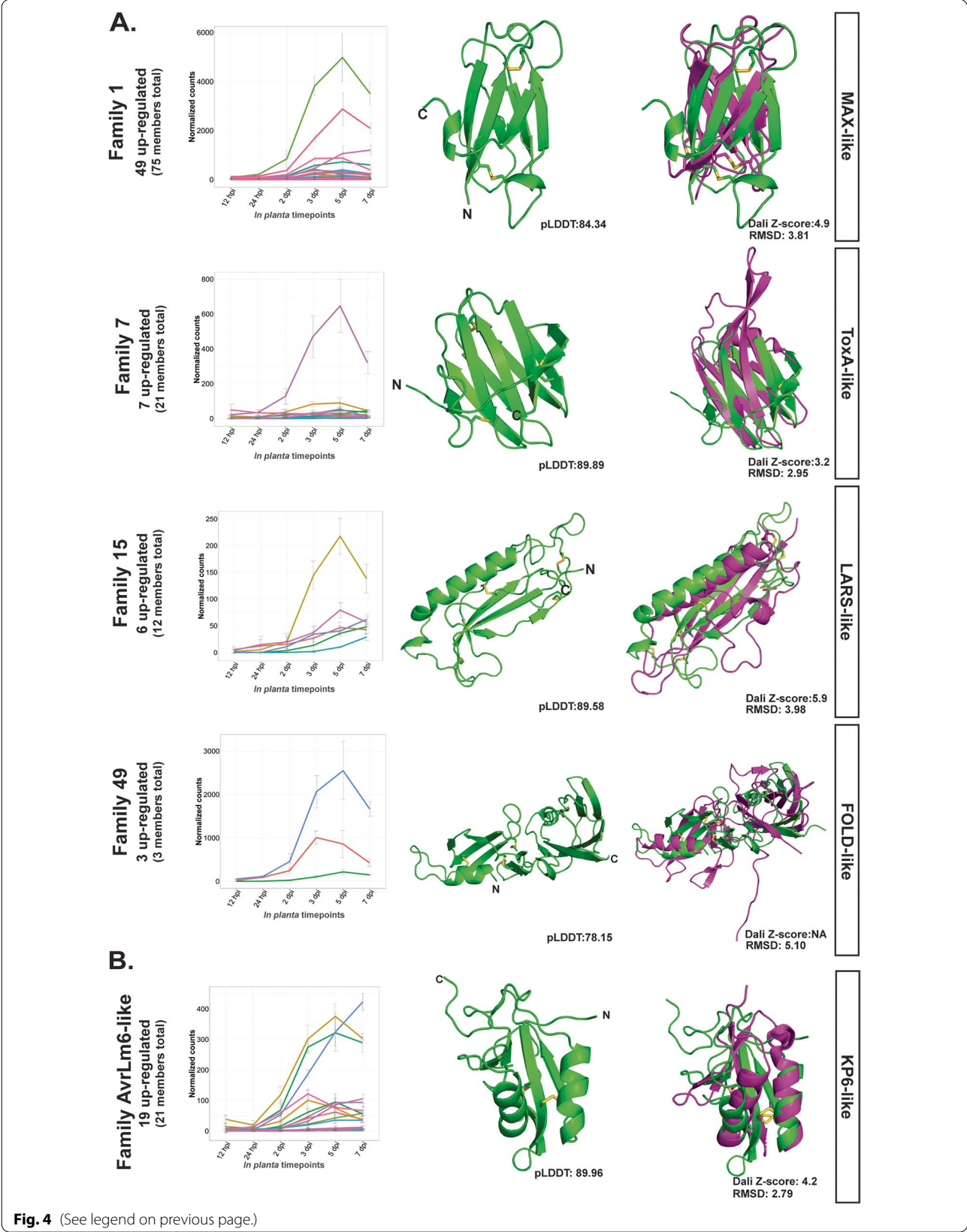


Fig. 4 (See legend on previous page.)

expressed during the temporal host infection-specific waves. In total, the tertiary structure was confidently predicted for the representative family member of 71 (~76%) EC families and 118 (~65%) EC singletons (Additional file 6).

Strikingly, many EC families were predicted to be structurally similar to one or more ECs or Avr effector proteins with solved tertiary structures from other plant-pathogenic fungi (Fig. 4, Additional file 1: Table S3, Additional file 2: Fig. S6). More specifically, 12 EC families were predicted to be structurally analogous to ECs or Avr effector proteins (Additional file 1: Table S3). Remarkably, many of these families were among the most expanded families in *V. inaequalis*. In contrast to the families, only three EC singletons had predicted structural similarity to ECs or Avr effector proteins from other plant-pathogenic fungi (Additional file 1: Table S3).

The representative member of the most expanded EC family in *V. inaequalis*, family 1, was confidently predicted to adopt a six-stranded β -sandwich fold with structural similarity to seven 'Magnaporthe Avr and ToxB-like' (MAX) effectors. The identified MAX effectors were the recently described MAX effector 6R5J [50] and the Avr effectors AvrPiz-t [51], Avr-Pia [52, 53], Avr-Pib [54], Avr1-CO39 [52] and Avr-Pik [55] from *M. oryzae*, as well as the host-selective toxin effector ToxB from *Pyrenophora tritici-repentis* [56], the fungal pathogen responsible for tan spot of wheat (Additional file 1: Table S3). A sequence alignment constructed from all family 1 members of *V. inaequalis*, as well as the identified MAX effectors from *M. oryzae*, revealed that these proteins lacked significant sequence similarity, with a maximum pairwise identity of only 13.9% observed between g13386 and Avr1-CO39 (Additional file 7). However, these sequence-diverse proteins did share the characteristic conserved disulfide bond between β 1 and β 5 that has previously been reported for MAX effectors [52] (Fig. 4A).

Similarly, two expanded families, families 7 and 28, as well as family 38 and a singleton, were predicted to share a common β -sandwich fold with structural similarity to the host-selective toxin effector ToxA from *P. tritici-repentis* [57], the Avr effector Avr2/Six3 from the wilt fungus *Fusarium oxysporum* [58, 59] and the AvrL567 (AvrL567-A and/or AvrL567-D) Avr effectors from the flax rust fungus *Melampsora lini* [60, 61] (Fig. 4A, Additional file 1: Table S3, Additional file 2: Fig. S6). Notably, despite sharing an overall β -sandwich fold with similar topology, these proteins were very diverse at the amino acid level, with a maximum sequence identity of 12.7% (Additional file 7). All of the ToxA-like families had a different number of β -sheets (Additional file 2: Fig. S6).

Another two families, families 15 and 47, were predicted to have structural similarity to the AvrLm4-7

and AvrLm5-9 Avr effectors from *L. maculans*, as well as the Ecp11-1 Avr candidate from *F. fulva* [62, 63] (Fig. 4A, Additional file 1: Table S3, Additional file 2: Fig. S6), which belong to the 'Leptosphaeria avirulence and suppressing (LARS)' structural effector family [63]. The *V. inaequalis* LARS-like effectors had the same predicted topology as AvrLm4-7 but with a variable number of β -sheets. At the amino acid level, the identified *V. inaequalis* LARS-like proteins shared only 7.8–16.9% sequence identity with AvrLm4-7, AvrLm5-9 and Ecp11-1 (Additional file 7). Interestingly, the *V. inaequalis* structures were predicted to be stabilized by a different number of disulfide bonds and lacked both the conserved disulfide bridge between the α -helix and β -strand, as well as the conserved WR(F/L/V)(R/K) motif, previously reported for the LARS effectors [63].

Finally, family 49 was predicted to adopt a two-domain fold similar to the Avr1/Six4 and Avr3/Six1 Avr effectors from *F. oxysporum* (Fig. 4A, Additional file 1: Table S3), which are the founding members of the recently identified *Fol* dual-domain (FOLD) structural family of effectors [64]. Despite the *V. inaequalis* FOLD-like family representative sharing only 17.6% amino acid identity with Avr1/Six4 (Additional file 7), the cysteine spacing pattern was conserved.

Interestingly, many expanded EC families had a KP6-like fold similar to the KP6 protein of the P6 virus from the corn smut fungus *Ustilago maydis* (Additional file 2: Fig. S7) [65, 66] and the Zt-KP6-1 EC from wheat blotch fungus *Zymoseptoria tritici* [67]. This included the AvrLm6-like family, families 2, 5, 23 and 26, as well as two singletons (Additional file 2: Fig. S7). All KP6-like proteins were predicted to have between two and four disulfide bridges, two α -helices and a variable number of β -sheets (Fig. 4B, Additional file 2: Fig. S7).

Unfortunately, many other EC proteins from *V. inaequalis* were too small to be included in Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) searches (or did not have a significant hit) using the Dali server. To gain further insights into their function, we investigated their general SCOPe fold classification and identified three families and one singleton that were predicted to adopt a knottin-like fold (Additional file 6). The most highly expressed member of family 11 (Additional file 2: Fig. S8) and a singleton were predicted to adopt a knottin-like fold with two β -sheets and three disulfide bonds that form an intramolecular knot (Additional file 2: Fig. S8). Additionally, families 12, 24 and 35 were predicted to adopt a knottin-like fold using the SCOPe database [68] (Additional file 2: Fig. S8). However, these proteins were not predicted to have a true intramolecular knot characteristic of knottin proteins based on our current AlphaFold2 predictions. Instead,

these proteins shared a common fold with cysteine-stabilized $\alpha\beta$ defensins, which are made up of a single α -helix and three β -strands.

In addition to these knottin-like proteins, the Ecp10-like family, along with the candidate Avr effector Ecp10-1 from *F. fulva* [44], were predicted to adopt a small compact β -folded structure that is structurally similar to the PAF protein from *Penicillium chrysogenum* (Additional file 2: Fig. S9) [69–71]. Furthermore, the Ecp39-like family, along with the Ecp39 EC from *F. fulva* [44], were predicted to adopt a crambin-like fold (Additional file 2: Fig. S9). Unlike that observed for the Ecp10-like family, no analogous protein structures were identified for the Ecp39-like family in the RCSB PDB using the Dali server. Lastly, the Ave1-like family, together with the Ave1 Avr effector from *V. dahliae* [19, 25], were predicted to adopt a double-psi β -barrel fold, stabilized by two conserved disulfide bonds, with high similarity to expansins (Additional file 2: Fig. S9).

Lastly, we investigated whether a relationship existed between the predicted fold types of EC proteins and the expression profiles of genes that encoded them. Consistent with the results mentioned above, ECs with a predicted HsbA-like fold were predominantly encoded by genes that peaked in expression during wave 2 of early infection (Fig. 3). In contrast, ECs with structural similarity to ECs and Avr effector proteins from other plant-pathogenic fungi predominantly peaked in expression during mid and mid-late infection (Fig. 3). Remarkably, all members of the FOLD-like and LARS-like families peaked in expression during mid-late infection (wave 4 and wave 5). Similarly, the majority of members from the MAX-like (65%), ToxA-like (70%), KP6-like (~59%) and knottin-like (~66%) families peaked in expression during wave 4.

Discussion

In this study, we present the first comprehensive transcriptome of *V. inaequalis* during colonization of apple, covering six biotrophic time points from early to late infection. In doing so, we have, to our knowledge, also provided the first comprehensive in planta transcriptome of a subcuticular fungal pathogen. Based on this transcriptome, we identified five temporal host infection-specific waves of gene expression for in planta upregulated genes of *V. inaequalis*. Here, genes demonstrated peak expression during one of three stages of host colonization corresponding to early (12 and 24 hpi; waves 1 and 2), mid (2 and 3 dpi; wave 3) and mid-late (5 and 7 dpi; waves 4 and 5) infection. These temporal gene expression waves were biologically distinct from each other and were enriched for different GO terms.

A key focus of our study was to understand the expression profile of EC genes that encode secreted, non-enzymatic proteins during host colonization, as these genes make up the bulk of effectors identified from plant-pathogenic fungi to date [4, 6]. Interestingly, during early host colonization, when *V. inaequalis* is either growing on the leaf surface or has just initiated subcuticular growth, only around 20% of the upregulated EC genes peaked in expression. In other plant-pathogenic fungi, however, the percentage of EC genes that peak in expression during early host colonization is much higher. For instance, in *U. maydis*, ~40% of genes encoding secreted proteins were found to be specifically induced during early host colonization [34]. One possible reason for this difference could be that, during the early infection stage, *V. inaequalis* predominantly colonizes the epicuticular wax above apple epidermal cells, perhaps on its way to colonizing the subcuticular environment. This may suggest that not all infections have yet resulted in close contact with the underlying epidermal cells and, consequently, the mass upregulation of genes encoding effector proteins with roles in suppressing host defences has not yet been initiated.

Among the EC genes that peaked in expression during early colonization, several belonged to the *HsbA* family. *HsbA* proteins have been suggested to recruit cutinases to hydrophobic surfaces [42] and, in *A. nidulans*, *HsbA* and *cutinase* genes are co-regulated by the same transcription factor [41, 42]. Consistent with this previous research, we observed that some *HsbA* and *cutinase* genes of *V. inaequalis* were co-expressed, suggesting that, as previously hypothesized, the *HsbA* proteins of *V. inaequalis* may recruit cutinases to facilitate the degradation and digestion of the hydrophobic apple cuticle during early host infection [19, 42, 72]. In line with this observation, early infection was enriched for the GO term ‘cutinase activity’ and many CE cutinase-encoding genes were upregulated. Altogether, these observations support previous reports showing that localized enzymatic hydrolysis is needed for penetration of the apple cuticle by *V. inaequalis* to facilitate access to the subcuticular environment [9, 73].

Other EC genes that peaked in expression during early host colonization included members of the *Gas1-like* family, which encode proteins with a ‘Egh16-like virulence factor’ domain, and members of the *Nod19* family, which encode proteins with a ‘stress upregulated Nod19’ domain. Recently, *Gas1-like* proteins have been shown to form part of the widely distributed fungal ‘effectors with chitinase activity’ (EWCA) family [74]. EWCA proteins are secreted chitinases without a characterized enzymatic domain in the carbohydrate-active enzyme (CAZyme) database that degrade immunogenic chitin

fragments to prevent chitin-triggered immunity in plants [74]. Thus, it is tempting to speculate that members of the Gas1-like family from *V. inaequalis* play a similar role upon access to the subcuticular environment. The function of proteins with a stress upregulated Nod19 domain is currently unknown. However, it has been suggested that these proteins are associated with responses to abiotic and biotic stress [75–78]. Based on these studies and the expression of the *Nod19* genes during early infection, the Nod19 family from *V. inaequalis* could play a role in modulating oxidative stress during early subcuticular colonization of apple. Crucially, other genes associated with oxidative stress tolerance, such as those encoding peroxidases, were also enriched during early infection, suggesting that modulation of oxidative stress is crucial for early host colonization by *V. inaequalis*.

Like the *HsbA* family, most members of the *CFEM* family peaked in expression during early host colonization. The *CFEM* domain is found in several fungal proteins, including those of plant pathogens [79], where it has been shown to confer a diverse range of functions ranging from the promotion or suppression of plant cell death and chlorosis [80–82] to the development of appressoria [83]. Consistent with this functional diversity, and because some members of the *CFEM* family from *V. inaequalis* instead demonstrated a peak level of expression during mid-late infection, it is likely that the *CFEM* proteins also play a diverse range of roles in *V. inaequalis* during colonization of apple.

The mid infection stage, which was characterized by the large-scale expansion and continued differentiation of subcuticular infection structures (i.e. stomata and runner hyphae), was enriched for GO terms associated with transmembrane transport. Interestingly, while many *EC* genes of *V. inaequalis* were highly expressed during this infection stage, very few displayed their peak level of expression here. This may suggest that the expression of most *EC* genes is still increasing during mid infection. However, it is important to point out that the mid infection wave was not well defined, overlapping partially with the mid-late infection waves. This is presumably due to the asynchronous nature of *V. inaequalis* infection.

Finally, during the mid-late infection stage, when *V. inaequalis* is heavily colonizing the subcuticular space, most *EC* genes peaked in expression. Based on this expression profile, and the fact that their expression steadily ramped up from the onset of the infection process, we believe that these genes likely play a key role in the establishment and maintenance of biotrophy. Intriguingly, many genes encoding GH enzymes associated with the degradation of the plant cell wall, such as pectin-degrading GH28 proteins, also peaked in expression during this infection stage. As nutrients in the subcuticular

environment are likely to be scarce throughout host colonization, it is anticipated that *V. inaequalis* meets a portion of its nutritional requirements through the degradation of the pectin-rich layer located between the cuticle and epidermal cells of apple using these enzymes [84]. Related to this, it is well known that fungal GH28 enzymes can be recognized as MAMPs, while host cell wall fragments released as a consequence of GH28 hydrolytic activity can be recognized as damaged-associated molecular patterns (DAMPs), by PRRs, to activate the plant immune system [7]. With this in mind, a subset of the *ECs* encoded by genes that peaked in expression during mid-late infection may function to suppress plant defences responses initiated by these GH28 enzymes.

Remarkably, many of the *EC* genes that peaked in expression during mid-late infection encoded proteins that belonged to expanded families. Such a phenomenon, where *ECs* are known to form part of expanded families, has also been observed in other lineage-specific pathogens, including *Blumeria graminis* [85, 86]. Although the relevance of *EC* family expansion to *V. inaequalis* is not yet well understood, it is anticipated that this expansion facilitates the diversification of effector function and enables the avoidance of recognition by cognate host R proteins [19]. In any case, repetitive elements are expected to play a major role in the expansion process, enabling both *EC* gene duplication and subsequent transposition to other regions of the *V. inaequalis* genome [87, 88]. In line with this, it has previously been shown that *ECs* of *V. inaequalis*, including members of the *Ave1*-like and *AvrLm6*-like *EC* families, tend to be closely associated with repetitive elements [13, 19]. Moreover, *EC* genes belonging to the same expanded families have been shown to cluster together in the *V. inaequalis* genome (this study), indicating that tandem duplication events might have occurred. To provide further insights into the process of *EC* family expansion in *V. inaequalis*, a chromosome-level genome assembly of this fungus is now required as, due to its highly fragmented nature (1012 scaffolds) [19], the current Illumina genome provides an incomplete picture of repeat composition and gene clusters.

To gain insights into the function of *EC* proteins from *V. inaequalis*, we used the de novo folding algorithm AlphaFold2 to predict their tertiary structures, as AlphaFold2 has been successfully benchmarked against effectors of characterized tertiary structure from other plant-pathogenic fungi [64, 89]. One of the main limitations when using AlphaFold2 is that proteins with a low number of homologous sequences in public databases normally result in predictions with low confidence scores. In an attempt to overcome this, we generated custom multiple sequence alignments (MSAs) that included

the amino acid sequences of all EC family members identified in this study, many of which were not available in public sequence databases, which greatly improved prediction scores (Additional file 2: Fig. S10).

Strikingly, many of the EC families, especially the expanded EC families, demonstrated predicted structural similarity to Avr effector proteins from other plant-pathogenic fungi. The biggest of these was the MAX-like family, which had predicted structural similarity to one of the largest effector/EC families from *M. oryzae*, the MAX family [90, 91]. Intriguingly, a recent computational study of secreted proteins from multiple plant pathogens based on AlphaFold2 concluded that the MAX fold was almost exclusive to *M. oryzae*, with a few members of this family also found in the vanilla black spot fungus *Colletotrichum orchidophilum* [91]. However, here we show that the MAX-like family has undergone massive expansion and diversification in *V. inaequalis*, highlighting the need for more comprehensive sampling of fungal species using AlphaFold2 to better understand the evolutionary origin and distribution of the MAX structural fold.

In *M. oryzae*, Avrs of the MAX effector family are translocated into host cells, where they are recognized by NLR R proteins [51, 54, 92–95]. Of these, Avr-PikD and Avr1-CO39 directly interact with their corresponding NLR R proteins; an interaction mediated through a heavy metal-associated (HMA) domain that is integrated into the R protein itself. Similarly, the MAX effector Avr-Pik binds and stabilizes an independent HMA protein to modulate host immunity [96]. Altogether, these studies suggest that the MAX fold could be well suited to interactions with HMA domains [97, 98]. It is therefore tempting to speculate that ECs of the MAX-like family from *V. inaequalis* are translocated into host cells, where they interact with HMA domain-containing proteins of apple. Certainly, as Rvi15, an R protein of apple that recognizes the AvrRvi15 Avr effector of *V. inaequalis*, is an NLR [99], it seems likely that a subset of Avrs from this fungus are translocated into host cells.

EC families and singletons with predicted structural similarity to ECs and Avr proteins from the ToxA-like family were also identified in *V. inaequalis*. Interestingly, like that observed in *V. inaequalis*, the recent computational study by Seong and Krasileva [91] showed that ToxA-like ECs are greatly expanded in the cereal stem rust fungus *Puccinia graminis*. The same study also showed that a further four species, *F. oxysporum*, *C. orchidophilum*, *V. dahliae* and *U. maydis*, have members of this family [91]. However, in the case of these four species, only a few members of the ToxA family could be identified. Thus, *V. inaequalis* appears to have one of the largest repertoires of ToxA-like ECs in fungal species investigated to date.

Like the MAX family, some members of the ToxA-like family (Avr2/Six3 and AvrL567) are translocated into plant cells, where they perform their virulence functions and are recognized by their corresponding R proteins [58, 100, 101]. For example, Avr2/Six3 of *F. oxysporum* functions together with another effector, Six5, to facilitate the cell-to-cell movement of effectors by increasing the size exclusion limit of plasmodesmata [102]. Other members, however, are thought to be apoplastic. For instance, an ortholog of ToxA from the wheat blotch fungus *Parasitagonospora nodorum* interacts with an integral membrane protein of wheat from the apoplast to facilitate a cell death reaction that involves the intracellular protein, Tsn1 [103]. As all ToxA-like effectors functionally characterized to date display different virulence functions, no insights into the possible function of ToxA-like proteins from *V. inaequalis* can be made. However, as the Rvi6 R protein of apple, which recognizes the AvrRvi6 Avr effector protein of *V. inaequalis*, is an RLP [104], it is possible that these or other ECs of this fungus identified in our study could be recognized as Avr determinants in the subcuticular environment.

Two other structural EC families in *V. inaequalis* were the LARS-like family [63] and the two-domain FOLD-like family [64]. Of note, only one small EC family of *V. inaequalis* was predicted to adopt the FOLD-like fold. However, it must be pointed out that proteins with this fold are known to be difficult to predict [91] and, as a consequence, some members may have been missed. Interestingly, while the precise functions of the LARS and FOLD effector families are currently unknown, both contain members that suppress the host immune response related to the recognition of another member. More specifically, in terms of the LARS family, AvrLm4-7 suppresses immune responses triggered by AvrLm3 and AvrLm5-9 [105, 106], while for the FOLD family, Avr1/Six4 suppresses immune responses triggered by Avr3/Six1 [107]. Together, this suggests that these protein structures may play a role in molecular mimicry to prevent detection [63, 64, 108]. It will be interesting to determine whether similar relationships are observed for the LARS-like and FOLD-like ECs of *V. inaequalis*.

Aside from the folds described above, an intriguing observation was that many of the EC families and singletons from *V. inaequalis* were predicted to adopt a KP6-like or knottin-like fold. The KP6 fold was first described in the antifungal KP6 protein from the P6 virus from *U. maydis* [65, 66] and, in our study, was predicted to be adopted by the AvrLm6-like family as well as AvrLm6 Avr effector protein from *L. maculans*. Notably, this fold is known to be adopted by the EC Zt-KP6-1 (6QPK) from *Z. tritici* [67] and has also been predicted to be adopted by many effectors such as the Avr effector AvrLm10 from

L. maculans [91, 109], the BAS4 effector from *M. oryzae* [91, 110], the Ecp28 EC family and Ecp29 EC from *F. fulva* [44], and the CbNip1 necrosis-inducing effector from the sugar beet leaf spot fungus *Cercospora beticola* [111]. Even more interesting, as observed in *V. inaequalis*, the KP6-like fold has been predicted to be the most abundant fold in the *M. oryzae* secretome [90, 91] and is widely shared among phytopathogens [91]. Altogether, this suggests that this KP6-like fold is a widely conserved structural family of effectors in different phytopathogens.

In terms of ECs from *V. inaequalis* that had a predicted knottin-like fold, several families and a singleton were identified. Knottins are small, ultra-stable proteins with at least three disulfide bridges that form an intramolecular knot known to provide stability in hostile conditions such as the plant apoplast [112]. Current evidence suggests that multiple fungal effectors adopt this fold. Indeed, the Avr effector Avr9 from *F. fulva* has previously been suggested to adopt a knottin fold, based on NMR [113] and cysteine bond connectivity [114] data. Furthermore, an EC from the poplar rust fungus *Melampsora larici-populina*, MLP124266, which is a homolog of the AvrPm4 Avr effector from *M. lini*, was recently shown to adopt this fold [115]. Other knottin-like families identified in our study appear to adopt a fold resembling defensins, similar to the VdAMP3 effector of *V. dahliae*, which has antifungal activity and facilitates microsclerotia formation [116].

Curiously, the Ecp39-like EC family from *V. inaequalis*, along with the Ecp39 EC from *F. fulva*, were predicted to adopt a crambin fold, which is commonly found in antimicrobial proteins [117], while the Ecp10-like family from *V. inaequalis*, together with the Ecp10-1 Avr candidate from *F. fulva*, were predicted to adopt a fold with similarity to the antimicrobial PAF protein from *P. chrysogenum* [69–71]. The large number of proteins from *V. inaequalis* that are predicted to have structural (KP6-like, knottin-like, Ecp39-like and Ecp10-like) or sequence (Ave1-like) similarity to antimicrobial proteins suggests that a considerable proportion of the ECs from this fungus may be dedicated to antagonistic fungus–microbe interactions to ward off microbial competitors.

It should be noted that as part of our EC prediction pipeline, we also identified four putative dikaritin RiPP precursor peptides that were encoded by genes displaying a peak level of expression during late infection. Dikaritins are a class of cyclic bioactive peptides that have recently been discovered in fungi [118] and have been hypothesized to play a role as effectors in promoting host colonization [6]. In line with this, Victorin, a host-selective dikaritin toxin from the necrotrophic oat blight fungus *Cochliobolus victoriae*, has been shown to be essential for pathogenicity in oat cultivars resistant to

the biotrophic crown rust fungus *Puccinia coronata* [46]. The late-expression profile of the *V. inaequalis* dikaritins, together with the finding that many RiPPs from plants have potent antimicrobial activity [119], may suggest that these peptides, perhaps in addition to some of the ECs described above, promote host colonization through the eradication of microbial competitors in preparation for saprobic growth inside fallen leaf litter.

Taken together, our study on *V. inaequalis*, along with previous studies on effector proteins from *M. oryzae*, *F. oxysporum* and other fungi [52, 63, 64, 90, 91], reinforces the idea that many sequence-diverse fungal effectors share common structural folds. This provides weight to the hypothesis that many fungal effector proteins have in fact originated from ancestral folds and suggests that the genes encoding these effectors have evolved through duplication, followed by sequence diversification, to encode sequence-unrelated but structurally similar proteins. Under this hypothesis, the effector proteins have evolved rapidly to a point where almost all sequence similarity, with the exception of residues involved in the maintenance of the overall structural fold, has been lost [52]. It is of course possible, however, that the appearance of common folds, at least in some instances, could be the result of convergent evolution, whereby certain similar folds have evolved independently in different fungi [52].

Specific protein folds may be common across fungal effector proteins as they provide a stable structural scaffold on which surface or loop features can be altered to enable functional diversification [108]. For those Avr effectors that are directly recognized by their corresponding R proteins, it may also be possible that these alterations extend to the evasion of host recognition. Another possibility is that particular structural folds are well suited to certain functions or to interactions with specific host components [108]. Most likely, though, both abovementioned scenarios are possible [108]. Future research focussing on the finer details of the distribution of structural effector families among both pathogenic and non-pathogenic fungi, and on the functional characterization of members within these families, will shed more light on this intriguing topic.

With regard to the ECs of *V. inaequalis* identified in our study, research is now needed to investigate whether these proteins play a role in virulence or avirulence. In many cases, ECs from plant-pathogenic fungi are screened for roles in virulence or avirulence in host plants through protein infiltrations or *Agrobacterium tumefaciens*-mediated transient expression assays (ATTAs). However, such experiments cannot be readily performed in apple. This is because protein solutions and *A. tumefaciens* cells cannot be easily infiltrated into apple leaves without causing mechanical damage. Furthermore, no

reliable method for ATTAs in apple has yet been developed. Often, in these situations, researchers instead opt to perform protein infiltrations or ATTAs in model non-host plants, such as *Nicotiana benthamiana* and *Nicotiana tabacum* [120, 121]. However, experiments involving non-host plants do not always provide useful information about the ECs in question. In support of this, we have screened close to 130 ECs of *V. inaequalis* (the majority of which formed part of the expanded protein families described in our study) for their ability to trigger or suppress cell death in *N. benthamiana* and *N. tabacum* using ATTAs. Using this approach, none of the tested ECs were able to trigger or suppress cell death (S. de la Rosa, unpublished results). One possible reason for this result is that many of the components targeted for modulation in apple by ECs of *V. inaequalis* are absent in *Nicotiana* species. This would not be surprising, as apple is from the Rosaceae family, while *Nicotiana* is from the Solanaceae family. Given this result, we believe that CRISPR-Cas9 technology, which has recently been applied to *V. inaequalis* [122], will provide an important way forward for the functional characterization of ECs and, in particular, those that form part of expanded families.

Conclusions

In conclusion, we have performed the first comprehensive gene expression analysis of a subcuticular pathogen, with a specific focus on genes encoding non-enzymatic proteinaceous ECs, during host colonization. In doing so, our study provides valuable new insights into the molecular mechanisms underpinning subcuticular host colonization by this largely understudied class of fungi, including *V. inaequalis*. Notably, in conjunction with structural modelling, we have also provided an enriched list of ECs from which effectors and Avr effectors of *V. inaequalis* can be identified and functionally characterized. Such a resource is desperately needed as, to date, there have been no publications reporting the cloning of Avr effector genes from this fungus. Once identified, these Avr effector genes will enable the real-time detection of resistance-breaking strains in the orchard. Should Avr effectors of *V. inaequalis* belong to expanded protein families, it may then be possible to engineer their cognate R proteins to recognize features common to the structural fold (direct recognition) or to monitor specific host components targeted by multiple members of the Avr family (indirect recognition). Certainly, with the recent development of CRISPR-Cas9 technology in *V. inaequalis* [122], the functional characterization of ECs, and in particular those that form part of expanded families, is now possible. Finally, our study has provided further evidence that many sequence-diverse fungal effectors share common structural folds. Given that the genomes of

many other *Venturia* species have now been sequenced [19, 22, 123–129], it will be interesting to determine whether specific effector folds are associated with subcuticular growth or the infection of specific host species.

Methods

V. inaequalis isolates

V. inaequalis isolate MNH120, also known as ICMP 13258 and Vi1 [19, 130], was used for all bright-field microscopy, transcriptome sequencing, gene predictions and AlphaFold2 protein structure predictions in this study. MNH120 is a race [1] isolate, meaning that it can only overcome resistance mediated by the *Rvi1* gene in apple [19]. This is presumably due to a mutated, absent, or non-expressed copy of the corresponding *AvrRvi1* gene (a functional copy of which is rare among isolates of *V. inaequalis* [131]). MNH120 is, however, anticipated to possess a functional copy of all other known *V. inaequalis* Avr effector genes (*AvrRvi2*–20), corresponding to all other known apple *R* genes (*Rvi2*–20) [19]. Gene predictions from *V. inaequalis* isolate 05/172 [21], which is of unknown race status [132], were used to assist the gene predictions in isolate MNH120.

Growth in culture for RNA sequencing

V. inaequalis isolate MNH120 was grown as a lawn culture from conidia on cellophane membranes (Waugh Rubber Bands, Wellington, New Zealand) [133] overlaying PDA (DifcoTM, NJ, USA) at 20°C for 7 days under white fluorescent lights (4,300 K) with a 16-h light/8 h dark photoperiod. Four culture plates, representing four independent biological replicates, were then flooded with 1 mL sterile distilled water and the fungal biomass was scraped from the membrane surface using a cell spreader. Following this step, fungal suspensions were transferred to independent microcentrifuge tubes, pelleted by centrifugation at 21,000×g for 1 min, snap frozen in liquid nitrogen and then ground to a powder in preparation for RNA extraction. The 7 dpi time point was chosen as *V. inaequalis* had produced enough biomass on the surface of cellophane membranes for adequate RNA extraction.

Plant infection assays for RNA sequencing and microscopy

Seeds from open-pollinated *M. × domestica* cultivar ‘Royal Gala’ (Hawke’s Bay, New Zealand), which is a cultivar susceptible to scab disease caused by *V. inaequalis*, were germinated at 4°C in moist vermiculite with 100 mg/ml Thiram fungicide (Kiwicare Corporation Limited; Christchurch, New Zealand) for approximately 2 months in the dark. Germinated seedlings were planted in potting mix (DaltonsTM premium potting mix; Daltons, Matamata, New Zealand) and grown under a 16 h light/8 h dark cycle with a Philips SON-T AGRO 400

Sodium lamp, at 20°C with ambient humidity. Inoculations with *V. inaequalis* isolate MNH120 were performed on freshly un-furled detached leaves from 4- to 6-week-old apple seedlings, as described previously [134], with the exception that 5 µl droplets of conidial suspension ($1 \times 10^5 \text{ ml}^{-1}$) were used to cover the entire leaf surface. At 12 and 24 hpi, as well as 2, 3, 5 and 7 dpi, four infected leaves, each from an independent seedling, were sampled to give four biological replicates. A microscopic evaluation of infection was then performed on harvested tips from these leaves. Here, leaf tips were cleared and stained according to a previously established protocol [135] and then visualized by bright-field microscopy, with images captured using a Leica DFC 295 digital camera and the Leica Application Suite X (LAS X). Immediately following tip harvesting, leaves were snap frozen in liquid nitrogen and then ground to a powder in preparation for RNA extraction.

RNA extraction and sequencing

Total RNA was extracted from samples of *V. inaequalis* isolate MNH120 grown in culture, as well as infected leaves, using a Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA), with DNA subsequently removed using DNase I (Invitrogen™, Thermo Fisher Scientific, MA, USA). RNA concentration and purity were quantified using a Nanodrop ND-1000 Spectrophotometer (NanoDrop Technologies, Rockland, DE, USA), while RNA integrity was assessed on the Agilent 2100 Bioanalyser (Agilent Technologies, Waldbronn, Germany) using an Agilent RNA 6000 Nano Kit in conjunction with Agilent 2100 Bioanalyzer software. Genomic DNA contamination was excluded by visualization of RNA on a 0.8% agarose gel and absence of polymerase chain reaction (PCR) amplification products specific to the *actin* gene of apple (GenBank Accession OU745002.1 [location 20,713,063–20,714,807]; primers RE45 [5′–TGACCGAATGAGCAAGGAAATTACT–3′] and RE64 [5′–TACTCAGCTTTGGCAATCCACATC–3′]) [11, 136]. Following these quality control checks, total RNA from each of the samples was sequenced on a HiSeq X platform at Novogene (Beijing, China) via the Massey Genome Service facility (Palmerston North, New Zealand; project number MGS00286). Here, only those RNA samples with an RNA Integrity Number (RIN) value of ≥ 3.5 were sequenced.

Gene prediction

The genome sequence of *V. inaequalis* isolate MNH120 [19] was downloaded from the Joint Genome Institute (JGI) MycoCosm portal (<https://mycocosm.jgi.doe.gov/Venin1/Venin1.home.html>) and a new gene catalogue predicted to accommodate those genes that may have

been missed in the initial annotation by Deng et al. [19] (summarized in Additional file 2: Fig. S1). Here, we combined three sources of information to generate a more complete gene catalogue. More specifically, for the first source of information, the complete set of predicted genes (coding sequences; CDSs) for *V. inaequalis* isolate 05/172 was downloaded from the National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/nuccore/QFBF00000000.1/>) and mapped to the MNH120 genome using GMAP v2021-02-22 [137] to generate a genome annotation file. Isolate 05/172 was used as the source of information for homology-based gene prediction, as it was deemed to have a more complete set of predicted EC genes than the previously predicted gene catalogue for isolate MNH120 [19], based on a crude assessment of how many EC family members were present in the non-redundant (nr) protein database using a tBLASTn search. For the second source of information, we used the RNA-seq data generated from isolate MNH120 in this study to predict open reading frames (ORFs), and thus CDSs, from transcript sequences. More specifically, high-quality RNA-seq reads (see the RNA-seq read analysis section below) from one biological replicate of each time point of *V. inaequalis* grown in planta and in culture were mapped to the MNH120 genome using HISAT2 v2.2.1 [138, 139], with unmapped reads filtered out using SAMtools v9.2.0 [140]. Then, a genome-guided de novo transcriptome assembly was performed with the mapped reads using Trinity v2.12.0 [141]. Likely CDSs were identified using Transdecoder v5.5.0 (<https://github.com/TransDecoder/TransDecoder>), a programme that predicts genes from transcript sequences, in conjunction with a minimum ORF length of 50 amino acids. To support gene predictions, a Pfam domain search was performed to ensure that translated ORFs with one or more characterized functional domains were retained. Additionally, to capture as many EC-encoding ORFs as possible, translated ORFs were scanned against a list of EC proteins identified in the previous gene annotation by Deng et al. [19], supplemented with an in-house database of putative MNH120 effector proteins (de la Rosa and Mesarich, unpublished). Here, a BLASTp *e*-value threshold of ≤ 0.05 was employed, with proteins identified using this analysis retained as likely CDSs. Finally, for the third source of information, the original gene catalogue predicted for isolate MNH120 by Deng et al. [19] was downloaded from the JGI MycoCosm portal as above. Together, the three sets of CDSs, derived from the three sources of information described above, were then loaded onto the MNH120 genome as different tracks in Geneious v9.05 [142], and manual curation was performed to create an updated gene catalogue based on a consensus prediction for each gene. More specifically,

predictions that were identical across at least two of the three sets of CDSs for a given gene were accepted as correct. Likewise, if for a given gene predictions were different across all CDSs, only the CDS supported by mapped RNA-seq reads (intron–exon boundaries) was selected. Finally, when a CDS was only predicted with one method, it was accepted as correct.

Prediction of protein functions

Protein functional domains were predicted using InterProScan v5.51-85.0 in conjunction with the Pfam, HAMAP, MOBIDB, PIRSE, PROSITE and SUPERFAMILY tools [143], while GO class predictions were carried out using Pannzer2 [144]. N-terminal signal peptides were predicted using SignalP v5.0 [145] and transmembrane (TM) domains were predicted using TMHMM v2.0 [146]. CAZymes were predicted using dbCAN2 in conjunction with the Hotpep, HMMER and DIAMOND tools [147]. Only CAZymes predicted with at least two of the three tools were retained for further analysis. Putative PCWDEs were manually identified based on their CAZy classification (<http://www.cazy.org>) [148], KEGG description and InterPro annotation. *RiPP* gene clusters were manually identified in the *V. inaequalis* MNH120 genome (Additional file 8: S1 Text) through the presence of a gene encoding a protein with a DUF3328 domain in close proximity to a gene encoding a dikaritin precursor peptide with an N-terminal signal peptide, followed by one or more perfect or imperfect tandem sequence repeats of at least 10 amino acid residues in length separated by putative kexin protease cleavage sites.

Prediction of effector candidates and effector candidate families

Small proteins of ≤ 400 amino acid residues in length with a predicted N-terminal signal peptide, but without a predicted TM domain or endoplasmic reticulum (ER) retention motif (HDEL/KDEL), were annotated as ECs. This list of ECs was supplemented with proteins of >400 amino acids in length with a predicted N-terminal signal peptide, but no predicted TM domain or ER retention motif, provided that they were predicted to be an effector using EffectorP v3.0 [38]. ECs were grouped into protein families using spectral clustering SCPS v0.9.8 [149]. The identified protein families were then manually curated by eye, taking into account conservation of the N-terminal signal peptide sequence, cysteine spacing, as well as conserved functional domains identified with InterProScan v5.51-85.0. To further refine the list of ECs, proteins with an enzymatic annotation by InterProScan v5.51-85.0 were discarded. In cases where only one or two ECs from a family were predicted to have an enzymatic domain, the EC was retained. To determine whether the ECs had

sequence similarity to other proteins, a BLASTp analysis using an *e*-value threshold of 0.05 was performed against the nr protein database at NCBI.

RNA-seq read analysis

Methods associated with this section are summarized in Additional file 2: Fig. S1. As a starting point for the analysis of RNA-seq reads, all unique genes of *V. inaequalis* isolate MNH120 were identified, and any additional genes that were identical in sequence, representing paralogs (or false gene duplications generated as a consequence of incorrect genome assembly by Deng et al. [19]), were masked to avoid the multimapping of RNA-seq reads. Here, a gff3 file containing only the duplicated genes was generated using AGAT [150], with the sequences subsequently masked using BEDTools v2.30.0 (maskfasta) [151]. Finally, a mappability mask was applied to the MNH120 genome to prevent the multimapping of reads to repetitive genomic regions, including those genes of *EC* families that are known to be highly similar in sequence. To generate the mappability mask, the MNH120 genome was fragmented into all possible 150-bp stretches and mapped to the MNH120 genome using the Burrows-Wheeler Aligner (bwa) [152] with a gap open penalty of 3, gap extension penalty of 3 and max mismatch penalty of 3. Finally, the resulting SAM file from was used to generate the mappability mask (<http://lh3lh3.users.sourceforge.net/download/seqbility-20091110.tar.bz2>). Raw RNA-seq reads were then filtered, in which adapter sequences, as well as reads with $>10\%$ Ns or $\geq 50\%$ low-quality (Qscore: 5) bases, were removed, and the final quality of reads was checked using fastQC v0.11.9 [153]. Next, filtered RNA-seq reads from all samples were mapped to the masked MNH120 genome using HISAT2 v2.2.1 [138, 139] and SAMtools v9.2.0 [140] was used to only keep those reads that mapped to the fungal genome. Uniquely mapped reads were counted using featureCounts from SubRead package v2.0.0 to generate a count matrix [154]. Results from all steps of the RNA-seq analysis were aggregated for quality control assessment using MultiQC v1.11 [155].

Differential gene expression and clustering

The count matrix (see the [RNA-seq read analysis](#) section) was imported to R and a differential gene expression analysis was performed with DESeq2 package v1.32.0 [156]. Pairwise comparisons from all samples were performed and genes with a $\log_2$ -fold change in expression of >1.5 and a *p*adj value of <0.01 during at least one in planta infection time point, relative to growth in culture, were considered significantly differentially expressed. Here, multiple testing correction was applied using the Benjamini-Hochberg (BH) method of the DESeq2 package, in

conjunction with a *p*adj value threshold of <0.01 . A PCA plot was then generated using the PCA function of the DESeq2 package. Genes that were upregulated at one or more in planta infection time points were selected for hierarchical clustering. For hierarchical clustering, RNA-seq read counts were first normalized using the rlog method from the DESeq2 package and scaled. Hierarchical clustering of the genes was performed using the hclust function according to the Ward.D2 and Euclidean distance methods, with the minimum number of clusters displaying a distinct expression profile during host infection identified in the resulting clustering dendrogram. Using quantitative guidance from the cutree function, the number of clusters was initially set to 10, then systematically reduced due to observed similarity between clusters, to give five distinct clusters with unique expression trends. Visualization of gene expression clusters (waves), with the expression trends plotted, was performed using ggplot2 v3.3.5 [157], while gene expression heatmaps were generated using Complexheatmap v2.9.1 [158]. Pearson correlation coefficients were calculated to investigate the co-expression of genes inside specific clusters.

GO and Pfam term enrichment analysis

GO predictions from Pannzer2 [144] with a predictive positive value (PPV) >5 were used for a GO term enrichment analysis across the five distinct gene expression waves. The GO term enrichment analysis was performed with topGO package v2.44.0 [159], using the total set of genes employed for clustering as background and Fisher's exact test for all GO terms: Biological Process (BP), Cellular Component (CC) and Molecular Function (MF). GO enrichment analysis results were visualized using ggplots2 v3.3.5 [157]. An enrichment test for Pfam domains was performed using Fisher's exact test, with all genes targeted in the clustering analysis used as background.

Structural modelling of protein tertiary structures

AlphaFold2 [160], in conjunction with the ColabFold notebook [161], was used to predict the protein tertiary structures of EC family members from *V. inaequalis* isolate MNH120. Here, only those family members that were encoded by genes upregulated in planta, relative to growth in culture, were used in this analysis, with only the most highly expressed member targeted for prediction (i.e. as a family representative). In each case, the mature amino acid sequence of the EC (i.e. without its predicted N-terminal signal peptide) was used as input. For the Avr1/Six4-like family, the published pro-domain [64] was also removed. For those ECs that had <30 proteins with sequence similarity in the NCBI database, as identified by a BLASTp analysis in conjunction with an

e-value threshold of ≤ 0.05 , a custom MSA was generated and used as input. These custom MSAs, which were built with up to 100 protein sequences, depending on how many protein sequences were available, included all mature EC family members that were unique to the updated MNH120 annotation (i.e. that could not be accessed by the AlphaFold2 search algorithms), as well as similar protein sequences identified through the BLASTp analysis at NCBI. To build the custom MSAs, all sequences were aligned using Clustal Omega [162, 163], with the alignments subsequently converted to the a3m format using ToolSeq [164, 165]. The only exception was for members of the FOLD-like family. Here, in an attempt to improve the structural prediction, the *F. oxysporum* Avr1/Six4 and Avr2/Six3 proteins were manually added to the input sequences to generate a custom MSA, even though these proteins were not identified in the initial BLASTp similarity search. For EC singletons, protein tertiary structures were predicted using AlphaFold2 open source code v2.0.1 and v2.1.0 [160], with preset casp14, max_template_date: 2020-05-14, using mature protein sequences as input. Again, only those ECs that were encoded by genes upregulated in planta, relative to growth in culture, were used in this analysis. All predicted protein tertiary structures with a pLDDT score of ≥ 70 were considered confident predictions. Protein structures with an pLDDT score of 50–60 that also had an intrinsically disordered region predicted with MobiDB-lite [166] or PrDos [167] were also considered confident predictions.

Predicted EC protein tertiary structures were screened against the RCSB PDB database to identify proteins with similar folds using the Dali server [49]. Here, all hits with a *Z*-score of ≥ 2 were considered similar. Protein tertiary structures were visualized and aligned using PyMol v2.5, in conjunction with the alignment plugin tool CEalign [168]. To further investigate similarities between protein tertiary structures, TM-align [169] was used to calculate a root-mean-square deviation (RMSD) value. Finally, the general fold of confidently predicted protein tertiary structures was investigated using RUPPEE [170, 171] against the SCOPe v2.08 database [68, 172]. Proteins predicted to have a knottin fold in the SCOPe database were assessed using Knotter 3D to determine whether they had a true knottin structure [112].

Abbreviations

Avr: Avirulence; BH: Benjamini-Hochberg; BP: Biological process; bwa: Burrows-Wheeler aligner; CAZyme: Carbohydrate-active enzyme; CC: Cellular component; CDS: Coding sequence; CE: Carbohydrate esterase; CFEM: Common fold in several fungal extracellular membrane proteins; CM: Cellophane membrane; C terminus: Carboxyl terminus; DAMP: Damaged-associated molecular pattern; dpi: Days post-inoculation; EC: Effector candidate; ER: Endoplasmic reticulum; EWCA: Effectors with chitinase activity; FOLD: *Fusarium*

oxysporum f. sp. *lycopersici* dual-domain; GEO: Gene Expression Omnibus; GH: Glycoside hydrolase; GO: Gene ontology; HMA: Heavy metal-associated; hpi: Hours post-inoculation; HsbA: Hydrophobic surface-binding protein A; IP: Invasion pattern; JGI: Joint Genome Institute; LAS X: Leica Application Suite X; LARS: *Leptosphaeria* avirulence and suppressing; MAMP: Microbe-associated molecular pattern; MAX: *Magnaporthe* Avr and ToxB-like; MF: Molecular function; MSA: Multiple sequence alignment; N terminus: Amino terminus; NCBI: National Center for Biotechnology Information; NeSI: New Zealand eScience Infrastructure; NLR: Nucleotide-binding leucine-rich repeat; nr: Non-redundant; ORF: Open reading frame; PCA: Principal component analysis; PCR: Polymerase chain reaction; PCWDE: Plant cell wall-degrading enzyme; PDA: Potato dextrose agar; PDB: Protein Data Bank; PPV: Predictive positive value; PRR: Pattern recognition receptor; QTL: Quantitative trait loci; R: Resistance; RCSB: Research Collaboratory for Structural Bioinformatics; RIN: RNA integrity number; RIPP: Ribosomally synthesized and post-translationally modified peptide; RLK: Receptor-like kinase; RLP: Receptor-like protein; RNA-seq: RNA sequencing; RMSD: Root-mean-square deviation; TM: Transmembrane.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-022-01442-9>.

Additional file 1: Table S1. RNA-seq transcriptome sequencing read statistics from this study. Samples used for the RNA-seq transcriptome sequencing experiment were derived from an infection time course of *Venturia inaequalis* on detached leaves from susceptible apple cultivar 'Royal Gala' at 12 and 24 hours post-inoculation (hpi), as well as 2, 3, 5 and 7 days post-inoculation (dpi), and during growth. **Table S2.** Effector candidate (EC) protein families or singletons from *Venturia inaequalis* that have sequence similarity to EC or avirulence (Avr) effector proteins from other plant-pathogenic fungi. **Table S3.** Effector candidates (ECs) of *Venturia inaequalis* (Vi) with predicted structural similarity to EC or avirulence (Avr) effector proteins from other plant-pathogenic fungi that have a characterized tertiary structure present in the RCSB PDB.

Additional file 2: Fig. S1. Bioinformatic pipeline used for transcriptome analysis and genome annotation. Total RNA was extracted from apple leaves infected with *Venturia inaequalis* at 12 and 24 hours post-inoculation (hpi), as well as 2, 3, 5 and 7 days post-inoculation (dpi). As a reference for growth in culture, total RNA was also extracted from *V. inaequalis* grown on the surface of cellophane membranes overlaying potato dextrose agar at 7 dpi. Four biological replicates were included per sample. PE: paired-end; CDS: coding sequence; CAzyme: carbohydrate-active enzyme; EC: effector candidate. **Fig. S2.** Plant cell wall-degrading enzyme (PCWDE)-encoding genes of *Venturia inaequalis* upregulated during infection of susceptible apple cultivar 'Royal Gala', relative to growth of the fungus in culture on the surface of cellophane membranes overlaying potato dextrose agar. **A.** Proportion of in planta upregulated PCWDE-encoding genes in each host infection-specific temporal expression wave. **B.** Heatmap of PCWDE-encoding genes upregulated in planta that demonstrate a peak level of expression during waves 1 and 2 of the early infection stage at 12 and 24 hours post-inoculation (hpi). **C.** Heatmap of PCWDE-encoding genes upregulated in planta that demonstrate a peak level of expression during wave 3 of the mid infection stage at 2 and 3 days post-inoculation (dpi) and waves 4 and 5 of the mid-late infection stage at 5 and 7 dpi. Block labels on the left indicate gene expression wave. Numbers in brackets indicate number of genes per wave. Gene expression data are scaled rlog-normalized counts across all samples (Z-score), averaged from four biological replicates. Labels on the right indicate carbohydrate-active enzyme (CAzyme) classification. Bar plots depict the maximum log₂ DESeq2-normalized count value across all in planta time points. AA: auxiliary activity; GH: glycoside hydrolase; CE: carbohydrate esterase; PL: polysaccharide lyase; CBM: carbohydrate-binding module. **Fig. S3.** Prediction of effectors from *Venturia inaequalis*. **A.** Pipeline for the identification of effector candidates (ECs) from *V. inaequalis*. **B.** Comparison of the number of ECs predicted from *V. inaequalis* isolate MNH120 in this study and a previous study [19]. The comparison is based on an exact protein sequence match. The previous study by Deng et al. [19] defined ECs as small proteins of <500 amino acid

residues in length with a signal peptide and no similarity to lytic enzymes.

Fig. S4. Gene families encoding proteinaceous effector candidates (ECs) of *Venturia inaequalis* that have members demonstrating different expression profiles during early and mid-late colonization of susceptible apple cultivar 'Royal Gala'. Expression data during host colonization are DESeq2-normalized counts, averaged from four biological replicates, with error bars representing standard deviation (hpi: hours post-inoculation; dpi: days post-inoculation). *HsbA*: Hydrophobic surface-binding protein A; *CFEM*: common fold in several fungal extracellular membrane proteins; *Cin1*: Cellophane-induced 1. **Fig. S5.** Expression of ribosomally synthesized and post-translationally modified peptide (RiPP) *dikaritin* gene clusters from *Venturia inaequalis* that are upregulated during colonization of susceptible apple cultivar 'Royal Gala', relative to growth of the fungus in culture on the surface of cellophane membranes overlaying potato dextrose agar. Heatmap gene expression data are scaled rlog-normalized counts across all samples (Z-score), averaged from four biological replicates. hpi: hours post-inoculation; dpi: days post-inoculation. Bar plot annotation depicts the maximum log₂ DESeq2-normalized count value across all in planta time points. Genes marked as ϵ putatively encode a protein with a DUF3382 domain or were annotated as a major-facilitator superfamily protein. Genes marked with * putatively encode a dikaritin precursor peptide. The gene marked with N encodes a protein with no characterized functional domain. **Fig. S6.** Effector candidates (ECs) from *Venturia inaequalis* with structural similarity to known avirulence (Avr) effector proteins from other plant-pathogenic fungi not shown in Fig. 4. Representative *V. inaequalis* protein structures (green; family 28 (g13172), family 38 (g9034), singleton g4288) aligned to ToxA from *Pyrenophora tritici-repentis* (1ZLE) (purple), LARS-like protein structure (green; family 47 (g24490)) aligned to AvrLm4-7 from *Leptosphaeria maculans* (7FPR). Protein tertiary structures predicted by AlphaFold2 are the most highly expressed member from each EC family. Disulfide bonds coloured in yellow. N: amino (N) terminus; C: carboxyl (C) terminus. pLDDT: predicted Local Distance Difference Test score (0–100); a pLDDT score of 70–100 is indicative of medium to high confidence. A Dali Z-score above 2 indicates 'significant similarities' between structures. RMSD: root-mean-square deviation. Gene expression data are from upregulated ECs during host colonization and are based on DESeq2-normalized counts, averaged from four biological replicates, with error bars representing standard deviation (hpi: hours post-inoculation; dpi: days post-inoculation). **Fig. S7.** Effector candidates (ECs) from *Venturia inaequalis* with structural similarity to killer protein 6 (KP6) from *Ustilago maydis* P6 virus not shown in Fig. 4. Representative *V. inaequalis* protein structures (green; family 2 (g11711), family 5 (g18375), family 23 (g4577), family 26 (g12079), singleton g4356, singleton g18322) aligned to EC Zt-KP6-1 from *Zymoseptoria tritici* (6QPK). Protein structures predicted by AlphaFold2 are the most highly expressed member from each EC family. Disulfide bonds coloured in yellow. N: amino (N) terminus; C: carboxyl (C) terminus. pLDDT: predicted Local Distance Difference Test score (0–100); a pLDDT score of 70–100 is indicative of medium to high confidence. A Dali Z-score above 2 indicates 'significant similarities' between structures. RMSD: root-mean-square deviation. Gene expression data are from upregulated ECs during host colonization and are based on DESeq2-normalized counts, averaged from four biological replicates, with error bars representing standard deviation (hpi: hours post-inoculation; dpi: days post-inoculation). **Fig. S8.** Effector candidates (ECs) from *Venturia inaequalis* with a predicted knottin-like fold. Representative *V. inaequalis* structures (green; family 11 (g8686), family 12 (g10808), family 24 (g22936), family 35 (g22623), singleton g8946). Protein structures predicted by AlphaFold2 are the most highly expressed member of each EC family. Disulfide bonds coloured in yellow. N: amino (N) terminus; C: carboxyl (C) terminus. pLDDT: predicted Local Distance Difference Test score (0–100); a pLDDT score of 70–100 is indicative of medium to high confidence. A Dali Z-score above 2 indicates 'significant similarities' between structures. Gene expression data are from upregulated ECs during host colonization are based on DESeq2-normalized counts, averaged from four biological replicates, with error bars representing standard deviation (hpi: hours post-inoculation; dpi: days post-inoculation). **Fig. S9.** Effector candidate (EC) families from *Venturia inaequalis* (Vi) with sequence similarity to effectors and characterized or candidate avirulence (Avr) effector proteins from other plant-pathogenic fungi. **A.** Expression data of EC genes during host colonization are

DESeq2-normalized counts, averaged from four biological replicates, with error bars representing standard deviation (hpi: hours post-inoculation; dpi: days post-inoculation). **B.** Protein tertiary structures of candidate virulence and avirulence (Avr) effector proteins predicted by AlphaFold2. Disulfide bonds coloured in yellow. *F. fulva* is *Fulvia fulva* and *V. dahliae* is *Verticillium dahliae*. Green tertiary structures represent the *V. inaequalis* protein; purple structures represent the EC/Avr from the other fungal pathogen; cyan, represents closest analogous structure in the RCSB PDB database. Sequence similarity was identified by reciprocal protein searches based on BLASTp (E-value <0.05). pLDDT: predicted Local Distance Difference Test score (0–100). A pLDDT score of 70–100 is indicative of medium to high confidence. A Dali Z-score above 2 indicates ‘significant similarities’ between proteins. RMSD: root-mean-square deviation. **Fig. S10.** Comparison of protein tertiary structures predicted for two effector candidates (ECs) of *Venturia inaequalis* using different AlphaFold2 methods. EC protein tertiary structures are coloured by amino acid pLDDT (predicted Local Distance Difference Test) score, with a high pLDDT score coloured in blue and a low pLDDT score coloured in red. ColabFold multiple sequence alignments (MSAs) were generated with MMseqs2 [175] and MSAs generated by AlphaFold2 open source CASP14 were assembled using HHblits and HMMER [176, 177]. Sequence coverage graphs of the MSAs used for the structural predictions were automatically generated by ColabFold [161].

Additional file 3: Differentially expressed genes of *Venturia inaequalis* during infection of susceptible apple cultivar ‘Royal Gala’, when compared to growth of the fungus in culture on the surface of cellophane membranes overlying potato dextrose agar.

Additional file 4: List of Pfam domains in proteins encoded by genes of the five distinct temporal expression waves of *Venturia inaequalis* that are upregulated during colonization of susceptible apple cultivar ‘Royal Gala’, when compared to growth of the fungus in culture on the surface of cellophane membranes overlying potato dextrose agar.

Additional file 5: List of effector candidates (ECs) identified from *Venturia inaequalis*, their family classification, and indication of whether EC family members cluster in the genome.

Additional file 6: List of protein tertiary structures from *Venturia inaequalis* with a confident AlphaFold2 prediction and list of predicted SCOPe folds.

Additional file 7: Matrix amino acid identity of the effector candidate (EC) structural families (MAX, ToxA, LARS, FOLD).

Additional file 8: S1 Text. Supplementary methods for the identification of dikaritin RiPP gene clusters in *Venturia inaequalis*.

Acknowledgements

We acknowledge use of the New Zealand eScience Infrastructure (NeSI) high-performance computing facilities, as well as their technical support and training services. In particular, we thank Dinindu Senanayake for his consulting support on the high-throughput prediction of protein tertiary structures. New Zealand’s national computing facilities are provided by NeSI and are funded jointly by NeSI’s collaborator institutions and through the Ministry of Science & Innovation’s Research Infrastructure programme. URL: <http://www.nesi.org.nz>. We thank Drs Erik Rikkerink and Jay Jayaraman for critically reviewing the manuscript, and Dr Simon Williams for providing the *F. oxysporum* Avr1/Six4 and Avr3/Six1 protein structure files ahead of public release.

Authors’ contributions

CHM, MR, JKB and KMP conceived the project. JKB, CHM and BM conducted plant infections and prepared samples for sequencing. MR, BH, BM and SdIR performed the bioinformatic analyses. CHM, JKB, MPC and REB provided critical input in experimental design and data analysis. MR, CHM, BH and JKB wrote the manuscript. All authors read, revised and approved the final manuscript.

Authors’ information

Twitter handles: @mercerocafort (Mercedes Rocafort); @C_H_Mesarich (Carl H. Mesarich)

Funding

MRF and CHM were supported by the Marsden Fund Council from Government funding (project ID 17-MAU-100), managed by Royal Society Te Apārangi. JKB and BM received funding from The New Zealand Institute for Plant and Food Research Limited, Strategic Science Investment Fund, Project number: 12070.

Availability of data and materials

The raw RNA-seq data generated in this study, as well as the count matrix and DESeq2-normalized read counts, have been deposited in the NCBI Gene Expression Omnibus (GEO), and are accessible through GEO Series accession number (GSE198244) [173]. The *V. inaequalis* MNH120 gene annotations and associated protein sequences generated in this study, as well as the output of AlphaFold2 (open source or ColabFold) with the PDB files for the predicted EC tertiary structures are available at zenodo (10.5281/zenodo.6233645) [174].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Laboratory of Molecular Plant Pathology/Bioprotection Aotearoa, School of Agriculture and Environment, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand. ²The New Zealand Institute for Plant and Food Research Limited, Mount Albert Research Centre, Auckland 1025, New Zealand. ³Bioprotection Aotearoa, School of Natural Sciences, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand. ⁴Department of Animal, Plant and Soil Sciences, La Trobe University, AgriBio, Centre for AgriBiosciences, La Trobe University, Bundoora, Victoria 3086, Australia.

Received: 29 March 2022 Accepted: 17 October 2022

Published online: 03 November 2022

References

1. Ristaino JB, Anderson PK, Bebber DP, Brauman KA, Cunniffe NJ, Fedoroff NV, et al. The persistent threat of emerging plant disease pandemics to global food security. *Proc Natl Acad Sci U S A*. 2021;118(23):e2022239118.
2. Cook DE, Mesarich CH, Thomma BPHJ. Understanding plant immunity as a surveillance system to detect invasion. *Annu Rev Phytopathol*. 2015;53(1):541–63.
3. Saijo Y, Loo EP, Yasuda S. Pattern recognition receptors and signaling in plant–microbe interactions. *Plant J*. 2018;93(4):592–613.
4. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal effectors and plant susceptibility. *Annu Rev Plant Biol*. 2015;66(1):513–45.
5. Rovenich H, Boshoven JC, Thomma BPHJ. Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. *Curr Opin Plant Biol*. 2014;20:96–103.
6. Rocafort M, Fudal I, Mesarich CH. Apoplastic effector proteins of plant-associated fungi and oomycetes. *Curr Opin Plant Biol*. 2020;56:9–19.
7. Bradley EL, Ökmen B, Doeblemann G, Henrissat B, Bradshaw RE, Mesarich CH. Secreted glycoside hydrolase proteins as effectors and invasion patterns of plant-associated fungi and oomycetes. *Front Plant Sci*. 2022;13:853106.
8. Lolle S, Stevens D, Coaker G. Plant NLR-triggered immunity: from receptor activation to downstream signaling. *Curr Opin Immunol*. 2020;62:99–105.

9. Bowen JK, Mesarich CH, Bus VG, Beresford RM, Plummer KM, Templeton MD. *Venturia inaequalis*: the causal agent of apple scab. *Mol Plant Pathol*. 2011;12(2):105–22.
10. Jha G, Thakur K, Thakur P. The *Venturia* apple pathosystem: pathogenicity mechanisms and plant defense responses. *J Biomed Biotechnol*. 2009;2009:680160.
11. Kucheryava N, Bowen JK, Sutherland PW, Conolly JJ, Mesarich CH, Rikkerink EH, et al. Two novel *Venturia inaequalis* genes induced upon morphogenetic differentiation during infection and *in vitro* growth on cellophane. *Fungal Genet Biol*. 2008;45(10):1329–39.
12. Nusbaum CJ, Keitt GW. A cytological study of host-parasite relations of *Venturia inaequalis* on apple leaves. *J Agric Res*. 1938;56(8):595–618.
13. Shiller J, Van de Wouw AP, Taranto AP, Bowen JK, Dubois D, Robinson A, et al. A large family of *AvrLm6*-like genes in the apple and pear scab pathogens, *Venturia inaequalis* and *Venturia pirina*. *Front Plant Sci*. 2015;6:980.
14. Manktelow D, Beresford R, Batchelor T, Walker J. Use patterns and economics of fungicides for disease control in New Zealand apples, International Conference on Integrated Fruit Production; 1995. p. 187–92.
15. Cox KD. Fungicide resistance in *Venturia inaequalis*, the causal agent of apple scab, in the United States. In: Ishii H, Hollomon DW, editors. Fungicide resistance in plant pathogens: principles and a guide to practical management. Tokyo: Springer Japan; 2015. p. 433–47.
16. Khajuria YP, Kaul S, Wani AA, Dhar MK. Genetics of resistance in apple against *Venturia inaequalis* (Wint.) Cke. *Tree Genet Genomes*. 2018;14(2):16.
17. Patocchi A, Wehrli A, Dubuis PH, Auwerkerken A, Leida C, Cipriani G, et al. Ten years of VINQUEST: first insight for breeding new apple cultivars with durable apple scab resistance. *Plant Dis*. 2020;104(8):2074–81.
18. Papp D, Singh J, Gadoury D, Khan A. New North American isolates of *Venturia inaequalis* can overcome apple scab resistance of *Malus floribunda* 821. *Plant Dis*. 2020;104(3):649–55.
19. Deng CH, Plummer KM, Jones DAB, Mesarich CH, Shiller J, Taranto AP, et al. Comparative analysis of the predicted secretomes of Rosaceae scab pathogens *Venturia inaequalis* and *V. pirina* reveals expanded effector families and putative determinants of host range. *BMC Genom*. 2017;18(1):339.
20. Passey TAJ, Armitage AD, Sobczyk MK, Shaw MW, Xu X. Genomic sequencing indicates non-random mating of *Venturia inaequalis* in a mixed cultivar orchard. *Plant Pathol*. 2020;69(4):669–76.
21. Passey TAJ, Armitage AD, Xu X. Annotated draft genome sequence of the apple scab pathogen *Venturia inaequalis*. *Microbiol Resour Annot*. 2018;7(12):e01062–18.
22. Le Cam B, Sargent D, Gouzy J, Anselem J, Bellanger M-N, Bouchez O, et al. Population genome sequencing of the scab fungal species *Venturia inaequalis*, *Venturia pirina*, *Venturia aucupariae* and *Venturia asperata*. *G3: Genes Genomes Genet*. 2019;9(8):2405–14.
23. Lichtner FJ, Jurick WM, Ayer KM, Gaskins VL, Villani SM, Cox KD. A genome resource for several North American *Venturia inaequalis* isolates with multiple fungicide resistance phenotypes. *Phytopathology*. 2020;110(3):544–6.
24. Fudal I, Ross S, Gout L, Blaise F, Kuhn ML, Eckert MR, et al. Heterochromatin-like regions as ecological niches for avirulence genes in the *Leptosphaeria maculans* genome: map-based cloning of *AvrLm6*. *Mol Plant Microbe Interact*. 2007;20(4):459–70.
25. de Jonge R, Peter van Esse H, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, et al. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc Natl Acad Sci U S A*. 2012;109(13):5110.
26. Snelders NC, Rovenich H, Petti GC, Rocafort M, van den Berg GCM, Vorholt JA, et al. Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. *Nat Plants*. 2020;6(11):1365–74.
27. Caffier V, Le Cam B, Expert P, Tellier M, Devaux M, Giraud M, et al. A new scab-like disease on apple caused by the formerly saprotrophic fungus *Venturia asperata*. *Plant Pathol*. 2012;61(5):915–24.
28. Latham AJR, A. E. Development of *Cladosporium caryigenum* in pecan leaves. *Phytopathology*. 1988;78(8):1104–8.
29. Lanza B, Ragnelli AM, Priore M, Aimola P. Morphological and histochemical investigation of the response of *Olea europaea* leaves to fungal attack by *Spillocaea oleagina*. *Plant Pathol*. 2017;66(8):1239–47.
30. Avrova A, Knogge W. *Rhynchosporium commune*: a persistent threat to barley cultivation. *Mol Plant Pathol*. 2012;13(9):986–97.
31. Jones P, Ayres PG. *Rhynchosporium* leaf blotch of barley studied during the subcuticular phase by electron microscopy. *Physiol Plant Pathol*. 1974;4(2):229–33.
32. Bleichert O, Debener T. Morphological characterization of the interaction between *Diplocarpon rosae* and various rose species. *Plant Pathol*. 2005;54(1):82–90.
33. Zhao H, Han Q, Wang J, Gao X, Xiao C-L, Liu J, et al. Cytology of infection of apple leaves by *Diplocarpon mali*. *Eur J Plant Pathol*. 2013;136(1):41–9.
34. Lanver D, Müller AN, Happel P, Schweizer G, Haas FB, Franitza M, et al. The biotrophic development of *Ustilago maydis* studied by RNA-Seq analysis. *Plant Cell*. 2018;30(2):300.
35. Gervais J, Plissonneau C, Linglin J, Meyer M, Labadie K, Cruaud C, et al. Different waves of effector genes with contrasted genomic location are expressed by *Leptosphaeria maculans* during cotyledon and stem colonization of oilseed rape. *Mol Plant Pathol*. 2017;18(8):1113–26.
36. Bradshaw RE, Guo Y, Sim AD, Kabir MS, Chettri P, Ozturk IK, et al. Genome-wide gene expression dynamics of the fungal pathogen *Dothistroma septosporium* throughout its infection cycle of the gymnosperm host *Pinus radiata*. *Mol Plant Pathol*. 2016;17(2):210–24.
37. Thakur K, Chawla V, Bhatti S, Swarnkar MK, Kaur J, Shankar R, et al. De novo transcriptome sequencing and analysis for *Venturia inaequalis*, the devastating apple scab pathogen. *PLoS One*. 2013;8(1):e53937.
38. Sperschneider J, Dodds PN. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol Plant Microbe Interact*. 2022;35(2):146–56.
39. Xue C, Park G, Choi W, Zheng L, Dean RA, Xu J-R. Two novel fungal virulence genes specifically expressed in appressoria of the rice blast fungus. *Plant Cell*. 2002;14(9):2107–19.
40. Kulkarni RD, Kelkar HS, Dean RA. An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. *Trends Biochem Sci*. 2003;28(3):118–21.
41. Garrido SM, Kitamoto N, Watanabe A, Shintani T, Gomi K. Functional analysis of FarA transcription factor in the regulation of the genes encoding lipolytic enzymes and hydrophobic surface binding protein for the degradation of biodegradable plastics in *Aspergillus oryzae*. *J Biosci Bioeng*. 2012;113(5):549–55.
42. Ohtaki S, Maeda H, Takahashi T, Yamagata Y, Hasegawa F, Gomi K, et al. Novel hydrophobic surface binding protein, HsbA, produced by *Aspergillus oryzae*. *Appl Environ Microbiol*. 2006;72(4):2407–13.
43. Mesarich CH, Schmitz M, Tremouilhac P, McGillivray DJ, Templeton MD, Dinglich AJ. Structure, dynamics and domain organization of the repeat protein Cin1 from the apple scab fungus. *Biochim Biophys Acta-Proteins Proteom*. 2012;1824(10):1118–28.
44. Mesarich CH, Ökmen B, Rovenich H, Griffiths SA, Wang C, Karimi Jashni M, et al. Specific hypersensitive response-associated recognition of new apoplastic effectors from *Cladosporium fulvum* in wild tomato. *Mol Plant Microbe Interact*. 2018;31(1):145–62.
45. Bolton MD, Van Esse HP, Vossen JH, De Jonge R, Stergiopoulos I, Stulemeijer IJE, et al. The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. *Mol Microbiol*. 2008;69(1):119–36.
46. Kessler SC, Zhang X, McDonald MC, Gilchrist CLM, Lin Z, Rightmyer A, et al. Victorin, the host-selective cyclic peptide toxin from the oat pathogen *Cochliobolus victoriae* is ribosomally encoded. *Proc Natl Acad Sci U S A*. 2020;117(39):24243.
47. Kessler SC, Chooi Y-H. Out for a RiPP: challenges and advances in genome mining of ribosomal peptides from fungi. *Nat Prod Rep*. 2022;39(2):222–30.
48. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9.
49. Holm L. Using dali for protein structure comparison. *Methods Mol Biol*. 2020;2112:29–42.
50. Hoh F, Padilla A, De Guillen K. New MAX effector from *Magnaporthe oryzae*; 2019. <https://doi.org/10.2210/pdb6R5J/pdb>.
51. Li W, Wang B, Wu J, Lu G, Hu Y, Zhang X, et al. The *Magnaporthe oryzae* avirulence gene *AvrPiz-t* encodes a predicted secreted protein that

- triggers the immunity in rice mediated by the blast resistance gene *Piz-t*. Mol Plant Microbe Interact. 2009;22(4):411–20.
52. de Guillen K, Ortiz-Vallejo D, Gracy J, Fournier E, Kroj T, Padilla A. Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. PLoS Pathog. 2015;11(10):e1005228.
 53. Ose T, Oikawa A, Nakamura Y, Maenaka K, Higuchi Y, Satoh Y, et al. Solution structure of an avirulence protein, AVR-Pia, from *Magnaporthe oryzae*. J Biomol NMR. 2015;63(2):229–35.
 54. Zhang X, He D, Zhao Y, Cheng X, Zhao W, Taylor IA, et al. A positive-charged patch and stabilized hydrophobic core are essential for avirulence function of AvrPib in the rice blast fungus. Plant J. 2018;96(1):133–46.
 55. De la Concepcion JC, Franceschetti M, Maqbool A, Saitoh H, Terauchi R, Kamoun S, et al. Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. Nat Plants. 2018;4(8):576–85.
 56. Nyarko A, Singarapu KK, Figueroa M, Manning VA, Pandelova I, Wolpert TJ, et al. Solution NMR structures of *Pyrenophora tritici-repentis* ToxB and its inactive homolog reveal potential determinants of toxin activity. J Biol Chem. 2014;289(37):25946–56.
 57. Sarma GN, Manning VA, Ciuffetti LM, Karplus PA. Structure of Ptr ToxA: An RGD-containing host-selective toxin from *Pyrenophora tritici-repentis*. Plant Cell. 2005;17(11):3190–202.
 58. Ma L, Cornelissen BJC, Takken FLW. A nuclear localization for Avr2 from *Fusarium oxysporum* is required to activate the tomato resistance protein I-2. Front Plant Sci. 2013;4:94.
 59. Di X, Cao L, Hughes RK, Tintor N, Banfield MJ, Takken FLW. Structure-function analysis of the *Fusarium oxysporum* Avr2 effector allows uncoupling of its immune-suppressing activity from recognition. New Phytol. 2017;216(3):897–914.
 60. Guncar G, Wang C-IA, Forwood JK, Teh T, Catanzariti A-M, Ellis JG, et al. The use of Co²⁺ for crystallization and structure determination, using a conventional monochromatic X-ray source, of flax rust avirulence protein. Acta Crystallogr F: Struct Biol. 2007;63(Pt 3):209–13.
 61. Wang C-IA, Gunčar G, Forwood JK, Teh T, Catanzariti A-M, Lawrence GJ, et al. Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. Plant Cell. 2007;19(9):2898–912.
 62. Blondeau K, Blaise F, Graille M, Kale SD, Linglin J, Ollivier B, et al. Crystal structure of the effector AvrLm4-7 of *Leptosphaeria maculans* reveals insights into its translocation into plant cells and recognition by resistance proteins. Plant J. 2015;83(4):610–24.
 63. Lazar N, Mesarich CH, Petit-Houdenot Y, Talbi N, Li de la Sierra-Gallay I, Zélie E, et al. A new family of structurally conserved fungal effectors displays epistatic interactions with plant resistance proteins. PLoS Pathog. 2022;18(7):e1010664.
 64. Yu DS, Outram MA, Smith A, McCombe CL, Khambalkar PB, Rima SA, et al. The structural repertoire of *Fusarium oxysporum* f. sp. *lycopersici* effectors revealed by experimental and computational studies. bioRxiv. 2021;2021.12.14.472499. <https://doi.org/10.1101/2021.12.14.472499>.
 65. Li N, Erman M, Pangborn W, Duax WL, Park C-M, Bruenn J, et al. Structure of *Ustilago maydis* killer toxin KP6 α -subunit: a multimeric assembly with a central pore. J Biol Chem. 1999;274(29):20425–31.
 66. Allen A, Chatt E, Smith TJ. The atomic structure of the virally encoded antifungal protein, KP6. J Mol Biol. 2013;425(3):609–21.
 67. Padilla A, Hoh F, De Guillen K. Zt-KP6-1: an effector from *Zymoseptoria tritici*; 2019. <https://doi.org/10.2210/pdb6QPK/pdb>.
 68. Fox NK, Brenner SE, Chandonia J-M. SCOPe: structural classification of proteins-extended, integrating SCOP and ASTRAL data and classification of new structures. Nucleic Acids Res. 2014;42(D1):D304–D9.
 69. Huber A, Hajdu D, Bratschun-Khan D, Gáspári Z, Varbanov M, Philippot S, et al. New antimicrobial potential and structural properties of PAFB: a cationic, cysteine-rich protein from *Penicillium chrysogenum* Q176. Sci Rep. 2018;8(1):1751.
 70. Marx F, Binder U, Leiter E, Pócsi I. The *Penicillium chrysogenum* antifungal protein PAF, a promising tool for the development of new antifungal therapies and fungal cell biology studies. Cell Mol Life Sci. 2008;65(3):445–54.
 71. Sonderegger C, Fizil Á, Burtscher L, Hajdu D, Muñoz A, Gáspári Z, et al. D19S mutation of the cationic, cysteine-rich protein PAF: novel insights into its structural dynamics, thermal unfolding and antifungal function. PLoS One. 2017;12(1):e0169920.
 72. Takahashi T, Maeda H, Yoneda S, Ohtaki S, Yamagata Y, Hasegawa F, et al. The fungal hydrophobin RolA recruits polyesterase and laterally moves on hydrophobic surfaces. Mol Microbiol. 2005;57(6):1780–96.
 73. Koller W, Parker DM, Becker CM. Role of cutinase in the penetration of apple leaves by *Venturia inaequalis*. Phytopathology. 1991;81(11):1375–9.
 74. Martínez-Cruz J, Romero D, Hierrezuelo J, Thon M, de Vicente A, Pérez-García A. Effectors with chitinase activity (EWCA), a family of conserved, secreted fungal chitinases that suppress chitin-triggered immunity. Plant Cell. 2021;33(4):1319–40.
 75. Gamas P, Niebel Fde C, Lescure N, Cullimore J. Use of a subtractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. Mol Plant Microbe Interact. 1996;9(4):233–42.
 76. Kimura M, Yamamoto YY, Seki M, Sakurai T, Sato M, Abe T, et al. Identification of *Arabidopsis* genes regulated by high light-stress using cDNA microarray. Photochem Photobiol. 2003;77(2):226–33.
 77. Doss RP. Treatment of pea pods with Bruchin B results in up-regulation of a gene similar to *MtN19*. Plant Physiol Biochem. 2005;43(3):225–31.
 78. Naya L, Paul S, Valdés-López O, Mendoza-Soto AB, Nova-Franco B, Sosa-Valencia G, et al. Regulation of copper homeostasis and biotic interactions by microRNA 398b in common bean. PLoS One. 2014;9(1):e84416.
 79. Zhang Z-N, Wu Q-Y, Zhang G-Z, Zhu Y-Y, Murphy RW, Liu Z, et al. Systematic analyses reveal uniqueness and origin of the CFEM domain in fungi. Sci Rep. 2015;5(1):13032.
 80. Zhao S, Shang X, Bi W, Yu X, Liu D, Kang Z, et al. Genome-wide identification of effector candidates with conserved motifs from the wheat leaf rust fungus *Puccinia tritici*. Front Microbiol. 2020;11:1188.
 81. Zhu W, Wei W, Wu Y, Zhou Y, Peng F, Zhang S, et al. BcCFEM1, a CFEM domain-containing protein with putative GPI-anchored site, is involved in pathogenicity, conidial production, and stress tolerance in *Botrytis cinerea*. Front Microbiol. 2017;8:1807.
 82. Wang J-x, Long F, Zhu H, Zhang Y, Wu J-y, Shen S, et al. Bioinformatic analysis and functional characterization of CFEM proteins in *Setosphaeria turcica*. J Integr Agric. 2021;20(9):2438–49.
 83. Choi W, Dean RA. The adenylate cyclase gene *MAC1* of *Magnaporthe grisea* controls appressorium formation and other aspects of growth and development. Plant Cell. 1997;9(11):1973–83.
 84. Skrzydeł J, Borowska-Wykret D, Kwiatkowska D. Structure, assembly and function of cuticle from mechanical perspective with special focus on perianth. Int J Mol Sci. 2021;22(8):4160.
 85. Sacristán S, Vigouroux M, Pedersen C, Skamnioti P, Thordal-Christensen H, Micali C, et al. Coevolution between a family of parasite virulence effectors and a class of LINE-1 retrotransposons. PLoS One. 2009;4(10):e7463.
 86. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science. 2010;330(6010):1543–6.
 87. Duthel JY, Mannhaupt G, Schweizer G, MK Sieber C, Münsterkötter M, Güldener U, et al. A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. Genome Biol Evol. 2016;8(3):681–704.
 88. Fouché S, Plissonneau C, Croll D. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr Opin Microbiol. 2018;46:34–42.
 89. Amoozadeh S, Johnston J, Meisrimler C-N. Exploiting structural modeling tools to explore host-translocated effector proteins. Int J Mol Sci. 2021;22(23):12962.
 90. Seong K, Krasileva KV. Computational structural genomics unravels common folds and novel families in the secretome of fungal phytopathogen *Magnaporthe oryzae*. Mol Plant Microbe Interact. 2021;34(11):1267–80.
 91. Seong K, Krasileva KV. Comparative computational structural genomics highlights divergent evolution of fungal effectors. bioRxiv. 2022;2022.05.02.490317. <https://doi.org/10.1101/2022.05.02.490317>.
 92. Cesari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, et al. The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. Plant Cell. 2013;25(4):1463–81.

93. Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, et al. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell*. 2009;21(5):1573–91.
94. Ashikawa I, Hayashi N, Yamane H, Kanamori H, Wu J, Matsumoto T, et al. Two adjacent nucleotide-binding site–leucine-rich repeat class genes are required to confer *Pikm*-specific rice blast resistance. *Genetics*. 2008;180(4):2267.
95. Zhang S, Wang L, Wu W, He L, Yang X, Pan Q. Function and evolution of *Magnaporthe oryzae* avirulence gene *AvrPib* responding to the rice blast resistance gene *Pib*. *Sci Rep*. 2015;5:11642.
96. Oikawa K, Fujisaki K, Shimizu M, Takeda T, Saitoh H, Hirabuchi A, et al. The blast pathogen effector AVR-Pik binds and stabilizes rice heavy metal-associated (HMA) proteins to co-opt their function in immunity. *bioRxiv*. 2020;2020.12.01.406389. <https://doi.org/10.1101/2020.12.01.406389>.
97. Guo L, Cesari S, de Guillen K, Chalvon V, Mammri L, Ma M, et al. Specific recognition of two MAX effectors by integrated HMA domains in plant immune receptors involves distinct binding surfaces. *Proc Natl Acad Sci U S A*. 2018;115(45):11637.
98. Franceschetti M, Maqbool A, Jiménez-Dalmaroni Maximiliano J, Pennington Helen G, Kamoun S, Banfield MJ. Effectors of filamentous plant pathogens: commonalities amid diversity. *Microbiol Mol Biol Rev*. 2017;81(2):e00066–16.
99. Schouten HJ, Brinkhuis J, Burgh A, Schaart JG, Groenwold R, Brogini GAL, et al. Cloning and functional characterization of the *Rvi15* (*Vr2*) gene for apple scab resistance. *Tree Genet Genomes*. 2013;10:251–60.
100. Rafiqi M, Gan PH, Ravensdale M, Lawrence GJ, Ellis JG, Jones DA, et al. Internalization of flax rust avirulence proteins into flax and tobacco cells can occur in the absence of the pathogen. *Plant Cell*. 2010;22(6):2017–32.
101. Manning VA, Hamilton SM, Karplus PA, Ciuffetti LM. The Arg-Gly-Asp-containing, solvent-exposed loop of Ptr ToxA is required for internalization. *Mol Plant Microbe Interact*. 2008;21(3):315–25.
102. Cao L, Blekemolen MC, Tintor N, Cornelissen BJC, Takken FLW. The *Fusarium oxysporum* Avr2-Six5 effector pair alters plasmodesmatal exclusion selectivity to facilitate cell-to-cell movement of Avr2. *Mol Plant*. 2018;11(5):691–705.
103. Dagvadorj B, Outram MA, Williams SJ, Solomon PS. The necrotrophic effector ToxA from *Parastagonospora nodorum* interacts with wheat NHL proteins to facilitate Tsn1-mediated necrosis. *Plant J*. 2022;110(2):407–18.
104. Belfanti E, Silfverberg-Dilworth E, Tartarini S, Patocchi A, Barbieri M, Zhu J, et al. The *HcrVf2* gene from a wild apple confers scab resistance to a transgenic cultivated variety. *Proc Natl Acad Sci U S A*. 2004;101(3):886–90.
105. Ghanbarnia K, Ma L, Larkan NJ, Haddadi P, Fernando WGD, Borhan MH. *Leptosphaeria maculans* AvrLm9: a new player in the game of hide and seek with AvrLm4-7. *Mol Plant Pathol*. 2018;19(7):1754–64.
106. Plissonneau C, Daverdin G, Ollivier B, Blaise F, Degraeve A, Fudal I, et al. A game of hide and seek between avirulence genes *AvrLm4-7* and *AvrLm3* in *Leptosphaeria maculans*. *New Phytol*. 2016;209(4):1613–24.
107. Houterman PM, Cornelissen BJC, Rep M. Suppression of plant resistance gene-based immunity by a fungal effector. *PLoS Pathog*. 2008;4(5):e1000061.
108. Outram MA, Figueroa M, Sperschneider J, Williams SJ, Dodds PN. Seeing is believing: Exploiting advances in structural biology to understand and engineer plant immunity. *Curr Opin Plant Biol*. 2022;67:102210.
109. Petit-Houdonot Y, Degraeve A, Meyer M, Blaise F, Ollivier B, Marais CL, et al. A two genes-for-one gene interaction between *Leptosphaeria maculans* and *Brassica napus*. *New Phytol*. 2019;223(1):397–411.
110. Mosquera G, Giraldo MC, Khang CH, Coughlan S, Valent B. Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice blast disease. *Plant Cell*. 2009;21(4):1273–90.
111. Ebert MK, Rangel LI, Spanner RE, Taliadoros D, Wang X, Friesen TL, et al. Identification and characterization of *Cercospora beticola* necrosis-inducing effector Cbnip1. *Mol Plant Pathol*. 2021;22(3):301–16.
112. Postic G, Gracy J, Périn C, Chiche L, Gelly J-C. KNOTTIN: the database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modeling. *Nucleic Acids Res*. 2018;46(D1):D454–D8.
113. Vervoort J, van den Hooven HW, Berg A, Vossen P, Vogelsang R, Joosten MHAI, et al. The race-specific elicitor AVR9 of the tomato pathogen *Cladosporium fulvum*: a cystine knot protein: sequence-specific ¹H NMR assignments, secondary structure and global fold of the protein. *FEBS Lett*. 1997;404(2):153–8.
114. van den Hooven HW, van den Burg HA, Vossen P, Boeren S, de Wit PJGM, Vervoort J. Disulfide bond structure of the AVR9 elicitor of the fungal tomato pathogen *Cladosporium fulvum*: evidence for a cystine knot. *Biochemistry*. 2001;40(12):3458–66.
115. de Guillen K, Lorrain C, Tsan P, Barthe P, Petre B, Saveleva N, et al. Structural genomics applied to the rust fungus *Melampsora larici-populina* reveals two candidate effector proteins adopting cystine knot and NTF2-like protein folds. *Sci Rep*. 2019;9(1):18084.
116. Snelders NC, Petti GC, van den Berg GCM, Seidl MF, Thomma BPHJ. An ancient antimicrobial protein co-opted by a fungal plant pathogen for in planta mycobiome manipulation. *Proc Natl Acad Sci U S A*. 2021;118(49):e2110968118.
117. Lee H-T, Lee C-C, Yang J-R, Lai JZC, Chang KY. A large-scale structural classification of antimicrobial peptides. *Biomed Res Int*. 2015;2015:475062.
118. Vogt E, Künzler M. Discovery of novel fungal RiPP biosynthetic pathways and their application for the development of peptide therapeutics. *Appl Microbiol Biotechnol*. 2019;103(14):5567–81.
119. Slazak B, Kapusta M, Strömstedt AA, Slomka A, Krychowiak M, Shariatgorji M, et al. How does the sweet violet (*Viola odorata* L.) fight pathogens and pests – cyclotides as a comprehensive plant host defense system. *Front Plant Sci*. 2018;9:1296.
120. Kettles GJ, Bayon C, Canning G, Rudd JJ, Kanyuka K. Apoplastic recognition of multiple candidate effectors from the wheat pathogen *Zymoseptoria tritici* in the nonhost plant *Nicotiana benthamiana*. *New Phytol*. 2017;213(1):338–50.
121. Ma L, Lukasik E, Gawehns F, Takken FLW. The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in *Nicotiana benthamiana* leaves. In: Bolton MD, Thomma BPHJ, editors. *Plant Fungal Pathogens: Methods and Protocols*. Totowa: Humana Press; 2012. p. 61–74.
122. Rocafor M, Arshed S, Hudson D, Sidhu JS, Bowen JK, Plummer KM, et al. CRISPR-Cas9 gene editing and rapid detection of gene-edited mutants using high-resolution melting in the apple scab fungus, *Venturia inaequalis*. *Fungal Biol*. 2021;126(1):35–46.
123. Chen C, Bock CH, Wood BW. Draft genome sequence of *Venturia carpophila*, the causal agent of peach scab. *Stand Genom Sci*. 2017;12:68.
124. Jaber MY, Bao J, Gao X, Zhang L, He D, Wang X, et al. Genome sequence of *Venturia oleaginea*, the causal agent of olive leaf scab. *Mol Plant Microbe Interact*. 2020;33(9):1095–7.
125. Johnson S, Jones D, Thrimawithana AH, Deng CH, Bowen JK, Mesarich CH, et al. Whole genome sequence resource of the Asian pear scab pathogen *Venturia nashicola*. *Mol Plant Microbe Interact*. 2019;32(11):1463–7.
126. Prokchorchik M, Won K, Lee Y, Choi ED, Segonzac C, Sohn KH. High contiguity whole genome sequence and gene annotation resource for two *Venturia nashicola* isolates. *Mol Plant Microbe Interact*. 2019;32(9):1091–4.
127. Winter DJ, Charlton ND, Krom N, Shiller J, Bock CH, Cox MP, et al. Chromosome-level reference genome of *Venturia effusa*, causative agent of pecan scab. *Mol Plant Microbe Interact*. 2020;33(2):149–52.
128. Zhou Y, Zhang L, Fan F, Wang ZQ, Huang Y, Yin LF, et al. Genome sequence of *Venturia carpophila*, the causal agent of peach scab. *Mol Plant Microbe Interact*. 2021;34(7):852–6.
129. Hamelin R. The *Venturia populina* CBS256.38 v1.0 genome 2022 [Available from: <https://mycocosm.jgi.doe.gov/Venpo1/Venpo1.home.html>].
130. Stehmann C, Pennycook S, Plummer KM. Molecular identification of a sexual interloper: the pear pathogen, *Venturia pirina*, has sex on apple. *Phytopathology*. 2001;91(7):633–41.
131. Caffier V, Patocchi A, Expert P, Bellanger M-N, Durel C-E, Hilber-Bodmer M, et al. Virulence characterization of *Venturia inaequalis* reference isolates on the differential set of *Malus* hosts. *Plant Dis*. 2014;99(3):370–5.
132. Xu X, Harvey A, Barbara DJ. Population variation of apple scab (*Venturia inaequalis*) within mixed orchards in the UK. *Eur J Plant Pathol*. 2013;135:97–104.

133. Parker D, Hilber U, Bodmer M, Smith F, Yao C, Köller W. Production and transformation of conidia of *Venturia inaequalis*. *Phytopathology*. 1995;85(1):87–91.
134. Win J, Greenwood DR, Plummer KM. Characterisation of a protein from *Venturia inaequalis* that induces necrosis in *Malus* carrying the *Vm* resistance gene. *Physiol Mol Plant Pathol*. 2003;62(4):193–202.
135. Bruzzese E, Hasan S. A whole leaf clearing and staining technique for host specificity studies of rust fungi. *Plant Pathol*. 1983;32(3):335–8.
136. Atkinson RG, Johnston SL, Yauk Y-K, Sharma NN, Schröder R. Analysis of *xyloglucan endotransglucosylase/hydrolase* (*XTH*) gene families in kiwifruit and apple. *Postharvest Biol Technol*. 2009;51(2):149–57.
137. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*. 2005;21(9):1859–75.
138. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–60.
139. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–15.
140. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008.
141. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52.
142. Kearsley M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28(12):1647–9.
143. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterPro-Scan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236–40.
144. Törönen P, Medlar A, Holm L. PANNZER2: a rapid functional annotation web server. *Nucleic Acids Res*. 2018;46(W1):W84–W88.
145. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019;37(4):420–3.
146. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305(3):567–80.
147. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46(W1):W95–W101.
148. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res*. 2022;50(D1):D571–d7.
149. Nepusz T, Sasidharan R, Paccanaro A. SCPS: a fast implementation of a spectral method for detecting protein families on a genome-wide scale. *BMC Bioinform*. 2010;11(1):120.
150. Jacques Dainat DH, Davis E, Crouch K, Sol L, Pascal-git, Tayyrov. AGAT: another gff analysis toolkit to handle annotations in any gtf/gff format. Zenodo; 2022. <https://doi.org/10.5281/zenodo.3549546>.
151. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
152. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2010;26(5):589–95.
153. Wingett SW, Andrews S. FastQ Screen: a tool for multi-genome mapping and quality control. F1000Research. 2018;7:1338.
154. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
155. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047–8.
156. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
157. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag; 2016.
158. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847–9.
159. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006;22(13):1600–7.
160. Ruff KM, Pappu RV. AlphaFold and implications for intrinsically disordered proteins. *J Mol Biol*. 2021;433(20):167208.
161. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022;19(6):679–82.
162. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7(1):539.
163. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res*. 2010;38:W695–W9.
164. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol*. 2018;430(15):2237–43.
165. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, et al. Protein sequence analysis using the MPI bioinformatics toolkit. *Curr Protoc Bioinformatics*. 2020;72(1):e108.
166. Necci M, Piovesan D, Dosztányi Z, Tosatto SCE. MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics*. 2017;33(9):1402–4.
167. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res*. 2007;35:W460–4.
168. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng*. 1998;11(9):739–47.
169. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 2005;33(7):2302–9.
170. Ayoub R, Lee Y. RUPEE: A fast and accurate purely geometric protein structure search. *PLoS One*. 2019;14(3):e0213712.
171. Ayoub R, Lee Y. Protein structure search to support the development of protein structure prediction methods. *Proteins: Struct Funct Genet*. 2021;89(6):648–58.
172. Chandonia J-M, Fox NK, Brenner SE. SCOPe: classification of large macromolecular structures in the structural classification of proteins—extended database. *Nucleic Acids Res*. 2019;47(D1):D475–D81.
173. Rocafort M, Bowen JK, Hassing B, Cox MP, McGreal B, de la Rosa S, Plummer KM, Bradshaw RE, Mesarich CH. Time series transcriptome of *Venturia inaequalis* (apple scab) infecting *Malus domestica* (apple). GEO 2022. [Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198244>].
174. Rocafort M, Bowen JK, Hassing B, Cox MP, McGreal B, de la Rosa S, et al. The *Venturia inaequalis* effector repertoire is expressed in waves, and is dominated by expanded families with predicted structural similarity to avirulence proteins. Zenodo; 2022. <https://doi.org/10.5281/zenodo.6233645>.
175. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017;35(11):1026–8.
176. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*. 2011;9(2):173–5.
177. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinform*. 2010;11(1):431.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.