bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The *Venturia inaequalis* effector repertoire is expressed in waves, and is dominated by expanded families with predicted structural similarity to avirulence proteins from other fungi

Mercedes Rocafort[1], Joanna K. Bowen[2], Berit Hassing[1], Murray P. Cox[3], Brogan McGreal[2], Silvia de la Rosa[1], Kim M. Plummer[4], Rosie E. Bradshaw[3] and Carl H. Mesarich[1,*]

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

[2]The New Zealand Institute for Plant and Food Research Limited, Mount Albert Research Centre, Auckland 1025, New Zealand.

[3]Bioprotection Aotearoa, School of Natural Sciences, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand.

[4]Department of Animal, Plant and Soil Sciences, La Trobe University, AgriBio, Centre for AgriBiosciences, La Trobe University, Bundoora, Victoria 3086, Australia.

[*]Corresponding author: Carl H. Mesarich (c.mesarich@massey.ac.nz). Postal address: School of Agriculture and Environment, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand.

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. However, in the presence of corresponding host resistance proteins, these effectors are recognized as avirulence determinants to activate plant defences. To develop durable control strategies against scab, a better understanding of the molecular mechanisms underpinning subcuticular growth by

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

V. inaequalis is required. Likewise, more information is needed on the role that effectors play in activating, suppressing or circumventing resistance protein-mediated defences.

Results: We generated the first comprehensive RNA-seq transcriptome of *V. inaequalis* during colonization of apple. Analysis of this transcriptome revealed five *in planta* gene expression clusters or waves corresponding to three specific infection stages: early, mid and mid-late infection. Early infection was characterized by genes encoding plant cell wall-degrading enzymes (PCWDEs) and proteins associated with oxidative stress responses. Mid infection was characterized by genes encoding transporter proteins. Finally, mid-late infection was characterized by genes encoding PCWDEs and effector candidates (ECs), with most ECs belonging to expanded protein families. To gain insights into function, AlphaFold2 was used to predict the tertiary structures of proteinaceous ECs. Strikingly, 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 structural effector families. In addition, several other ECs, including an EC family with amino acid similarity to the AvrLm6 effector from *Leptosphaeria maculans*, were predicted to adopt a KP6/ferredoxin-like fold. Thus, proteins with a KP6/ferredoxin-like fold may represent yet another structural family of effectors shared among plant-pathogenic fungi.

Conclusions: Our study reveals the transcriptomic profile underpinning subcuticular growth by *V. inaequalis*, and reinforces the idea that fungal effectors share a limited number of structural folds. Importantly, our study also provides an enriched list of *V. inaequalis* ECs that can be investigated for roles in virulence and avirulence, and raises the possibility that apple resistance proteins can be engineered to recognize EC family folds or host components targeted by multiple EC family members.

Keywords

Venturia inaequalis; apple scab fungus; biotrophic subcuticular pathogen; effectors and effector families; virulence and avirulence; RNA-seq transcriptome; AlphaFold2 protein tertiary structure predictions.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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). Following the recognition of these MAMPs, plant defence responses are initiated that slow or halt growth of the fungal pathogen [3-5]. To circumvent or suppress these defence responses, plant-pathogenic fungi must secrete a collection of virulence factors, termed effectors, into their hosts [6-8]. These effectors are typically proteinaceous, but can also be secondary metabolites and small RNAs [6]. 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 resistance (R) proteins, to activate plant defences. Such effectors are termed avirulence (Avr) effectors because their recognition often 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% [11-13]. During biotrophic host colonization, *V. inaequalis* exclusively colonizes the subcuticular environment without penetrating the underlying plant epidermal cells [9, 14, 15]. It is here that the fungus develops specialized infection structures, known as stromata and runner hyphae [9, 14]. Stromata give rise to asexual conidia, but are also likely required for nutrient acquisition and effector secretion [9, 14, 16], while runner hyphae enable the fungus to radiate out from the initial site of host penetration, acting as a base from which

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

77 additional stromata can be differentiated [14]. At the end of the season, in autumn, *V. inaequalis*
78 switches to saprobic growth inside fallen leaves, where it undergoes sexual reproduction [9].

79 Scab disease is largely controlled through fungicides, which greatly accelerate the
80 development of fungicide resistance in *V. inaequalis* [17]. Scab-resistant apple cultivars, developed
81 through the incorporation of *R* genes, represent a more sustainable disease control option [12].
82 However, as races of *V. inaequalis* have been identified that can overcome resistance in apple
83 mediated by single *R* genes [12, 18], it is likely that multiple *R* genes will need to be stacked in these
84 cultivars to provide durable disease resistance. For this to be successful, prior knowledge of the
85 molecular mechanisms used by the fungus to overcome *R* gene-mediated resistance, including an
86 understanding of how these mechanisms impact effector function and pathogen fitness, will be
87 required. So far, however, there have been no publications describing the cloning of *Avr* effector genes
88 from *V. inaequalis.* Crucially, the genomes of multiple *V. inaequalis* isolates have been sequenced [11,
89 19-24] and bioinformatic studies have identified a large catalogue of effector candidates (ECs) from
90 which Avr effectors can be identified [20]. Most of these ECs belong to expanded families comprising
91 multiple members [20]. These include ECs of the AvrLm6-like family, which share amino acid sequence
92 similarity to AvrLm6 [16], an Avr effector from the fungal pathogen *Leptosphaeria maculans* (blackleg
93 of canola) [25], as well as ECs of the Ave1-like family [20], which share amino acid sequence similarity
94 to Ave1, an Avr effector from the fungal pathogen *Verticillium dahliae* (Verticillium wilt disease) [26].

95 To develop durable control strategies against scab disease, a better understanding of the
96 molecular mechanisms underpinning subcuticular growth by *V. inaequalis* is also required. To date,
97 subcuticular growth has been largely understudied, even though it is exhibited by many plant-
98 pathogenic fungi, including other crop-infecting members of the *Venturia* genus [14, 27-29], as well
99 as, for example, *Rhynchosporium* (scald disease of graminaceous plants) [30, 31] and *Diplocarpon* (e.g.
100 rose black spot) [32, 33]. In recent years, host colonization by plant-pathogenic fungi has been studied
101 by transcriptomic analysis [34-36]. However, comprehensive transcriptomic studies focusing on the
102 subcuticular parasitic strategy are not available. Indeed, while previous expression data derived from

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

interactions between *V. inaequalis* and susceptible apple have been published [13, 20], these data have only been 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 gene expression waves associated with general fungal and effector biology. 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 from this fungus that can be investigated for potential roles in virulence and avirulence.

Results

Different stages of subcuticular host infection by *V. inaequalis* display distinct gene expression profiles

To investigate changes in *V. inaequalis* gene expression during host colonization, we set up an infection time course involving detached leaves of susceptible apple cultivar *M.* x *domestica* 'Royal Gala'. This time course involved 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 was compared to growth of the fungus in culture at 7 dpi.

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 (**Fig. 1A**). These appressoria were produced at the end of germ tubes or directly from conidia and were frequently surrounded by extracellular matrix (ecm) (**Fig. 1A**). At 24 hpi, primary hyphae had developed, indicating that subcuticular colonization was underway. Then, by 2 and 3 dpi, *V. inaequalis* had developed subcuticular runner hyphae from stromata, and stromata had undergone a rapid expansion in size through non-polar division (**Fig. 1A**). By 5 dpi, fungal biomass had accumulated extensively in the subcuticular environment and more stromata had started to develop from runner hyphae (**Fig.**

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

129 **1A**). Often, these stromata had formed conidiophores, from which conidia had developed (**Fig. 1A**).
130 Finally, at 7 dpi, conidia of *V. inaequalis* had started to rupture through the plant cuticle (**Fig. 1A**) and,
131 at this time point, the first macroscopic olive-brown lesions, indicative of heavy sporulation, were
132 apparent. This represented the end of the biotrophic infection stage, as detached apple leaves had
133 started to decay after this time point.

134 Inspection of the RNA-seq data revealed that the percentage of reads mapping to the
135 *V. inaequalis* genome [20] increased as the infection time course progressed (**Additional file 1: Table**
136 **S1**). Furthermore, the biological replicates clustered robustly within time points, and a clear distinction
137 between the early and mid-late infection stages, as well as between the *in planta* and in-culture
138 growth conditions, was observed (**Fig. 1B**).

139

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

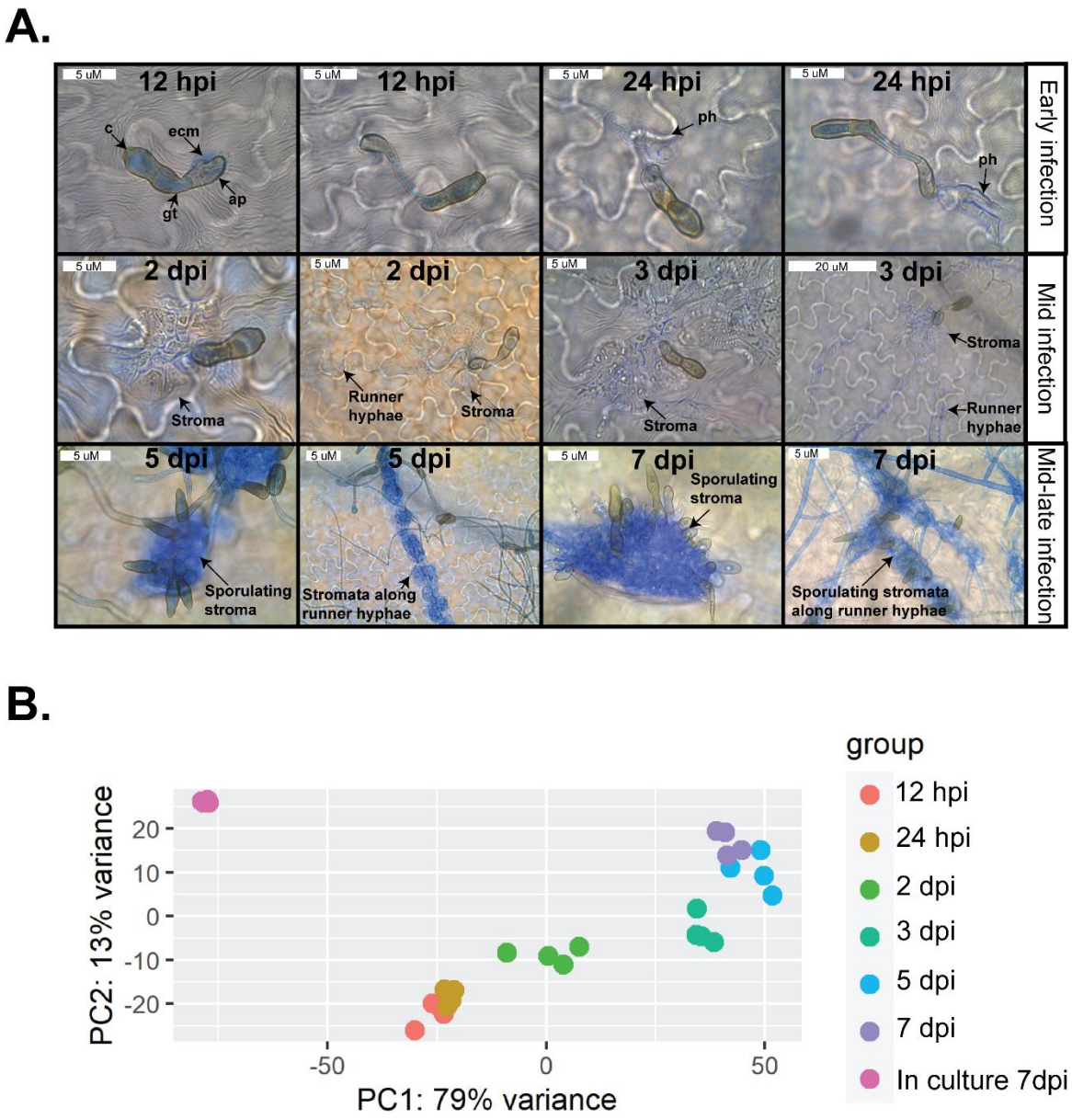


Fig. 1 Microscopic and transcriptomic profile of *Venturia inaequalis* during infection of detached leaves from susceptible apple cultivar 'Royal Gala'. **A.** Microscopic evaluation of *V. inaequalis* during colonization of apple leaves. Infection structures observed by bright-field microscopy were 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.** 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 overlaying potato dextrose agar. Four biological replicates were used for each *in planta* time point and the in culture growth condition. hpi: hours post-inoculation; dpi: days post-inoculation.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

149

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

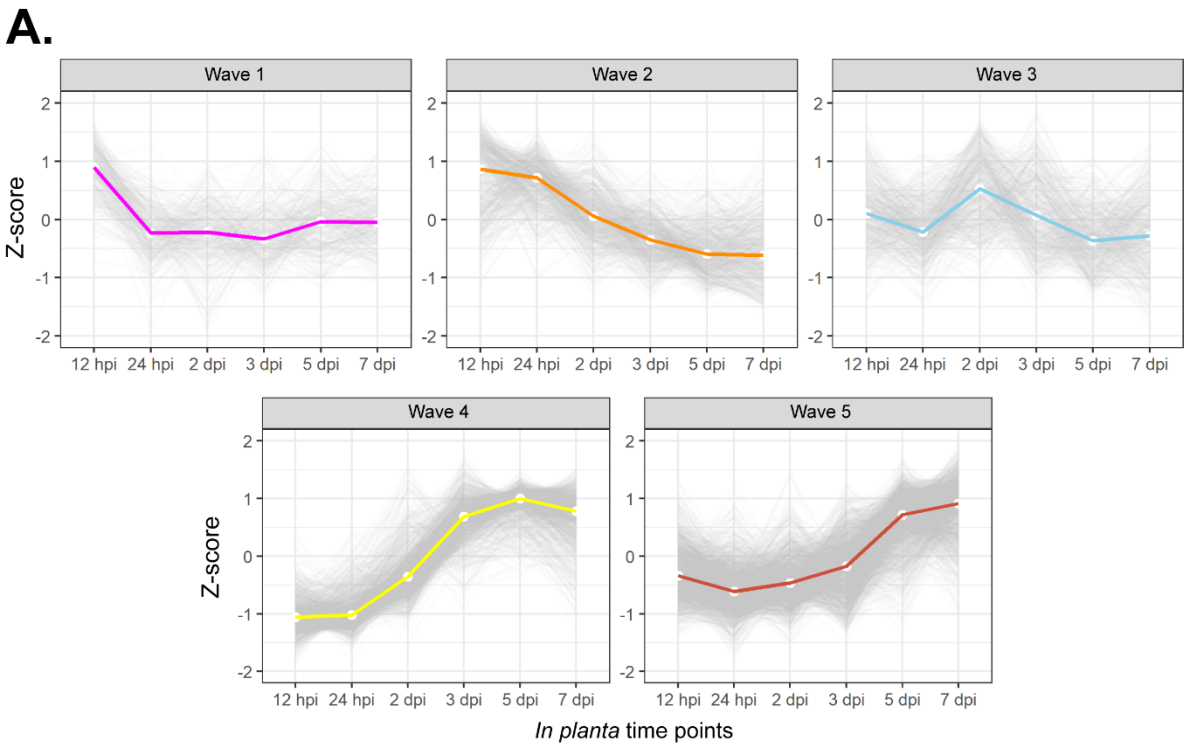
151 We set out to identify which genes of *V. inaequalis* are up-regulated during host colonization,
152 when compared to growth in culture. For this purpose, we updated the current gene catalogue for
153 isolate MNH120 [20] to increase the total number of annotated genes, including those that encode
154 ECs, which are notoriously difficult to predict in fungi (**Additional file 2: Fig. S1**). Here, our approach
155 was to predict as many genes as possible and, as a consequence, it is expected that many spurious
156 genes were included in the annotation. In total, 24,502 genes, excluding splice variants, were
157 predicted and, of these, 3,563 were up-regulated at one or more *in planta* time points (*p* value of 0.01
158 and $\log_2$-fold change of 1.5) (**Additional file 3**).

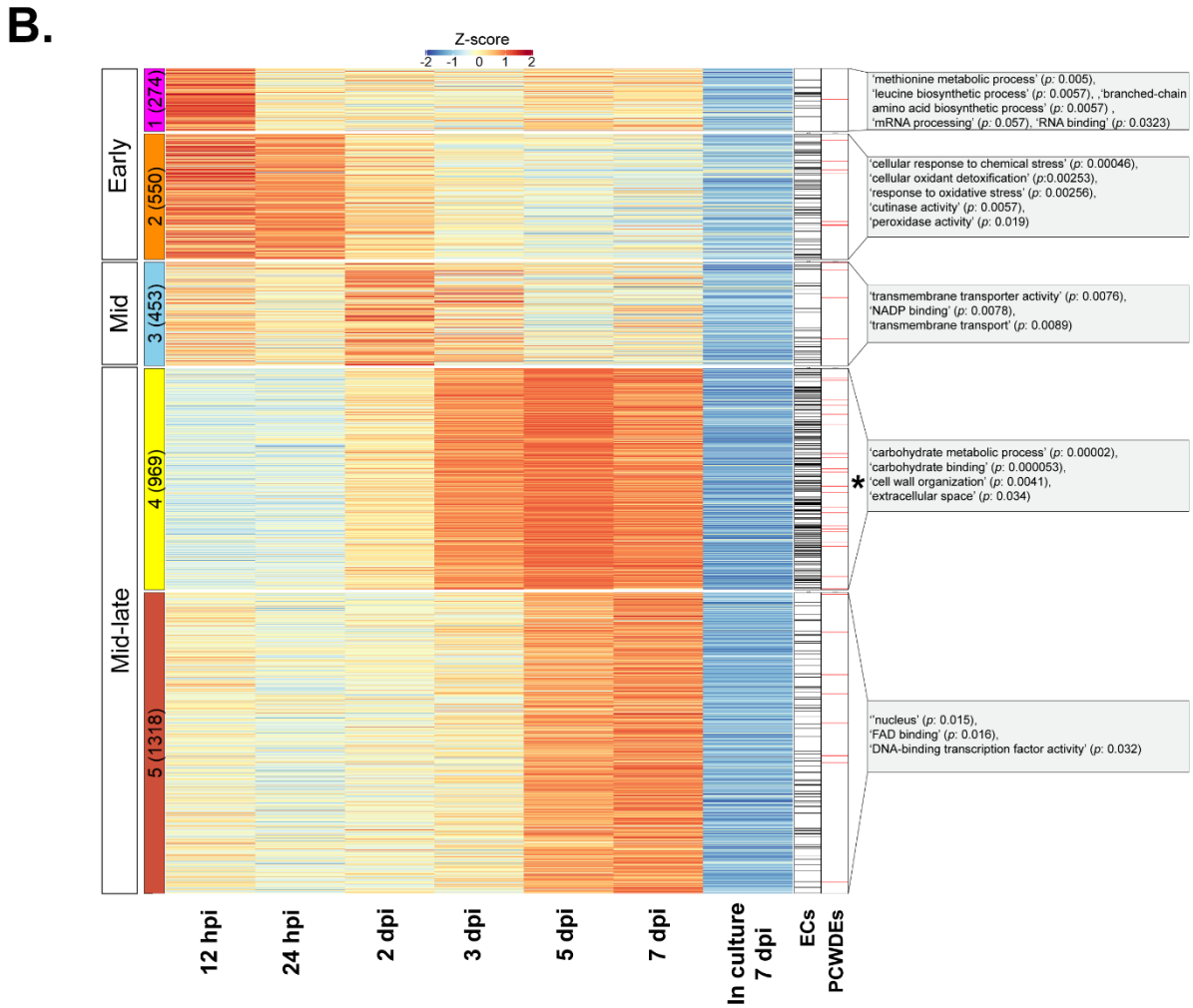
159 The total set of *in planta* up-regulated genes was used to identify host infection-specific gene
160 expression clusters, henceforth referred to as temporal waves. Here, all expression data was scaled
161 across all samples (Z-score) to visualize the gene expression deviation from the overall mean. For
162 clustering, the parameters were set to specifically identify the minimum number of gene clusters for
163 which a distinct gene expression profile could be observed (**Fig. 2A**). In total, five distinct host
164 infection-specific gene expression waves (**Fig. 2A**), representing three separate infection stages (**Fig.**
165 **2B**), were identified (**Additional file 4: Fig. S2**). Waves 1 (274 genes) and 2 (550 genes) represented
166 early infection (**Figs. 2A and 2B**). Genes of these two waves demonstrated their highest level of
167 expression at 12 hpi, with expression largely plateauing (wave 1) or trending downwards (wave 2)
168 throughout the remaining infection time points (**Figs. 2A and 2B**). Wave 3 (453 genes) represented
169 mid infection. This wave contained genes that had a peak level of expression at 2 dpi (**Figs. 2A and**
170 **2B**). Finally, waves 4 (959 genes) and 5 (1,318 genes) represented mid-late infection (**Figs. 2A and 2B**).
171 Genes of wave 4 displayed their lowest level of expression at 12 and 24 hpi, with expression strongly
172 increasing through 2 and 3 dpi, peaking at 5 dpi (**Figs. 2A and 2B**). A similar profile was also observed
173 for genes of wave 5, but with expression strongly increasing from 3 dpi and peaking at 7 dpi (**Figs. 2A**
174 **and 2B**).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

175 To determine which biological processes are overrepresented in the five host infection-
176 specific gene expression waves, gene ontology (GO) (**Fig. 2B, Additional file 5: Fig. S3, Additional file**
177 **6**) and protein family (Pfam) enrichment analyses (**Additional file 7**) were performed. Early infection
178 (waves 1 and 2) was characterized by GO terms associated with high metabolic activity and responses
179 to oxidative stress. In contrast, mid infection (wave 3) was characterized by GO terms associated with
180 transmembrane transport, while mid-late infection (waves 4 and 5) was mostly characterized by GO
181 terms associated with carbohydrate metabolism (**Fig. 2B**).
182

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.





183

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 2 Genes of *Venturia inaequalis* up-regulated during infection of susceptible apple cultivar 'Royal Gala', relative to growth in culture, belong to one of five distinct expression waves. **A.** Expression profile of the five distinct expression waves at 12 and 24 hours post-inoculation (hpi), as well as 2, 3, 5 and 7 days post-inoculation (dpi), relative to growth in culture (7 dpi). **B.** Heatmap of all *V. inaequalis* genes up-regulated *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 up-regulated *in planta* were clustered using hclust method Ward.D2 and Euclidian distance. Coloured block labels on the left indicate gene expression waves. Numbers in brackets indicate number of genes per wave. Black asterisk indicates the wave significantly enriched for ECs (*p* value: 3.644e-14). Grey boxes indicate enriched gene ontology (GO) terms per wave highlighted in the main text. ECs: effector candidates; PCWDEs: plant cell wall-degrading enzymes; *p*: value.

Plant cell wall-degrading enzyme (PCWDE)- and transporter-encoding genes of *V. inaequalis* are up-regulated during colonization of the subcuticular environment

To understand the nutritional requirements of *V. inaequalis* during colonization of the subcuticular environment, we investigated the expression profile of genes that putatively encode PCWDEs and transporters during the five host infection-specific waves (**Figs. 3 and 4**). Within the early infection wave 1, only four genes encoding PCWDEs, mostly of the auxiliary activity (AA) enzyme class, that are thought to aid the activity of subsequent classes of PCWDEs, were expressed. Within wave 2, an increase in the overall number of genes encoding PCWDEs, mostly of the AA and carbohydrate esterase (CE) classes (e.g. CE5 cutinases), was observed. Within wave 3 (mid infection), only six PCWDE-encoding genes were expressed, representing a mixture of AA and CE enzymes. Later, within wave 4, a striking increase in the number of genes encoding PCWDEs, mostly of the glycoside hydrolase (GH) class (e.g. GH28s), was observed. Finally, during wave 5 (mid-late infection), genes encoding GH and AA enzymes predominated.

The expression profile of genes encoding sugar and nitrogen transporters also differed during early and mid-late infection. During early infection (waves 1 and 2), only two genes encoding sugar

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transporters, together with seven genes encoding nitrogen-associated transporters, were up-regulated. Most of the up-regulated genes encoding nitrogen transporters belonged to the oligopeptide (OPT) transporter family. Many OPT transporter-encoding genes were found to be co-expressed highly with aspartyl protease-encoding genes ($R > 0.8$). During mid infection (wave 3), only two genes encoding sugar transporters were expressed and genes encoding nitrogen transporters started increasing expression. During mid-late infection (waves 4 and 5), there was also a striking increase in the expression of genes encoding sugar and nitrogen transporters (**Fig. 4**).

Taken together, the differences observed between the expression of genes encoding PCWDEs and transporters across the different waves likely reflect the differences in the nutritional requirements by *V. inaequalis* during early and mid-late subcuticular host colonization.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

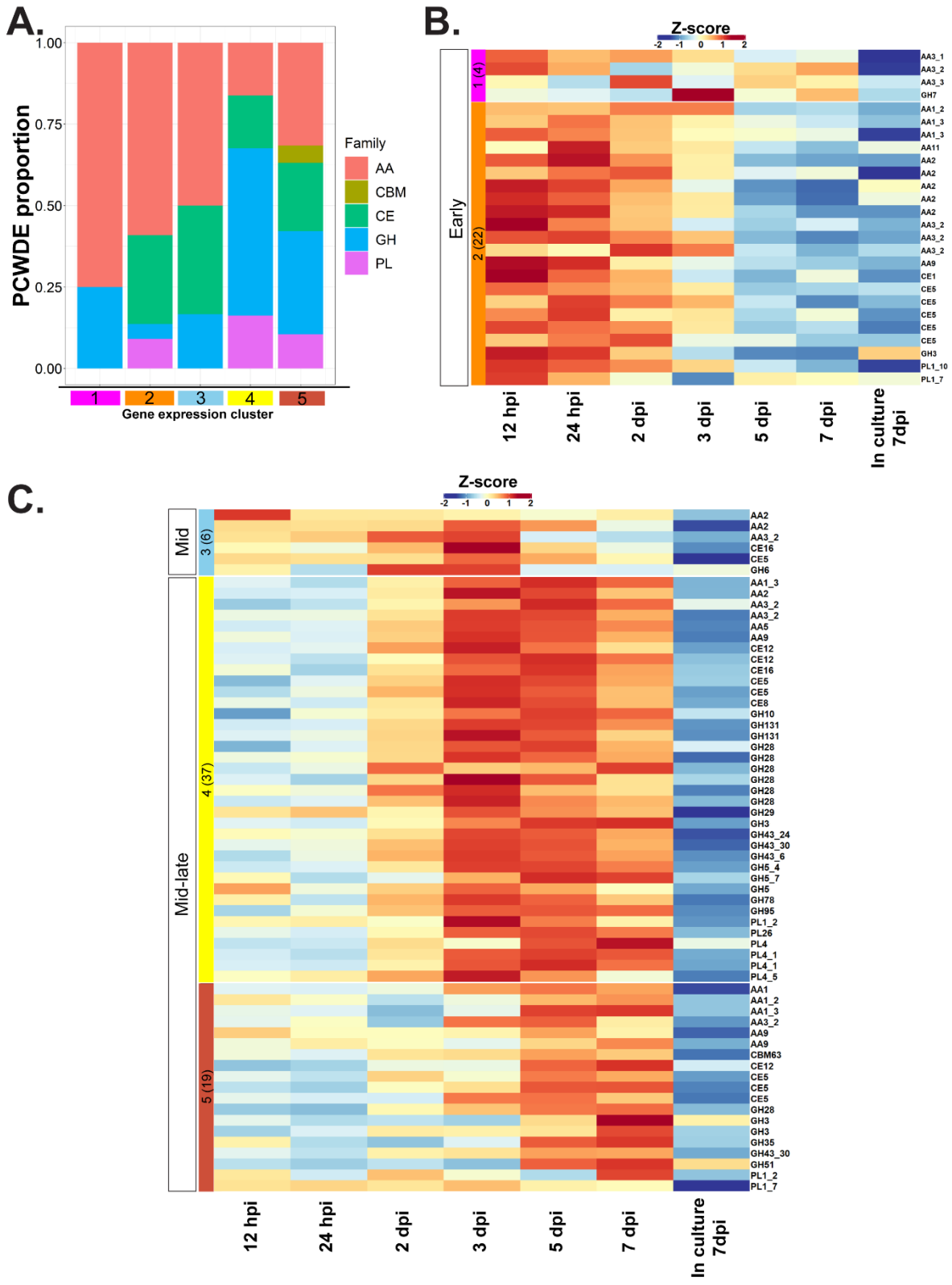


Fig. 3 Plant cell wall-degrading enzyme (PCWDE)-encoding genes of *V. inaequalis* up-regulated during infection of susceptible apple cultivar 'Royal Gala', relative to growth in culture. **A.** Proportion of *in planta* up-regulated PCWDE-encoding genes in each host infection expression wave. **B.** Heatmap of PCWDE-encoding genes up-

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

regulated *in planta* during early infection waves 1 and 2. **C.** Heatmap of PCWDE-encoding genes up-regulated during mid infection (wave 3) and mid-late infection (waves 4 and 5). Labels on the right indicate expression profiles corresponding to genes encoding carbohydrate-active enzymes (CAZymes). Block labels on the left indicate gene expression wave. 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. AA: auxiliary activity; GH: glycoside hydrolase; CE: carbohydrate esterase; PL: polysaccharide lyases; CBM: carbohydrate-binding module. Numbers in brackets indicate number of genes per wave.

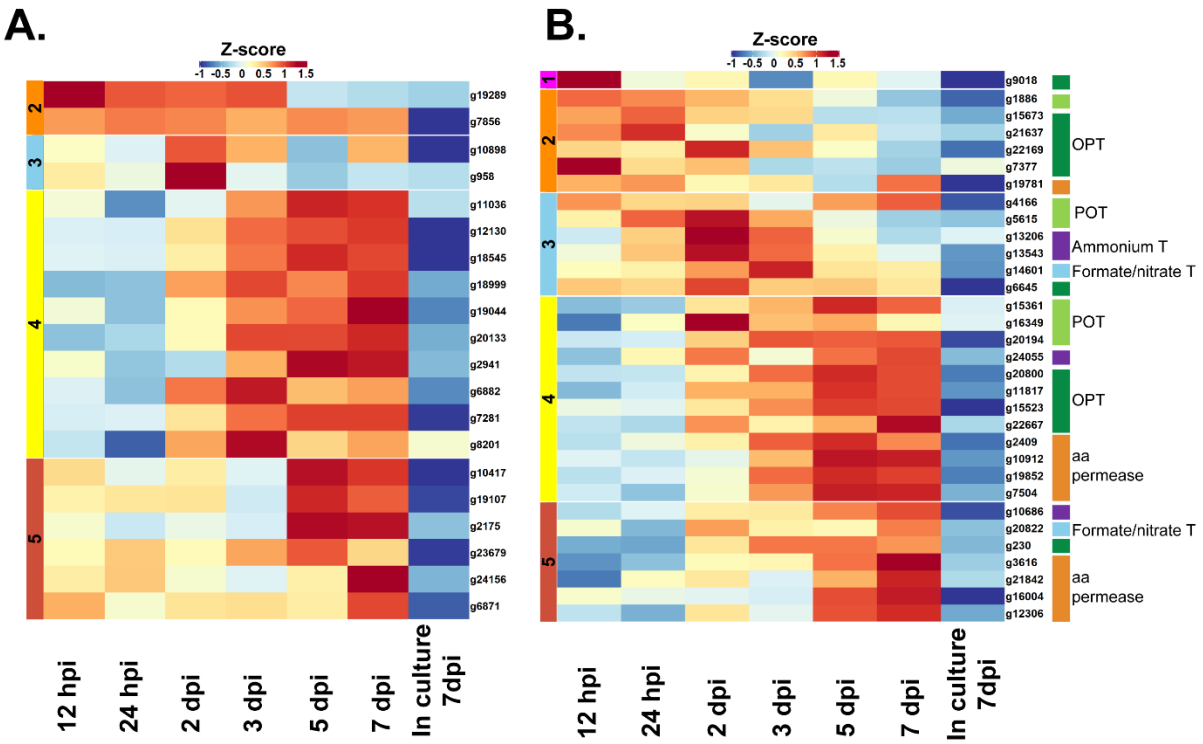


Fig. 4 Sugar- (**A.**) and nitrogen-associated transporter (**B.**) genes of *Venturia inaequalis* up-regulated during infection of susceptible apple cultivar 'Royal Gala', relative to growth in culture. Gene expression data are scaled rlog-normalized counts across all samples (Z-score), averaged from four biological replicates. hpi: hours post-inoculation; dpi: days post-inoculation. Sugar transporters in **A.** (PF0083), nitrogen associated transporters in **B.** OPT: oligopeptide transporter (PF03169); POT: proton-dependent oligopeptide transporter (PF00854); Ammonium T: ammonium transporter (PF00909); aa permease: amino acid permease (PF13620); Formate/nitrate T: formate/nitrate transporters (PF01226).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

241

242 **Genes encoding components of the predicted *V. inaequalis* effector repertoire are predominantly**
243 **expressed in the early and mid-late waves**

244 To determine whether the predicted *V. inaequalis* effector repertoire is expressed in temporal
245 waves during host colonization, the expression of genes encoding putative effectors was examined,
246 including genes responsible for the biosynthesis of secondary metabolites and genes encoding
247 proteinaceous ECs. A total of 26 secondary metabolite gene clusters were predicted using antiSMASH
248 v6.0.1; however, only seven were up-regulated during host-colonization. Of these, two were non-
249 ribosomal peptide synthetase-like (NRPS-like) clusters, two were type I polyketide synthase (T1PKS)
250 clusters, two were indole clusters, and one was a siderophore cluster (**Additional file 8: Fig. S4**). All of
251 the up-regulated secondary metabolite genes, except the lowly expressed T1PKS1 core biosynthetic
252 gene *g15330*, were expressed during the mid-late infection waves (waves 4 and 5) (**Additional file 8:**
253 **Fig. S4**).

254 Ribosomally-synthesized and post-translationally modified peptide (RiPP) gene clusters
255 cannot be predicted using antiSMASH. Therefore, RiPP biosynthetic clusters present in the *V.*
256 *inaequalis* genome were manually identified by searching for genes that encode a protein with a
257 DUF3328 (or DUF3328-like) domain, as this domain is known to be associated with RiPP biosynthesis
258 [37]. In total, nine dikaritin RiPP clusters were identified, with each cluster containing one or more
259 DUF3328 protein-encoding genes as well as a RiPP precursor gene (**Additional file 9: Fig S5**). Three of
260 the dikaritin RiPP clusters were made up of genes that were highly expressed and up-regulated during
261 host-colonization. Of these, all were expressed during the mid-late infection waves (4 and 5). The most
262 highly expressed dikaritin precursor gene (*g7830*, from the dikaritin-2 cluster) was expressed during
263 wave 4. This gene corresponds to the previously identified gene, *cellophane-induced* (*Cin) 3*, which
264 was formerly considered to be a repeat-containing EC protein [14, 20].

265

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

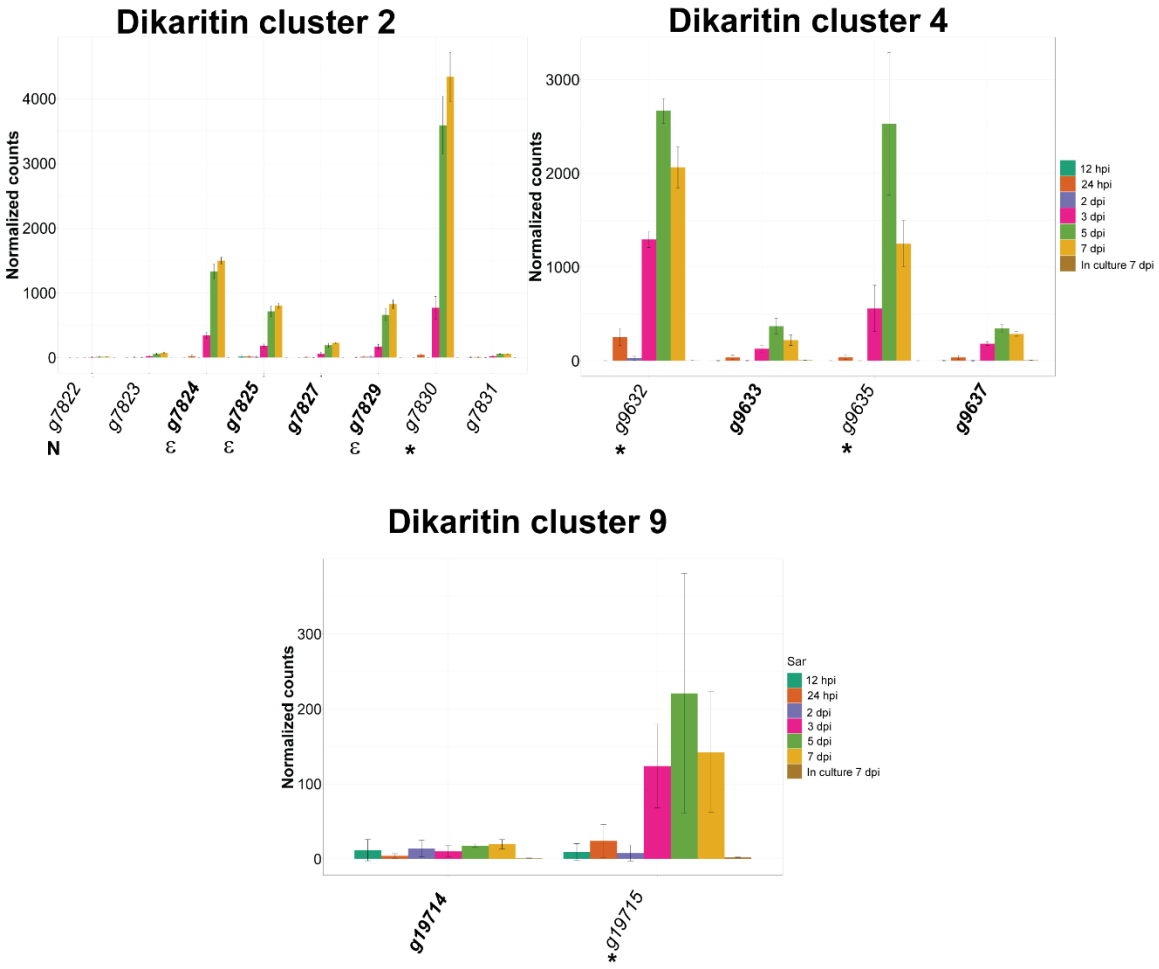


Fig. 5 Expression of ribosomally-synthesized and post-translationally modified peptide (RiPP) dikaritin gene clusters from *Venturia inaequalis* that are up-regulated during colonization of susceptible apple cultivar 'Royal Gala', relative to growth in culture. Gene expression data are DESeq2-normalized counts, averaged from four biological replicates per infection time point *in planta*, with error bars representing standard deviation (hpi, hours post-inoculation; dpi, days post-inoculation) and during growth in culture. Genes marked with bold characters encode putative DUF3382 proteins. 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.

Proteinaceous ECs were investigated including genes that encode small secreted proteins of ≤400 amino acid residues or larger secreted proteins with an effector prediction by EffectorP v3.0. In total, 1,369 genes encoding putative proteinaceous ECs were identified (**Additional file 10: Fig. S6**).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Amino acid similarities among these ECs were investigated using BLASTp, and the ECs grouped into families using spectral clustering. Based on this analysis, 759 of the proteinaceous EC were grouped into 118 families ranging in size from 2 to 75 members. Of these, 32 families were found to be expanded with five or more members. Only 610 of the proteinaceous ECs were singletons that did not belong to any family (**Additional file 11**). Of the genes encoding proteinaceous ECs, 686 were both up-regulated *in planta* and expressed during one of the early or mid-late infection waves. More specifically, early infection (waves 1 and 2) contained transcripts from 139 *EC* genes, while the mid-late infection stage (waves 4 and 5) contained transcripts from 485 *EC* genes (**Fig. 6**). The early infection waves were mostly composed of singletons, while the mid-late infection waves were mostly composed of expanded families (**Fig. 6**).

Most of the *EC* genes expressed during early infection (waves 1 and 2) encoded proteins that lacked predicted functional domains. However, two genes from wave 2 encoded proteins with an Egh16-like virulence factor domain and had amino acid sequence similarity to the appressorium-specific Gas1 effector from the rice blast fungus *Magnaporthe oryzae* [38], hereafter named the Gas1-like family. Three of four identified expanded gene families expressed during the early infection waves also encoded proteins with predicted functional domains. These were family 8 with a stress-up regulated Nod19 domain (PF07712), hereafter named the Nod19 family (39 members), family 13 with a hydrophobic surface-binding protein A (HsbA) domain (PF12296), hereafter named the HsbA family (15 members), and family 37 with a common fold in several fungal extracellular membrane proteins (CFEM) domain (PF05730) [39], hereafter named the CFEM family (16 members).

The Nod19 family is the third most expanded EC family in *V. inaequalis*, and most Nod19 family members were encoded by genes that were both up-regulated *in planta* and expressed during early host colonization (waves 1 and 2) (**Fig. 6 and Additional file 12**: **Fig. S7A**). Alignment of all members from this family with Nod19 domain-containing proteins from *Medicago truncatula* (barrel medic) and the potato late blight oomycete *Phytophthora infestans* revealed that they all contained the conserved motif H-X-H-X-GG-X18/20-Y (**Additional file 12**: **Fig. S7B**). The HsbA and CFEM families are mostly

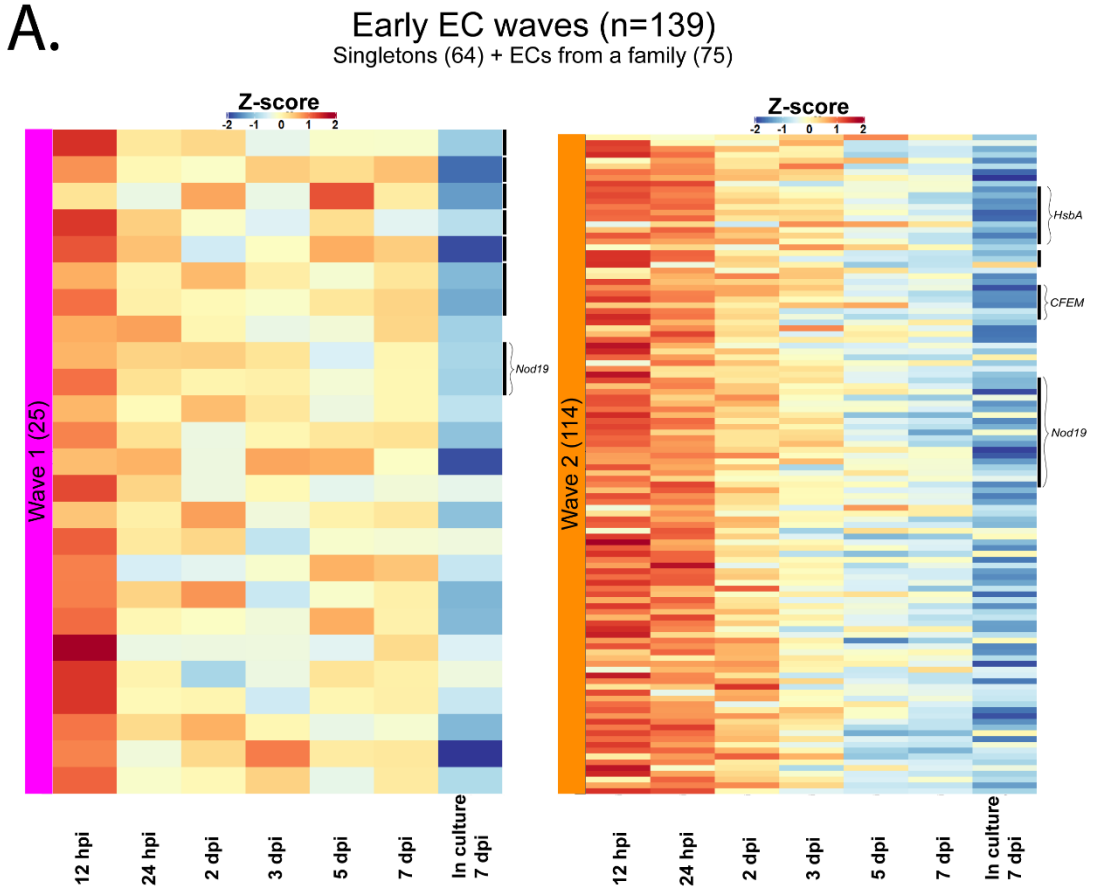
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

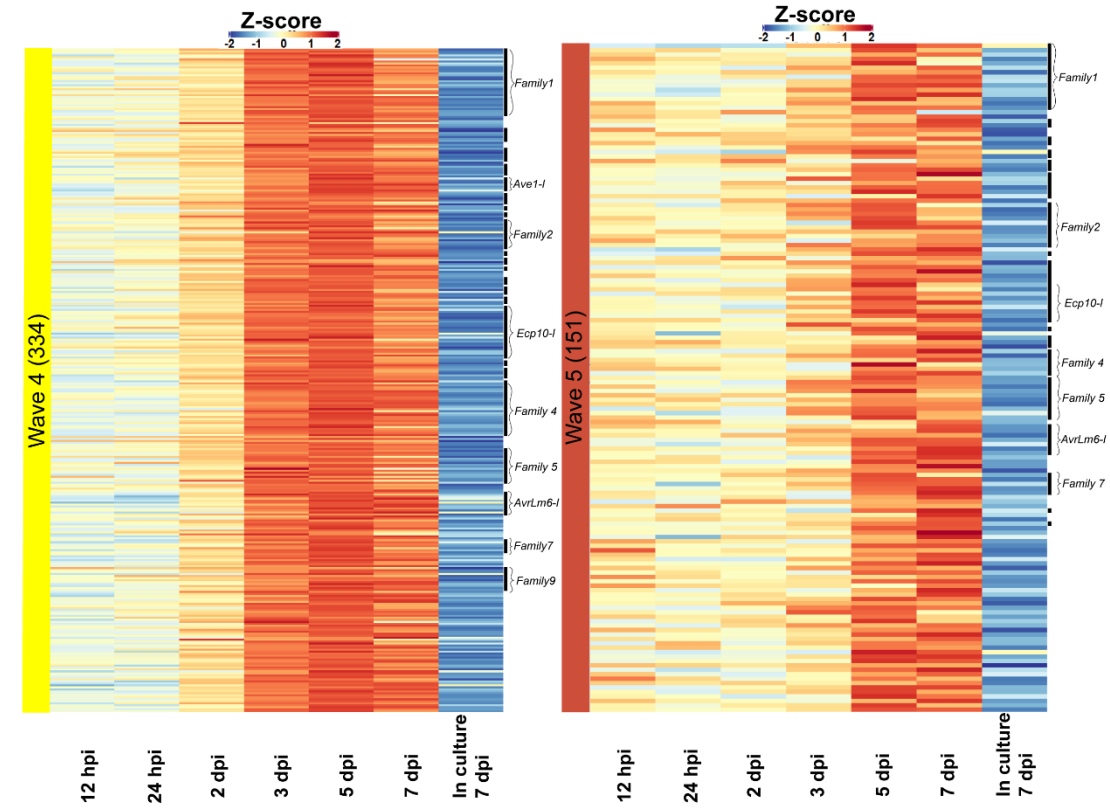
305 expressed during early host-colonization (**Fig. 6**). However, each of these families had genes with
306 contrasting expression profiles, with most members expressed during early infection (waves 1 and 2)
307 and a few members expressed during mid-late infection (waves 4 and 5) (**Additional file 13**: **Fig. S8**).
308 Interestingly, the *HsbA* genes expressed during early host colonization appeared to be co-expressed
309 with cutinases (R > 0.8).

310 Another family with contrasting expression among members during host colonization was the
311 *Cin* 1 family (**Additional file 13**: **Fig. S8**) [40]. This family contains the *Cin1* gene (*g8385*), which encodes
312 a cysteine-rich protein with eight repeats, and two *Cin1-like* genes (*Cin1-like 1*: *g10529*, *Cin1-like 2*:
313 *g13013*), which encode smaller proteins with only one repeat. The Cin1 family has no homology to
314 known proteins and is specific to the *Venturia* genus (**Additional file 14: Fig. S9**). *Cin1* was expressed
315 during mid-late infection (wave 4), and it was the most highly expressed gene during mid-late host
316 colonization. In contrast, the *Cin1-like* genes were expressed during early infection (wave 2)
317 (**Additional file 14**: **Fig. S9**).

318

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.





319

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 6 Genes encoding proteinaceous effector candidates (ECs) of *Venturia inaequalis* are expressed in waves during colonization of susceptible apple cultivar 'Royal Gala'. Early infection (**A.**) waves 1 and 2 are composed of 139 EC-encoding genes, and the mid-late infection (**B.**) waves 4 and 5 are composed of 485 EC-encoding genes. Black annotations on the right indicate that the proteins encoded by these genes belong to an expanded EC family (≥5 members), with brackets highlighting large families and families with amino acid sequence similarity to avirulence (Avr) effector proteins from other plant-pathogenic fungi. hpi: hours post-inoculation; dpi: days post-inoculation.

The mid-late infection waves (waves 4 and 5) possessed the largest number of transcripts from genes encoding proteinaceous ECs, with 485 genes expressed from 2 to 7 dpi. Indeed, wave 4 was enriched in ECs (*p* value: 3.644e-14). Notably, most expanded families, including those with amino acid similarity to effectors or Avr effectors from other plant-pathogenic fungi, were encoded by genes that were expressed during waves 4 and 5 (**Fig. 6, Additional file 15: Table S2**). For example, the *AvrLm6-like* family [16] had 21 up-regulated genes *in planta*, with most expressed during waves 4 and 5 (**Fig. 7A**). Likewise, the *Ave1-like* family [20] had eight genes up-regulated *in planta,* with most expressed during wave 4 (**Fig. 7A**). One of the genes in this family, *g18550*, was the second most highly expressed EC gene during mid infection, and encoded a protein with the highest amino acid similarity to Ave1 (39.3% pair wise identity). Another example was the *Ecp10-like* family, which encoded proteins with amino acid sequence similarity to Ecp10-1, an Avr EC from the tomato leaf mold fungus *Cladosporium fulvum*. The *Ecp10-1* family represented the second most expanded family in *V. inaequalis* with 61 members and, of these, 45 were up-regulated during waves 4 and 5 (**Fig. 7A**). Similarly, a family of six proteins with amino acid sequence similarity to Ecp39 from *C. fulvum* [41], hereafter named the Ecp39-like family, were predominantly encoded by genes that were expressed during waves 4 and 5 (**Fig. 7A**). Finally, an *Ecp6-like* gene (singleton), which encoded a protein with amino acid similarity to the Ecp6 effector from *C. fulvum*, was very highly expressed during wave 4 (**Fig. 7A**).

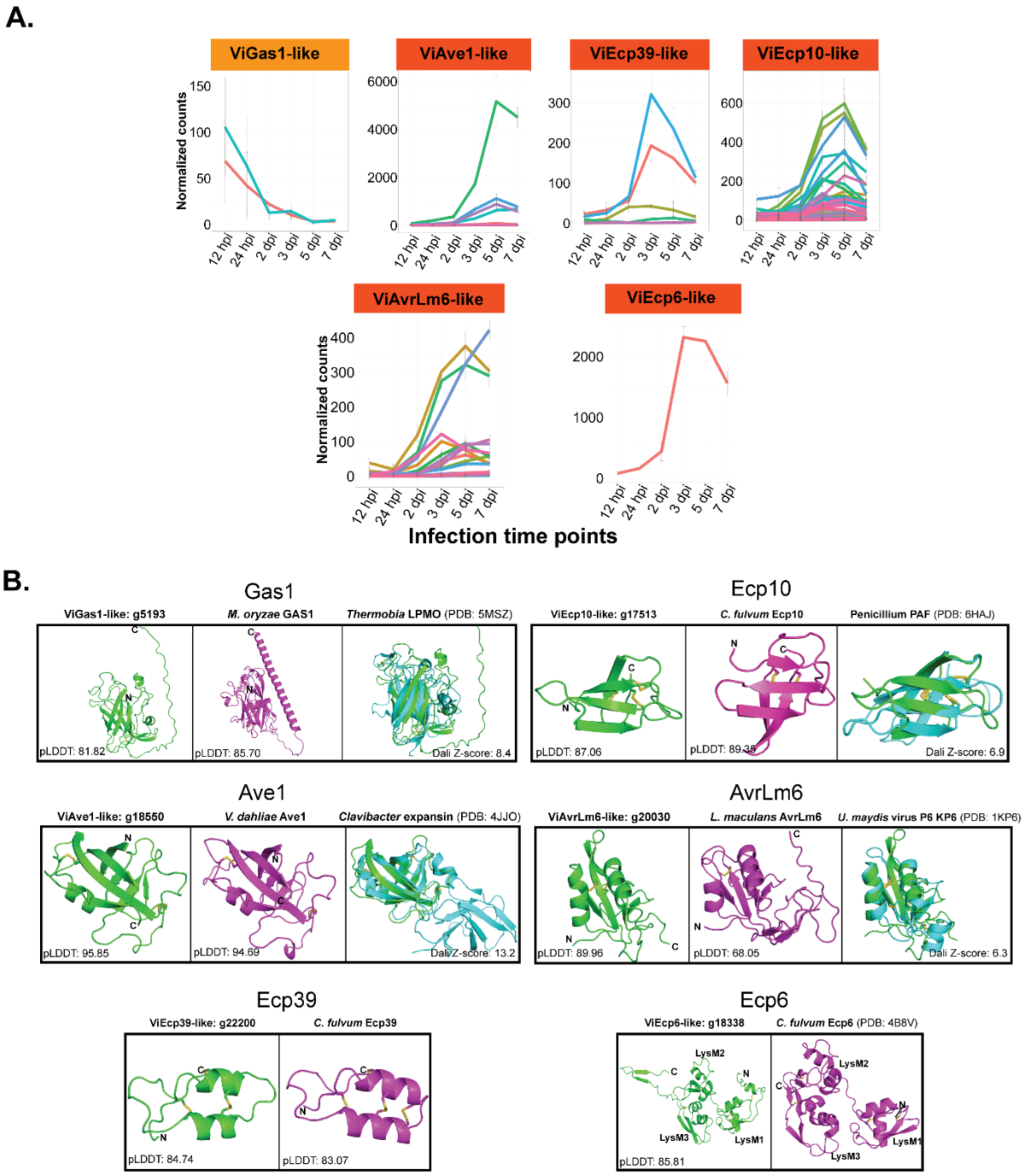
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

346 Interestingly, most expanded EC families encoded by genes expressed during waves 4 and 5
347 did not have amino acid sequence similarity to other proteins. Of these, family 1 was the most
348 expanded EC family in *V. inaequalis*. This family was encoded by 75 genes and, of these, 54 were up-
349 regulated *in planta* (**Fig. 8**). Likewise, family 2 was encoded by 32 genes, and of these, 28 were up-
350 regulated *in planta* (**Fig. 8**). In most cases, one or a few family members were very highly expressed,
351 when compared with the other family members, during host colonization (**Fig. 8**).
352
353 **Several expanded effector candidate families of *V. inaequalis* have predicted structural similarity to**
354 **Avr effectors from other plant-pathogenic fungi**
355 To gain insights into the putative function of proteinaceous ECs, we predicted their tertiary
356 structures using AlphaFold2 [42], which has been successfully benchmarked against pathogen
357 effectors of known tertiary structure [43, 44], and then investigated these structures for similarity to
358 proteins of characterized tertiary structure (and in some cases, function) using the Dali server [45].
359 This analysis was specifically performed on the most highly expressed member from each EC family
360 (which we referred to as the main family member) as well as each singleton, expressed during the host
361 infection-specific waves. In total, the tertiary structure was confidently predicted for the most highly
362 expressed member of 75% of EC families (94 families) and 64% of EC singletons (118 singletons)
363 (**Additional file 16**).
364 We investigated the predicted tertiary structures of EC families from *V. inaequalis* that had
365 amino acid sequence similarity to known effectors and characterized or candidate Avr effector
366 proteins from other plant-pathogenic fungi (**Additional file 17: Table S3**) (**Fig. 7**). During early
367 infection, the most highly expressed *Gas1-like* gene (**Fig. 7A**), as well as the similar *Gas1* effector from
368 *M. oryzae*, were predicted to encode a protein with an immunoglobin-like β-sandwich fold. These
369 structures were similar to chitin-binding lytic polysaccharide monooxygenase (LPMO) proteins (**Fig.**
370 **7B**). During the mid-late infection waves, the most highly expressed gene of the *Ecp10-like* family (**Fig.**
371 **7A**) was predicted to encode a protein with a compact β-folded structure, stabilized by three disulfide

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bonds, that is structurally similar to the antimicrobial PAFB protein from *Penicillium chrysogenum* [46] (**Fig. 7B**). *C. fulvum* Ecp10-1 on the other hand was predicted to adopt a smaller compact β-folded structure with only two disulphide bonds (**Fig. 7B**). The most highly expressed *AvrLm6-like* family gene (**Fig. 7A**), along with AvrLm6 from *L. maculans*, were predicted to encode a protein with a ferredoxin-like fold, stabilized by two disulfide bonds, that is structurally similar to the virally encoded *Ustilago maydis* (corn smut pathogen) P6 virus killer toxin α-subunit (KP6α) [47, 48] (**Fig. 7B**), as well as an EC from the Septoria tritici blotch pathogen *Zymoseptoria tritici* [49] (Zt-PK6-1) (**Fig. 8B**). The protein encoded by the most highly expressed *V. inaequalis Ave1-like* gene (**Fig. 7A**), together with Ave1, were predicted to adopt a double-psi β-barrel fold, stabilized by two conserved disulfide bonds, with high similarity to expansins (**Fig. 7B**). Both the most highly expressed *Ecp39-like* gene (**Fig. 7A**) and *CfEcp39* gene from *C. fulvum* were predicted to encode a protein that adopts a crambin-like fold (**Fig. 7B**). However, no analogous protein structures were found in the RCSB PDB using the predicted tertiary structure of the Ecp39-like protein as query. Finally, the ViEcp6-like protein was predicted to adopt a similar tertiary structure to the previously reported three-domain lysin motif (LysM) structure of Ecp6 from *C. fulvum* [50] (**Fig. 7B**).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



388

389 **Fig. 7** Effector candidate (EC) families and singletons from *Venturia inaequalis* with amino acid sequence
390 similarity to effectors and characterized or candidate avirulence (Avr) effector proteins from other plant-
391 pathogenic fungi. **A.** Expression data of ECs during host colonization are DESeq2-normalized counts, averaged
392 from four biological replicates, with error bars representing standard deviation (hpi: hours post-inoculation; dpi:
393 days post-inoculation). Light orange box indicates that the EC family belongs to the early expression waves (wave
394 1 or 2) and dark orange box indicates that the EC family belongs to the mid-late expression waves (wave 4 or 5).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B. Protein tertiary structures of candidate virulence and avirulence (Avr) effector proteins predicted by AlphaFold2, except Ecp6 from *Cladosporium fulvum* (RCSB PDB id: 4B8V). Disulfide bonds coloured in yellow. *V. dahliae* is *Verticillium dahliae*; *U. maydis* is *Ustilago maydis*; *L. maculans* is *Leptosphaeria maculans*. 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. Amino acid 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.

We investigated the predicted fold of the most highly expressed member of the *V. inaequalis* EC families and singletons that were novel and did not share obvious amino acid similarity with other ECs and/or Avr effectors. 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. 8, Additional file 17: Table S3**, **Additional file 18: Fig. S9**). Of the 94 EC families up-regulated *in planta*, 71 could be assigned a structure with confidence and of those, 12 families were structurally analogous to ECs and/or Avr effectors (**Additional file 17: Table S3**). Remarkably, many of these families were among the most expanded families in *V. inaequalis*, while only three singletons had structural similarity to other ECs and/or Avr effectors (**Table 1**).

The main 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 new MAX effector 6R5J [51] and the Avr effectors AvrPiz-t [52], Avr-Pia [53, 54], Avr-Pib [55], Arv1-CO39 [53] and Avr-Pik [56] from the rice blast fungus *M. oryzae*, as well as the ToxB virulence effector from *Pyrenophora tritici-repentis* [57], the fungal pathogen responsible for tan spot of wheat (**Additional file 17: Table S3**). An amino acid sequence alignment constructed from all family 1 protein members from *V. inaequalis*, as well as the identified MAX effectors from *M. oryzae*, revealed that these proteins lacked significant amino

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

421 acid sequence similarity, with a maximum pairwise identity of only 13.9% observed between *g13386*
422 and Avr1-CO39 (**Additional file 19**). However, these sequence-diverse proteins did share the
423 characteristic conserved disulphide bond between β1 and β5 that has previously been reported for
424 MAX effectors [53] (**Fig. 8A**).

425 Two additional expanded families, family 7 with 21 members and family 28 with 6 members,
426 as well as a small family with three members (family 38) and a singleton, shared a common β-sandwich
427 fold with structural similarity to the virulence effector ToxA from *P. tritici-repentis* [58], the Avr
428 effectors Avr2/Six3 from the wilt fungus *Fusarium oxysporum* [59], and AvrL567-A from the flax rust
429 fungus *Melampsora lini* [60] (**Fig. 8A, Additional file 17: Table S3, Additional file 18: Fig. S9**). Notably,
430 despite sharing an overall β-sandwich fold with similar topology, these proteins were very diverse at
431 the amino acid level, with a maximum amino acid identity of 12.7% (**Additional file 19**). All of the ToxA-
432 like families had a different number of β-sheets: family 7 with 10 β-sheets, family 28 with nine β-
433 sheets, family 38 with eight β-sheets and the singleton with nine β-sheets (**Additional file 18: Fig. S9**).

434 Another two families, family 15 with 12 members and family 47 with two members, were
435 found to have structural similarity to AvrLm4-7 and AvrLm5-9 from *L. maculans*, as well as Ecp11-1
436 from *C. fulvum* [61, 62] (**Fig. 8A, Additional file 17: Table S3, Additional file 18: Fig. S9**), which belong
437 to the *Leptosphaeria* AviRulence-Supressing (LARS) structural superfamily [62]. AvrLm4-7 has four
438 stranded β-sheets [61]; however, the *V. inaequalis* LARS-like effectors had the same predicted
439 topology but with a variable number of β-sheets. Specifically, members of family 15 had five β-sheets,
440 while members of family 47 had three β-sheets. The identified *V. inaequalis* LARS-like proteins share
441 an amino acid identity of only 7.8–16.9% with AvrLm4-7, AvrLm5-9 and Ecp11-1 (**Additional file 19**).
442 Interestingly, these structures are stabilized by a different number of disulfide bonds, and they lack
443 both the conserved disulfide bridge between the α-helix and β-strand, as well as the conserved
444 WR(F/L/V)(R/K) motif, previously reported for the LARS effectors (Lazar et al., 2020).

445 Family 49, with 3 members, was predicted to adopt a two-domain fold similar to the Avr1/Six4
446 and Avr3/Six1 Avr effectors from *F. oxysporum* (**Fig. 8A, Additional file 17: Table S3**), which are the

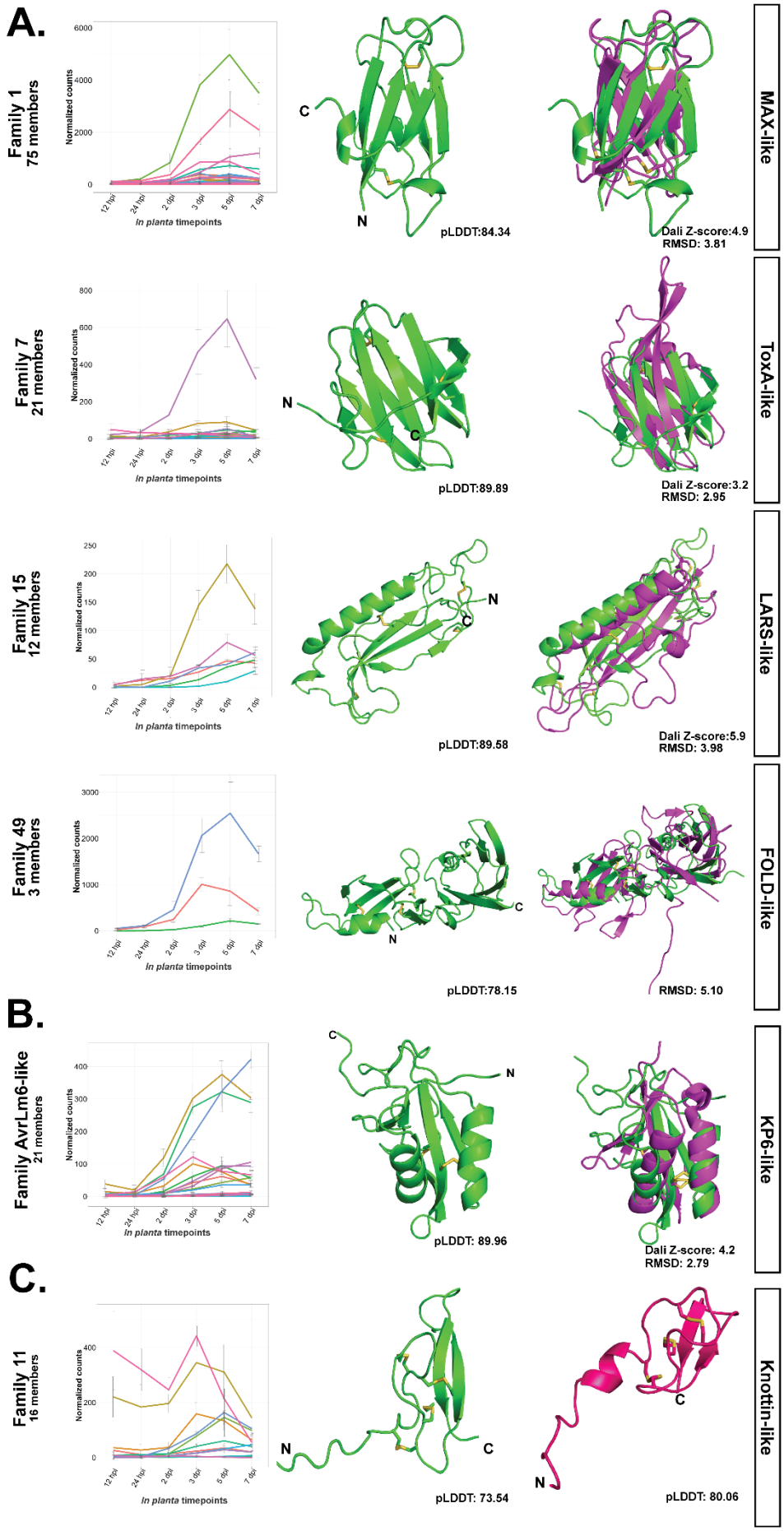
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

447 founding members of the *Fol* dual-domain (FOLD)-family effector family recently identified in *F.*
448 *oxysporum* [43]. The *V. inaequalis* FOLD-like family representative shared 17.6% amino acid identity
449 with Avr1/Six4 (**Additional file 19**) and, as part of this, the cysteine spacing pattern was conserved.
450 Many expanded families had a ferredoxin-like fold similar to the KP6 protein of the *U. maydis*
451 P6 virus (**Additional file 20: Fig. S9**) [47, 48] and the Zt-KP6-1 EC from *Z. tritici* [49]. This included the
452 representative from the AvrLm6-like family (**Fig. 8B**). Additionally, family 2 with 32 members, family 5
453 with 36 members, family 23 with 5 members, family 26 with 4 members, and two singletons were
454 predicted to adopt a KP6-like fold (**Additional file 20: Fig. S9**). All KP6-like proteins had two to four
455 disulphide bridges, two α-helices and a variable number of β-sheets (**Fig. 8B, Additional file 20: Fig.**
456 **S9**).
457 Many EC families were made up of proteins that were too small to be included in RCSB PDB
458 searches using the Dali server or that did not have any significant Dali hit. We investigated the general
459 structural classification of all proteins with a confidently predicted tertiary structure in this category
460 and identified three EC families and one EC singleton that were predicted to adopt a knottin-like fold
461 (**Additional file 16**). The most highly expressed member of family 11 (16 members) (**Fig. 8C**) and a
462 singleton were predicted to adopt a knottin-like fold with two β-sheets, as well as three disulfide
463 bonds that form an intramolecular knot (**Additional file 21: Fig. S10**). Additionally, family 12 with 17
464 members, family 24 with 12 members, and family 35 with 5 members, were predicted to adopt a
465 knottin-like fold using the SCOPe database [63] (**Additional file 21: Fig. S10**). However, these proteins
466 were not predicted to have a true intramolecular knot characteristic of knottin proteins based on our
467 current AlphaFold2 predictions. Instead, these proteins shared a common fold with cysteine-stabilized
468 αβ defensins, which are made up of a single α-helix and three β-strands [64]. In line with previous
469 reports [65, 66], Avr9 from *C. fulvum* was predicted to adopt a knottin-like fold (**Fig. 8C**).
470 Finally, we investigated which predicted protein SCOPe folds were most frequent in the ECs
471 expressed during the early (waves 1 and 2) and mid-late infection (waves 4 and 5) stages. During the
472 early waves, we observed that ECs had a great diversity of predicted folds. The most frequently

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

predicted fold in the early infection wave with a pLDDT score of 70 or higher was the 'STAT-like' fold (**Additional file 22: Fig. S11**). The 'STAT-like' fold was observed for the most highly expressed member of the HsbA-like family and family 22. In contrast, during the mid-late infection waves, many ECs were predicted to share similar folds. The most frequently predicted fold observed was the 'immunoglobin-like β-sandwich' fold, followed by 'ferredoxin-like', 'parallel coiled-coil' and 'knottin-like' folds (**Additional file 22: Fig. S11**). The 'immunoglobin-like β-sandwich' fold is made up of the EC families with a predicted MAX-like or ToxA-like fold, and interestingly, most of the genes encoding these ECs are expressed during wave 4. Likewise, the 'ferredoxin-like fold', which includes the KP6-like ECs, were also all encoded by genes that are expressed during infection wave 4. In conclusion, genes expressed during the mid-late infection waves encoded more ECs with similar folds than the early infection waves, and many of these had structural similarity to Avr effectors or ECs from other plant-pathogenic fungi.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



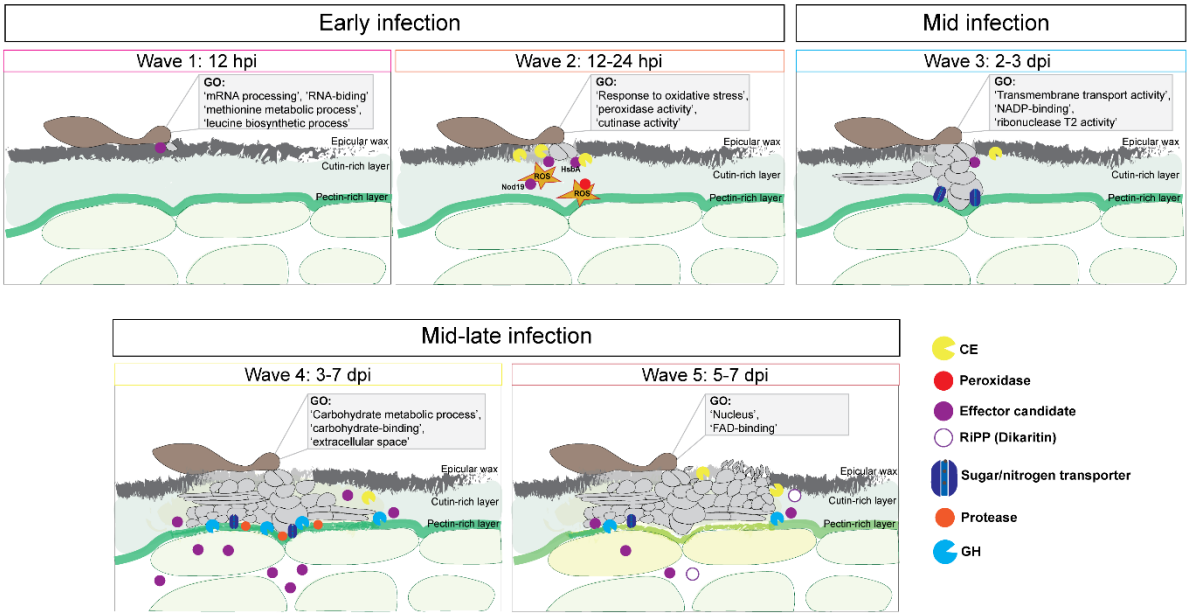
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 8 Predicted tertiary structures of effector candidates (EC) from *Venturia inaequalis*. **A.** Representative EC family members with structural similarity to known 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). 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). **C.** Predicted EC family members with a knottin-like fold. Knottin-like representative predicted family 11 (green; g8686) protein structure. Predicted structure of Avr9 from *Cladosporium fulvum* (pink). Gene expression data of up-regulated ECs 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.

Discussion

In this study, we present the first comprehensive temporal transcriptome of *V. inaequalis* during colonization of apple, covering six biotrophic time points from early (12 hpi) to late (7 dpi) infection. In doing so, we have, to our knowledge, also provided the first comprehensive *in planta* transcriptome of a subcuticular fungal pathogen. Based on our analysis of this transcriptome, we identified five host infection-specific gene expression waves for *V. inaequalis* that represent three infection stages: early, mid and mid-late infection (**Fig. 9**). These host infection-specific gene expression waves were biologically distinct from each other and were enriched for different GO terms.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



514

515 **Fig. 9** Summary of *Venturia inaequalis* host infection stages during biotrophic colonization of susceptible apple
516 leaves. GO: gene ontology; RiPP: ribosomally synthesized and post-translationally modified peptide; CE:
517 carbohydrate esterase, GH: glycoside hydrolase; hpi: hours post-inoculation; dpi: days post-inoculation.

518

519 Early infection was characterized by the expression of genes associated with metabolic
520 processes (wave 1) and responses to oxidative stress (wave 2). More specifically, wave 1 was enriched
521 for GO terms associated with RNA-binding and the synthesis of branched amino acids. Branched amino
522 acids play an important role in fungal pathogenesis [67-69], with leucine metabolism shown to be
523 crucial for host infection by *M. oryzae* and *Fusarium graminearum* [68], and methionine biosynthesis
524 found to be associated with the regulation of multiple cellular processes [69, 70]. The early expression
525 of genes associated with the synthesis of branched amino acids and methionine could suggest a similar
526 crucial role for these amino acids in the pathogenesis of *V. inaequalis*.

527 Wave 2 was enriched for GO terms associated with cutinase activity and the mitigation of
528 chemical and oxidative stress. Here, a large number of CE-encoding genes, specifically of the CE5 and
529 CE1 families, were up-regulated during early host colonization, which is in line with previous reports
530 showing that localized enzymatic hydrolysis is needed for penetration of the apple cuticle by

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

V. inaequalis [9, 71]. Interestingly, we observed that genes encoding proteins with an HsbA domain were co-expressed with cutinase-encoding genes during early infection, suggesting that, as was previously hypothesized, these putative hydrophobic surface-binding proteins may recruit cutinases to facilitate the degradation and digestion of the hydrophobic apple cuticle [20, 72, 73].

Wave 2 was additionally characterized by genes associated with oxidative stress tolerance, such as those encoding peroxidases, while genes encoding proteins with a 'stress up-regulated Nod19' domain were also abundant, forming part of an expanded Nod19 EC family in *V. inaequalis*. The function of proteins with a stress up-regulated Nod19 domain is currently unknown. However, these proteins have been suggested to be associated with responses to abiotic and biotic stress [74-77]. Based on these studies and the expression of Nod19 domain-containing genes during the early infection waves, the Nod19 proteins of *V. inaequalis* could represent a novel family of effectors that play a role in modulating oxidative stress during early infection of apple.

Interestingly, only around 20% of the *EC* genes from *V. inaequalis* were up-regulated during early infection and, with the exception of genes encoding the Nod19, HsbA and CFEM EC families, were mainly singletons. The CFEM family is defined by the presence of a CFEM domain with eight conserved cysteines. This domain is commonly found in proteins of pathogenic fungi [78] and has been shown to display diverse functions [79-82]. Some CFEM proteins promote cell death and chlorosis [79, 80], while others suppress plant cell death [81] or are associated with the development of appressoria [82]. The *CFEM* family in *V. inaequalis* has a diverse expression profile, indicating that genes of this family might play distinct roles during host colonization. Another EC family expressed during early host colonization was the *Gas1-like* family. Recently, Gas1-like proteins have been observed to form part of the widely distributed fungal effectors with chitinase activity (EWCA) family [83]. The tertiary structure of the representative *V. inaequalis* Gas1-like protein was predicted to be structurally similar to different chitin-binding LPMOs, for which EWCA proteins have been predicted to adopt the same fold [83]. EWCA proteins are secreted chitinases that degrade immunogenic chitin fragments to

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

suppress chitin-triggered immunity in plants [83]. Thus, it is tempting to speculate that members of the Gas1-like protein family from *V. inaequalis* play a similar role.

Mid infection (wave 3) was characterized by the expansion of subcuticular infection structures (stromata and runner hyphae), and was enriched for GO terms associated with transmembrane transport. Given that mid infection was characterized by the development of infection structures, we suggest that this wave contains genes that are highly expressed in stromata, mostly transporters for nitrogen metabolism and morphogenesis. However, it is important to point out that the mid infection wave was not well defined, overlapping partially with the mid-late infection waves 4 and 5. This is likely due to the asynchronous nature of *V. inaequalis* infection.

Mid-late infection (waves 4 and 5) was characterized by an extensive change in lifestyle and nutrition by *V. inaequalis*, as well as the expression of a large number of *EC* genes. More specifically, in wave 4, there was sudden increase in the number and expression of GH-encoding genes associated with the degradation of the plant cell wall, such as the pectin-degrading GH28s. Nutrients in the cuticle are likely scarce throughout host colonization; therefore, we suggest that *V. inaequalis* obtains nutrients by degradation of the pectin-rich layer between the cuticle and epidermal cells using these enzymes [84]. Another interesting PCWDE is a GH10 endoxylanase, which is encoded by a gene that is highly expressed during this mid-late infection stage. The importance of endoxylanases in virulence has been reported in some necrotrophic and hemibiotrophic fungi [85-88], and a GH10 endoxylanase from the necrotrophic fungus *Valsa mali* is important for colonization of apple [89]. Notably, the up-regulation of genes encoding PCWDEs during mid-late infection coincided with the up-regulation of genes encoding sugar and nitrogen-associated transporters. As many fungi only utilize nitrates as a source of nitrogen during nitrogen starvation, the up-regulation of genes encoding nitrate transporters at 2 dpi likely reflects the scarcity of nitrogen available in the subcuticular environment [90] and suggests the start of nitrogen starvation for *V. inaequalis*. The up-regulated genes encoding nitrogen transporters are co-expressed with six genes encoding aspartyl proteases; these proteases may degrade extracellular apple proteins to obtain oligopeptides that are then taken up by OPT

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

582 transporters to be used as a nitrogen source. Co-regulation of genes encoding OPT transporters and
583 aspartyl proteases has been reported in *Candida albicans* and *U. maydis* [34, 91].

584 Finally, wave 5, which represents the end of mid-late infection, was enriched for GO terms
585 that could be associated with the end of the biotrophic growth phase by *V. inaequalis*. In line with this,
586 *PCWDE* genes expressed during this late infection stage mainly encoded AA enzymes. It is possible
587 that these AA enzymes facilitate maceration of the plant cell wall at the end of the biotrophic stage,
588 following leaf senescence, as a means of nutrition. A subset of cutinase-encoding genes was also
589 expressed during late infection, which may assist *V. inaequalis* to rupture through the plant cuticle to
590 enable the release of conidiospores from infection structures.

591 Mid-late infection was also characterized by the expression of secondary metabolite genes,
592 especially those that encode RiPPs, a class of cyclic bioactive peptides recently discovered in fungi
593 [92]. Three dikaritin RiPP clusters were up-regulated during host colonization, and all were highly
594 expressed during late infection at 5–7 dpi. RiPPs have been hypothesized to play a role as effectors in
595 promoting host colonization [6]. Notably, the host-selective toxin Victorin from the necrotrophic
596 fungal pathogen *Cochliobolus victoriae* is a RiPP that has been reported to be essential for
597 pathogenicity in oat cultivars resistant to the biotrophic crown rust fungus *Puccinia coronata* [93].
598 Additionally, two putative RiPP precursor genes from the endophytic fungus *Epichloë festucae* have
599 been shown to be highly expressed during the mutualistic interaction with perennial ryegrass. In this
600 case, deletion of these genes did not disrupt the mutualist interaction, suggesting that these RiPPs
601 could have a redundant role or might be involved in antagonistic fungus–microbe interactions [94,
602 95]. The late-expression profile of the *V. inaequalis* RiPPs, together with the finding that many RiPPs
603 from plants have potent antimicrobial activity, may suggest that these peptides promote host
604 colonization through the eradication of microbial competitors in preparation for saprobic growth
605 inside fallen leaf litter [6, 96, 97].

606 Remarkably, the mid-late infection wave also had the largest number of up-regulated genes
607 encoding ECs (~70%), and many of these belonged to expanded protein families. The reason behind

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

608 the large expansion of EC families in *V. inaequalis* is still unknown; however, it has been suggested
609 that the expansion of effectors within families might facilitate diversification of effector function or
610 enable avoidance of recognition by cognate host immune receptors [20]. Expansion of EC families has
611 also been reported in other lineage-specific pathogens, such as *Blumeria graminis* [98, 99]. Given that
612 most EC families of *V. inaequalis* did not have amino acid sequence similarity to ECs from other plant-
613 pathogenic fungi, and because their amino acid sequences did not provide any clues regarding
614 potential functions, we used the *de novo* folding algorithm AlphaFold2 to predict their tertiary
615 structures. One of the main limitations when using AlphaFold2 is that proteins with low numbers of
616 homologous sequences in public databases normally result in predictions with low confidence scores.
617 In an attempt to overcome this limitation, we generated custom multiple sequence alignments
618 (MSAs), which greatly improved prediction scores (**Additional file 23: Fig. S12**). In total, we were able
619 to confidently predict the tertiary structures of ~75% of EC families and 65% of singletons up-regulated
620 *in planta*. Strikingly, many of the EC families had predicted structural similarity to Avr effectors from
621 other plant-pathogenic fungi. This was in addition to ECs from *V. inaequalis* that had amino acid
622 sequence similarity to Avr or candidate Avr effectors from *V. dahliae* (Ave1), *L. maculans* (AvrLm6),
623 and *C. fulvum* (Ecp6 and Ecp10-1). Interestingly, the genes encoding these EC families were
624 predominantly expressed during mid-late infection, suggesting that they may play a key role in the
625 establishment and maintenance of biotrophy.

626 The largest EC family with structural similarity to Avr effectors was the MAX-like effector
627 family, which has undergone drastic expansion in *V. inaequalis* (family 1, 75 members). In *M. oryzae*,
628 Avrs in the MAX effector superfamily have been shown to be translocated into plant cells, where they
629 are recognized by NLR R proteins [52, 55, 100-103]. Of these, Avr-PikD and Avr1-CO39 directly interact
630 with their corresponding NLR R proteins. This is mediated through an interaction with a heavy metal-
631 associated (HMA) domain that is integrated into the R protein itself, albeit through different
632 interactions. Remarkably, it has recently been shown that the MAX effector Avr-Pik binds and
633 stabilizes an HMA protein from rice to modulate host immunity [104]. Altogether, the MAX fold could

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

be well suited to interactions with HMA domains [105, 106]. It is therefore tempting to speculate that members of the MAX-like effector family from *V. inaequalis* are translocated into the plant cell, where they interact with HMA domain-containing proteins of apple involved in plant defence. In terms of expression, the *V. inaequalis MAX-like* genes are expressed at a later infection time point than the *M. oryzae MAX* effectors [53], which are expressed during early infection [53]. This difference in expression could be associated with the extended biotrophic phase of *V. inaequalis*, when compared to *M. oryzae*, which is an intracellular hemibiotrophic fungus.

Also identified in *V. inaequalis* with structural similarity to Avr and EC proteins from other plant-pathogenic fungi was the ToxA-like family. In total, three EC families and one singleton were predicted to adopt a ToxA-like fold in *V. inaequalis*. The ToxA-like structural superfamily includes the necrotrophic effector ToxA from *P. tritici-repentis* [58, 107], as well as the Avr effectors Avr2 from *F. oxysporum* [59, 108] and AvrL567 (AvrLm567-A and AvrLm567-D) [109, 110] from *M. lini*. Although all members of the ToxA-like structural family described above have been shown to be translocated into plant cells [108, 111-113], it remains to be determined if this is the case for members of the ToxA-like superfamily from *V. inaequalis*. Certainly, as R proteins from apple active against *V. inaequalis* are known to be either a PRR (Rvi6) or an NLR (Rvi15) [114, 115], it seems likely that a subset of Avrs from this fungus will likely be translocated into plants cells.

A third structural superfamily that was identified in *V. inaequalis* was the LARS-like superfamily [62]. We identified two families that have the general LARS fold, common to the Avr effectors AvrLm4-7 [61], AvrLm5-7 and AvrLm3 from *L. maculans*, as well as the Avr effector candidate Ecp11-1 from *C. fulvum* [62]. As in *V. inaequalis*, the LARS effector genes of *L. maculans* and *C. fulvum* are expressed during biotrophic host colonization [41, 62]. The *V. inaequalis* LARS-like effectors have the same predicted topology as the abovementioned effectors, with a variable number of β-sheets, but are quite different at the amino acid level and do not share the previously described conserved WR(F/L/V)(R/K) motif [62]. Another family in *V. inaequalis*, the Six4-like family, was predicted to adopt

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

659 the two-domain FOLD-like structure, recently identified in *F. oxysporum* [43]. The C-domains of these
660 effectors also have structural similarity to proteins from the ToxA-like structural family [43].
661 Two other folds were found to be abundant among ECs of *V. inaequalis* that were encoded by
662 genes expressed during the mid infection wave. These were the ferredoxin-like/KP6-like and knottin-
663 like folds. Notably, the AvrLm6-like family, including the Avr effector AvrLm6 from *L. maculans*, as well
664 as the candidate effector Zt-KP6-1 (6qpk) from *Z. tritici*, appear to adopt the KP6-like fold [49]. This
665 fold was first described in the KP6 protein from the P6 virus from *U. maydis,* an antifungal toxic protein
666 [47, 48]. Similarly, different ECs from *C. fulvum* and a necrosis-inducing effector from *Cercospora*
667 *beticola* are predicted to adopt this fold [41, 116]. Intriguingly, through structural modelling, the
668 ferredoxin-like fold was also found to be the most abundant fold in the *M. oryzae* secretome, and was
669 predicted for the BAS4 effector protein from this fungus [117]. Altogether, this suggests that this
670 ferredoxin/KP6-like fold may be a widely conserved structural family in different phytopathogens.
671 In terms of ECs that had a predicted knottin-like fold, family 11 of *V. inaequalis*, along with
672 one singleton, were identified. Knottins are small, ultra-stable proteins with at least three disulfide
673 bridges that form an intramolecular knot known to provide stability in hostile conditions such as the
674 plant apoplast [118]. Current evidence suggests that multiple fungal effectors adopt this fold. Indeed,
675 as previously suggested based on NMR [66] and cysteine bond connectivity [65] data, we predicted
676 that Avr9, an Avr effector from *C. fulvum*, adopts a knottin fold. Furthermore, an EC from *Melampsora*
677 *larici-populina*, MLP124266, which is a homolog of AvrPm4, an Avr effector from *M. lini*, was recently
678 shown to adopt this fold [119]. The other knottin-like families identified in our study (families 12, 24
679 and 35) appear to adopt a fold resembling defensins, similar to the VdAMP3 effector of *V. dahliae*,
680 which has antifungal activity and facilitates microsclerotia formation [97]. Whether the ECs from
681 *V. inaequalis* with KP6/ferredoxin-like and knottin-like folds also possess antimicrobial activity remains
682 to be determined, but could be a future focus of research.
683 Taken together, our study on *V. inaequalis*, along with previous studies on effector proteins
684 from *M. oryzae*, *F. oxysporum* and other fungi [43, 53, 62, 117, 120], reinforce the idea that fungal

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

effectors are often sequence-diverse, but share a limited number of structural folds. The presence of common structural folds in effectors without obvious sequence similarity could be the result of diversifying selection, where the effectors have evolved rapidly to a point where almost all amino acid sequence similarity, with the exception of residues involved in the maintenance of the overall structural fold, is lost [53]. Alternatively, the appearance of common folds could be the result of convergent evolution, indicating that similar folds have evolved independently in different fungi [53]. Common protein folds might have arisen in many fungal effector proteins as they provide a stable structural scaffold that is well suited to hostile plant environments and/or for interaction with conserved host targets. Future knowledge about the distribution of these folds among phytopathogens and the functional characterization of these effectors, will shed more light on this intriguing topic.

Conclusions

In conclusion, we have performed the first comprehensive gene expression analysis of a subcuticular pathogen during host plant colonization, which will provide valuable new insights into the molecular mechanisms underpinning subcuticular growth by this largely understudied class of fungi, including *V. inaequalis*. 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 and, in particular, have provided further evidence that many fungal effectors are highly diverse in sequence, but very conserved at the structural level. The identification of ECs from *V. inaequalis* with sequence and/or structural similarity to Avr effectors from other plant-pathogenic fungi is of great interest as, to date, there are no publications reporting the cloning of *Avr* effector genes from this fungus. The identification of these *Avr* effector genes is of paramount importance, as they can provide information on how *V. inaequalis* overcomes or suppresses *R* gene-mediated resistance in apple, and can 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 be possible to engineer their cognate R

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

711 proteins to recognize features common to the structural fold (direct recognition) or to monitor specific
712 host components targeted by multiple members of the Avr family (indirect recognition). Indeed, such
713 an approach could also be applied more globally to provide resistance against multiple plant-
714 pathogenic fungi possessing effectors with shared effector folds or host virulence targets. Certainly,
715 with the recent development of CRISPR-Cas9 technology in *V. inaequalis* [121], the functional
716 characterization of ECs, and in particular those that form part of expanded EC families, is now possible.
717

718 **Materials and methods**

719 ***V. inaequalis* isolate**

720 The race (1) *V. inaequalis* isolate MNH120, also known as ICMP 13258 and Vi1 [20, 122], was used in
721 this study.
722

723 **Growth in culture for RNA sequencing**

724 *V. inaequalis* was grown as a lawn culture from conidia on cellophane membranes (Waugh Rubber
725 Bands, Wellington, New Zealand) [123] overlaying potato dextrose agar (PDA; Difco™, NJ, USA) at 20°C
726 for 7 days under white fluorescent lights (4,300 K) with a 16 h light/8 h dark photoperiod. Four culture
727 plates, representing four independent biological replicates, were then flooded with 1 mL sterile
728 distilled water and the fungal biomass scraped from the membrane surface using a sterile cell
729 spreader. Following this step, fungal suspensions were transferred to independent microcentrifuge
730 tubes, pelleted by centrifugation at 21,000 x g for 1 min, snap frozen in liquid nitrogen, and then
731 ground to a powder in preparation for RNA extraction.
732

733 **Plant infection assays for RNA sequencing and microscopy**

734 Seeds from open-pollinated *M.* x *domestica* cultivar 'Royal Gala' (Hawke's Bay, New Zealand) were
735 germinated at 4°C in moist vermiculite with 100 mg/ml Thiram fungicide (Kiwicare Corporation
736 Limited; Christchurch, New Zealand) for approximately two months in the dark. Germinated seedlings

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were planted in potting mix (Daltons™ 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 were performed on freshly un-furled detached leaves from 4- to 6-week old apple seedlings, as described previously [124], with the exception that 5 µl droplets of conidial suspension (1 x 10^5 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 [125], and then visualised 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* 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 visualisation of RNA on a 0.8% agarose gel and absence of polymerase chain reaction (PCR) amplification products specific to the *actin* gene of apple (primers RE45 [TGACCGAATGAGCAAGGAAATTACT] and RE64 [TACTCAGCTTTGGCAATCCACATC]) [14]. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

763

764 **Gene prediction**

765 The genome sequence and associated gene annotations of *V. inaequalis* isolate MNH120 [20] were
766 downloaded from the Joint Genome Institute MycoCosm portal
767 (https://mycocosm.jgi.doe.gov/Venin1/Venin1.home.html). New genes were predicted to
768 accommodate genes that may have been missed in the initial annotation (summarized in **additional**
769 **file 2: Fig S1**). For this purpose, we used a three-step approach. In the first step, coding sequences
770 (CDSs) from *V. inaequalis* isolate 05/172, which were predicted as part of a previous study by [19],
771 were downloaded from the National Center for Biotechnology Information (NCBI;
772 https://www.ncbi.nlm.nih.gov/nuccore/QFBF00000000.1/) and mapped to the MNH120 genome
773 using GMAP v2021-02-22 [126]. Here, CDSs from 05/172 were used instead of CDSs previously
774 predicted for MNH120 [20], as the former was deemed to contain a higher number of CDSs
775 corresponding to *EC* genes. In the second step, good quality RNA-seq reads (see RNA-seq read analysis
776 section) from one biological replicate of each time point of *V. inaequalis* grown *in planta* and in culture
777 were mapped to the MNH120 genome using HISAT2 v2.2.1 [127, 128], with unmapped reads filtered
778 out using SAMtools v9.2.0 [129]. Then, a genome-guided *de novo* transcriptome assembly was
779 performed with reads mapped to the *V. inaequalis* genome using Trinity v2.12.0 [130]. Likely CDS
780 regions were identified using Transdecoder v5.5.0 (https://github.com/TransDecoder/TransDecoder)
781 in conjunction with a minimum length open frame (ORF) of 50 amino acids. To maximize the
782 identification of ORFs encoding proteins with characterized functional domains, translated ORFs were
783 scanned against the Pfam database. Additionally, to capture as many ORFs encoding ECs as possible,
784 translated ORFs were scanned against a list of putative effector proteins identified by [20],
785 supplemented with an in-house database of putative MNH120 effector proteins (de la Rosa and
786 Mesarich, unpublished). Finally, in the third step, the original gene prediction for *V. inaequalis* isolate
787 MNH120 [20], together with the newly annotated CDSs from steps 1 and 2, were loaded as different
788 tracks in Geneious v9.05 [131], and manual curation was performed to create consensus gene

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

789 predictions. The aim of this pipeline was to predict as many genes as possible. As a consequence, it is
790 likely that many spurious genes were predicted.
791
792 **Annotation of protein functions**
793 Protein functional annotations were predicted using InterProScan v5.51-85.0 in conjunction with the
794 Pfam, HAMAP, MOBIDB, PIRSF, PROSITE and SUPERFAMILY tools [132], while GO class predictions
795 were carried out using Pannzer2 [133]. N-terminal signal peptides were predicted using SignalP v5.0
796 [134] and transmembrane (TM) domains were predicted using TMHMM v2.0 [135]. CAZymes were
797 predicted using dbCAN2 in conjunction with the Hotpep, HMMER and DIAMOND tools [136]. Only
798 CAZymes predicted with at least two of the three tools were retained for further analysis. Putative
799 PCWDEs were manually identified based on their CAZy classification, KEGG description and InterPro
800 annotation. Secondary metabolite clusters were predicted using antiSMASH v6.0.1 for fungi [137-142]
801 and RiPP gene clusters in *V. inaequalis* were manually identified (**Additional file 24: S1 Text**), RiPP
802 clusters were defined by the presence of a DUF3328 domain and precursor gene (i.e. a gene that
803 encodes a precursor peptide with an N-terminal signal peptide, followed by one or more tandem
804 sequence repeats that are often separated by putative kexin protease cleavage sites).
805
806 **Prediction of effector candidates and effector candidate families**
807 Small proteins of ≤400 amino acid residues in length with a predicted N-terminal signal peptide, but
808 without a predicted transmembrane domain (TM) or endoplasmic reticulum (ER) retention motif
809 (HDEL/KDEL), were annotated as ECs. The list of ECs was supplemented with proteins of >400 amino
810 acids in length with a predicted N-terminal signal peptide, but no predicted TM domain or ER retention
811 motif, provided that they were predicted to be an effector using effectorP v3.0 [143]. ECs were
812 grouped into protein families using spectral clustering SCPS 0.9.8 [144]. The identified protein families
813 were then manually curated by eye, taking in to account conservation of the N-terminal signal peptide
814 sequence, cysteine spacing, as well as conserved functional domains identified with InterProScan

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 amino acid sequence similarity to other proteins, a BLASTp analysis using an e-value threshold of 0.05 was performed against the NCBI non-redundant (nr) protein database.

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, a mappability mask was applied to the *V. inaequalis* MNH120 genome to prevent the multimapping of reads to repetitive genomic regions (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 ≥50% low quality (Qscore: 5) bases, were removed. Quality of all reads was then checked using fastQC v0.11.9 [145]. Next, filtered RNA-seq reads from all samples were mapped to the masked MNH120 genome using HISAT2 v2.2.1 [127, 128]. Here, reads were filtered with SAMtools v9.2.0 [129] to keep only those reads that mapped to the fungal genome. Uniquely mapped reads were counted using featureCounts from SubRead package v2.0.0 [146]. Results from all steps of the RNA-seq analysis were aggregated for quality control assessment using MultiQC v1.11 [147].

Differential gene expression and clustering

The count matrix was imported to R and a differential gene expression analysis was performed with DESeq2 package v1.32.0 [148]. Pairwise comparisons from all samples were performed and genes with $\log_2$fold change of 1.5 and a *p* value of 0.01 during at least one *in planta* infection time point, relative to growth in culture, were considered to be significantly differentially expressed. A PCA plot was then generated using the PCA function of the DESeq2 package. Genes that were identified to be up-regulated at one or more *in planta* infection time points were selected for hierarchical clustering. For

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

clustering, expression counts were normalized using the rlog method from the DESeq2 package and scaled. Hierarchical clustering of the genes was performed using the hclust function, method Ward.D2 and Euclidian distance. Visualization of gene expression clusters/waves was performed using ggplots2 v3.3.5 [149], while gene expression heatmaps were generated using Complexheatmap v2.9.1 [150]. To further investigate the co-expression of genes inside a wave, their correlation was calculated using Pearson method.

GO and Pfam term enrichment analysis

GO predictions from Pannzer2 [133] 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 [151], using the total set of genes employed for clustering as background and a 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 [149]. An enrichment test for Pfam domains was performed using a Fisher's exact test, with all genes targeted in the clustering analysis used as background.

Structural modelling of protein tertiary structures

AlphaFold2 [152], in conjunction with the ColabFold notebook [153] and default parameters, was used to predict the protein tertiary structures of EC family members from *V. inaequalis*. Here, only those family members that were encoded by genes up-regulated *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 peptide) was used as input. For the Avr1/Six4-like family, the published pro-domain [43] was also removed. For those ECs that had <30 proteins with amino acid 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. To generate custom MSAs, mature

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

EC family members and similar proteins identified through the BLASTp analysis were aligned using Clustal Omega [154, 155], with alignments subsequently converted to the a3m format using ToolSeq [156, 157]. The only exception was for members of the Avr1/Six4-like family. Here, in an attempt to improve the structural prediction, the *F. oxysporum* Six4 and 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 version 2.0.1 and 2.1.0 [152], with preset casp14, max_template_date: 2020-05-14, using mature protein sequences as input. Again, only those ECs that were encoded by genes up-regulated *in planta*, relative to growth in culture, were used in this analysis. All predicted protein tertiary structures with a pLDDT score of 70 or higher were considered confident predictions. Protein structures with an pLDDT score of 50–60 that also had an intrinsically disordered region predicted with MobiDB-lite [158] or PrDos [159] were also considered confident predictions.

Predicted EC protein tertiary structures were screened against the Research Collaboratory for Structural Bioinformatics (RCSB) PDB database to identify proteins with similar folds using the Dali server [45]. Here, all hits with a Z-score of ≥2 were considered to be similar. Protein tertiary structures were visualized and aligned using PyMol v2.5, in conjunction with the alignment plugin tool CEalign [160]. To further investigate similarities between protein tertiary structures, TM-align [161] was used to calculate root-mean-square deviation (RMSD). Finally, the general fold of confidently predicted protein tertiary structures was investigated using RUPEE [162, 163] against the SCOPe v2.08 database [63, 164]. 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 [118].

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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). The *V. inaequalis* MNH120 gene annotations and associated proteins sequences generated in this study, as well as the output of AlphaFold2 (open source or ColabFold) with the PDB files for the predicted ECs tertiary structures are available at zenodo (10.5281/zenodo.6233645).

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 and Avr3 protein structure files ahead of public release.

References

1. Ristaino JB, Anderson PK, Bebber DP, Brauman KA, Cunniffe NJ, Fedoroff NV, Finegold C, Garrett KA, Gilligan CA, Jones CM *et al*: **The persistent threat of emerging plant disease pandemics to global food security**. *Proceedings of the National Academy of Sciences* 2021, **118**(23):e2022239118.
2. Cook DE, Mesarich CH, Thomma BPHJ: **Understanding plant immunity as a surveillance system to detect invasion**. *Annual Review of Phytopathology* 2015, **53**(1):541-563.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Macho Alberto P, Zipfel C: **Plant PRRs and the activation of innate immune signaling**. *Molecular Cell* 2014, **54**(2):263-272.
4. Lolle S, Stevens D, Coaker G: **Plant NLR-triggered immunity: from receptor activation to downstream signaling**. *Current Opinion in Immunology* 2020, **62**:99-105.
5. Bolton MD: **Primary metabolism and plant defense—Fuel for the fire**. *Molecular Plant-Microbe Interactions* 2009, **22**(5):487-497.
6. Rocafort M, Fudal I, Mesarich CH: **Apoplastic effector proteins of plant-associated fungi and oomycetes**. *Current Opinion in Plant Biology* 2020, **56**:9-19.
7. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R: **Fungal effectors and plant susceptibility**. *Annual Review of Plant Biology* 2015, **66**(1):513-545.
8. Rovenich H, Boshoven JC, Thomma BPHJ: **Filamentous pathogen effector functions: of pathogens, hosts and microbiomes**. *Current Opinion in Plant Biology* 2014, **20**:96-103.
9. Bowen JK, Mesarich CH, Bus VG, Beresford RM, Plummer KM, Templeton MD: ***Venturia inaequalis*: the causal agent of apple scab**. *Molecular Plant Pathology* 2011, **12**(2):105-122.
10. Le Cam B, Parisi L, Arene L: **Evidence of two formae speciales in *Venturia inaequalis*, responsible for apple and pyracantha scab**. *Phytopathology* 2002, **92**(3):314-320.
11. Le Cam B, Sargent D, Gouzy J, Amselem J, Bellanger M-N, Bouchez O, Brown S, Caffier V, De Gracia M, Debuchy R *et al*: **Population genome gequencing of the scab fungal species *Venturia inaequalis*, *Venturia pirina*, *Venturia aucupariae* and *Venturia asperata***. *G3: Genes, Genomes, Genetics* 2019, **9**(8):2405-2414.
12. Khajuria YP, Kaul S, Wani AA, Dhar MK: **Genetics of resistance in apple against *Venturia inaequalis* (Wint.) Cke**. *Tree Genetics & Genomes* 2018, **14**(2):16.
13. Thakur K, Chawla V, Bhatti S, Swarnkar MK, Kaur J, Shankar R, Jha G: ***De novo* transcriptome sequencing and analysis for *Venturia inaequalis*, the devastating apple scab pathogen**. *PLOS ONE* 2013, **8**(1):e53937.
14. Kucheryava N, Bowen JK, Sutherland PW, Conolly JJ, Mesarich CH, Rikkerink EH, Kemen E, Plummer KM, Hahn M, Templeton MD: **Two novel *Venturia inaequalis* genes induced upon morphogenetic differentiation during infection and *in vitro* growth on cellophane**. *Fungal Genetics and Biology* 2008, **45**(10):1329-1339.
15. Nusbaum CJ, Keitt GW: **A cytological study of host-parasite relations of *Venturia inaequalis* on apple leaves**. *Journal of Agricultural Research* 1938, **56**(8):595-618.
16. Shiller J, Van de Wouw AP, Taranto AP, Bowen JK, Dubois D, Robinson A, Deng CH, Plummer KM: **A large family of *AvrLm6-like* genes in the apple and pear scab pathogens, *Venturia inaequalis* and *Venturia pirina***. *Frontiers in Plant Science* 2015, **6**(980).
17. Manktelow D, Beresford R, Batchelor T, Walker J: **Use patterns and economics of fungicides for disease control in New Zealand apples**. In: *International Conference on Integrated Fruit Production: 1995*. 187-192.
18. Patocchi A, Wehrli A, Dubuis PH, Auwerkerken A, Leida C, Cipriani G, Passey T, Staples M, Didelot F, Philion V *et al*: **Ten years of VINQUEST: first insight for breeding new apple cultivars with durable apple scab resistance**. *Plant Disease* 2020, **104**(8):2074-2081.
19. Passey TAJ, Armitage AD, Xu X: **Annotated draft genome sequence of the apple scab pathogen *Venturia inaequalis***. *Microbiology Resource Announcements* 2018, **7**(12):e01062-01018.
20. Deng CH, Plummer KM, Jones DAB, Mesarich CH, Shiller J, Taranto AP, Robinson AJ, Kastner P, Hall NE, Templeton MD *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 Genomics* 2017, **18**(1):339.
21. 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**. *Molecular Plant-Microbe Interactions* 2019, **32**(9):1091-1094.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

22. Bock CH, Chen C, Yu F, Stevenson KL, Wood BW: **Draft genome sequence of *Fusicladium effusum*, cause of pecan scab**. *Standards in Genomic Sciences* 2016, **11**(1):36.
23. Chen C, Bock CH, Wood BW: **Draft genome sequence of *Venturia carpophila*, the causal agent of peach scab**. *Standards in Genomic Sciences* 2017, **12**:68-68.
24. Jaber MY, Bao J, Gao X, Zhang L, He D, Wang X, Wang A, Wang Z, Wang B: **Genome sequence of *Venturia oleaginea*, the causal agent of olive leaf scab**. *Molecular Plant-Microbe Interactions* 2020, **33**(9):1095-1097.
25. Fudal I, Ross S, Gout L, Blaise F, Kuhn ML, Eckert MR, Cattolico L, Bernard-Samain S, Balesdent MH, Rouxel T: **Heterochromatin-like regions as ecological niches for avirulence genes in the *Leptosphaeria maculans* genome: map-based cloning of *AvrLm6***. *Molecular Plant-Microbe Interactions* 2007, **20**(4):459-470.
26. de Jonge R, Peter van Esse H, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, Zhang Z, Usami T, Lievens B, Subbarao KV *et al*: **Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing**. *Proceedings of the National Academy of Sciences* 2012, **109**(13):5110.
27. Caffier V, Le Cam B, Expert P, Tellier M, Devaux M, Giraud M, Chevalier M: **A new scab-like disease on apple caused by the formerly saprotrophic fungus *Venturia asperata***. *Plant Pathology* 2012, **61**(5):915-924.
28. Latham AJR, A. E.: **Development of *Cladosporium caryigenum* in pecan leaves**. *Phytopathology* 1988, **78**(8):1104-1108.
29. Lanza B, Ragnelli AM, Priore M, Aimola P: **Morphological and histochemical investigation of the response of *Olea europaea* leaves to fungal attack by *Spilocaea oleagina***. *Plant Pathology* 2017, **66**(8):1239-1247.
30. Avrova A, Knogge W: ***Rhynchosporium commune*: a persistent threat to barley cultivation**. *Molecular Plant Pathology* 2012, **13**(9):986-997.
31. Jones P, Ayres PG: **Rhynchosporium leaf blotch of barley studied during the subcuticular phase by electron microscopy**. *Physiological Plant Pathology* 1974, **4**(2):229-233.
32. Blechert O, Debener T: **Morphological characterization of the interaction between *Diplocarpon rosae* and various rose species**. *Plant Pathology* 2005, **54**(1):82-90.
33. Zhao H, Han Q, Wang J, Gao X, Xiao C-L, Liu J, Huang L: **Cytology of infection of apple leaves by *Diplocarpon mali***. *European Journal of Plant Pathology* 2013, **136**(1):41-49.
34. Lanver D, Müller AN, Happel P, Schweizer G, Haas FB, Franitza M, Pellegrin C, Reissmann S, Altmüller J, Rensing SA *et al*: **The biotrophic development of *Ustilago maydis* studied by RNA-Seq analysis**. *The Plant Cell* 2018, **30**(2):300.
35. Gervais J, Plissonneau C, Linglin J, Meyer M, Labadie K, Cruaud C, Fudal I, Rouxel T, Balesdent M-H: **Different waves of effector genes with contrasted genomic location are expressed by *Leptosphaeria maculans* during cotyledon and stem colonization of oilseed rape**. *Molecular Plant Pathology* 2017, **18**(8):1113-1126.
36. Bradshaw RE, Guo Y, Sim AD, Kabir MS, Chettri P, Ozturk IK, Hunziker L, Ganley RJ, Cox MP: **Genome-wide gene expression dynamics of the fungal pathogen *Dothistroma septosporum* throughout its infection cycle of the gymnosperm host *Pinus radiata***. *Molecular Plant Pathology* 2016, **17**(2):210-224.
37. Ye Y, Ozaki T, Umemura M, Liu C, Minami A, Oikawa H: **Heterologous production of asperipin-2a: proposal for sequential oxidative macrocyclization by a fungi-specific DUF3328 oxidase**. *Organic & Biomolecular Chemistry* 2019, **17**(1):39-43.
38. 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**. *The Plant Cell* 2002, **14**(9):2107-2119.
39. Kulkarni RD, Kelkar HS, Dean RA: **An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins**. *Trends in Biochemical Sciences* 2003, **28**(3):118-121.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40. Mesarich CH, Schmitz M, Tremouilhac P, McGillivray DJ, Templeton MD, Dingley AJ: **Structure, dynamics and domain organization of the repeat protein Cin1 from the apple scab fungus**. *Biochimica et Biophysica Acta - Proteins and Proteomics* 2012, **1824**(10):1118-1128.
41. Mesarich CH, Ökmen B, Rovenich H, Griffiths SA, Wang C, Karimi Jashni M, Mihajlovski A, Collemare J, Hunziker L, Deng CH *et al*: **Specific hypersensitive response-associated recognition of new apoplastic effectors from *Cladosporium fulvum* in wild tomato**. *Molecular Plant-Microbe Interactions* 2018, **31**(1):145-162.
42. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A *et al*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**(7873):583-589.
43. Yu DS, Outram MA, Smith A, McCombe CL, Khambalkar PB, Rima SA, Sun X, Ma L, Ericsson DJ, Jones DA *et al*: **The structural repertoire of *Fusarium oxysporum* f. sp. *lycopersici* effectors revealed by experimental and computational studies**. *bioRxiv* 2021:2021.2012.2014.472499.
44. Amoozadeh S, Johnston J, Meisrimler C-N: **Exploiting structural modelling tools to explore host-translocated effector proteins**. *International Journal of Molecular Sciences* 2021, **22**(23):12962.
45. Holm L: **Using dali for protein structure comparison**. *Methods in Molecular Biology* 2020, **2112**:29-42.
46. Huber A, Hajdu D, Bratschun-Khan D, Gáspári Z, Varbanov M, Philippot S, Fizil Á, Czajlik A, Kele Z, Sonderegger C *et al*: **New antimicrobial potential and structural properties of PAFB: a cationic, cysteine-rich protein from *Penicillium chrysogenum* Q176**. *Scientific Reports* 2018, **8**(1):1751.
47. Li N, Erman M, Pangborn W, Duax WL, Park C-M, Bruenn J, Ghosh D: **Structure of *Ustilago maydis* Killer Toxin KP6 α-Subunit: A multimeric assembly with a central pore**. *The Journal of Biological Chemistry* 1999, **274**(29):20425-20431.
48. Allen A, Chatt E, Smith TJ: **The atomic structure of the virally encoded antifungal protein, KP6**. *Journal of Molecular Biology* 2013, **425**(3):609-621.
49. Padilla A, Hoh, F., De guillen, K.: **Zt-KP6-1: an effector from *Zymoseptoria tritici***. 2019.
50. Sánchez-Vallet A, Saleem-Batcha R, Kombrink A, Hansen G, Valkenburg D-J, Thomma BPHJ, Mesters JR: **Fungal effector Ecp6 outcompetes host immune receptor for chitin binding through intrachain LysM dimerization**. *eLife* 2013, **2**:e00790.
51. Hoh F, Padilla, A., De Guillen, K.: **New MAX effector from *Magnaporthe oryzae***. 2019.
52. Li W, Wang B, Wu J, Lu G, Hu Y, Zhang X, Zhang Z, Zhao Q, Feng Q, Zhang H *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***. *Molecular Plant-Microbe Interactions* 2009, **22**(4):411-420.
53. 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 Pathogens* 2015, **11**(10):e1005228.
54. Ose T, Oikawa A, Nakamura Y, Maenaka K, Higuchi Y, Satoh Y, Fujiwara S, Demura M, Sone T, Kamiya M: **Solution structure of an avirulence protein, AVR-Pia, from *Magnaporthe oryzae***. *Journal of Biomolecular NMR* 2015, **63**(2):229-235.
55. Zhang X, He D, Zhao Y, Cheng X, Zhao W, Taylor IA, Yang J, Liu J, Peng YL: **A positive-charged patch and stabilized hydrophobic core are essential for avirulence function of AvrPib in the rice blast fungus**. *The Plant Journal* 2018, **96**(1):133-146.
56. De la Concepcion JC, Franceschetti M, Maqbool A, Saitoh H, Terauchi R, Kamoun S, Banfield MJ: **Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen**. *Nature Plants* 2018, **4**(8):576-585.
57. Nyarko A, Singarapu KK, Figueroa M, Manning VA, Pandelova I, Wolpert TJ, Ciuffetti LM, Barbar E: **Solution NMR structures of *Pyrenophora tritici-repentis* ToxB and its inactive**

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

homolog reveal potential determinants of toxin activity. *The Journal of Biological Chemistry* 2014, **289**(37):25946-25956.

58. Sarma GN, Manning VA, Ciuffetti LM, Karplus PA: **Structure of Ptr ToxA: An RGD-containing host-selective toxin from *Pyrenophora tritici-repentis*** *The Plant Cell* 2005, **17**(11):3190-3202.
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 Phytologist* 2017, **216**(3):897-914.
60. Guncar G, Wang C-IA, Forwood JK, Teh T, Catanzariti A-M, Ellis JG, Dodds PN, Kobe B: **The use of Co2+ for crystallization and structure determination, using a conventional monochromatic X-ray source, of flax rust avirulence protein**. *Crystallization Communications* 2007, **63**(Pt 3):209-213.
61. Blondeau K, Blaise F, Graille M, Kale SD, Linglin J, Ollivier B, Labarde A, Lazar N, Daverdin G, Balesdent MH *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**. *The Plant Journal* 2015, **83**(4):610-624.
62. Lazar N, Mesarich CH, Petit-Houdenot Y, Talbi N, Li de la Sierra-Gallay I, Zélie E, Blondeau K, Gracy J, Ollivier B, Blaise F *et al*: **A new family of structurally conserved fungal effectors displays epistatic interactions with plant resistance proteins**. *bioRxiv* 2020:2020.2012.2017.423041.
63. 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 Research* 2014, **42**(D1):D304-D309.
64. Dias RdO, Franco OL: **Cysteine-stabilized αβ defensins: From a common fold to antibacterial activity**. *Peptides* 2015, **72**:64-72.
65. 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):3458-3466.
66. Vervoort J, van den Hooven HW, Berg A, Vossen P, Vogelsang R, Joosten MHAJ, de Wit PJGM: **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 Letters* 1997, **404**(2):153-158.
67. Steyer Joel T, Downes Damien J, Hunter Cameron C, Migeon Pierre A, Todd Richard B, Goldman Gustavo H: **Duplication and functional divergence of branched-chain amino acid biosynthesis genes in *Aspergillus nidulans***. *mBio* 2021, **12**(3):e00768-00721.
68. Que Y, Yue X, Yang N, Xu Z, Tang S, Wang C, Lv W, Xu L, Talbot NJ, Wang Z: **Leucine biosynthesis is required for infection-related morphogenesis and pathogenicity in the rice blast fungus *Magnaporthe oryzae***. *Current Genetics* 2020, **66**(1):155-171.
69. Saint-Macary ME, Barbisan C, Gagey MJ, Frelin O, Beffa R, Lebrun MH, Droux M: **Methionine biosynthesis is essential for infection in the rice blast fungus *Magnaporthe oryzae***. *PLOS ONE* 2015, **10**(4):e0111108.
70. Shao W, Yang Y, Zhang Y, Lv C, Ren W, Chen C: **Involvement of *BcStr2* in methionine biosynthesis, vegetative differentiation, multiple stress tolerance and virulence in *Botrytis cinerea***. *Molecular Plant Pathology* 2016, **17**(3):438-447.
71. Koller W, Parker DM, Becker CM: **Role of cutinase in the penetration of apple leaves by *Venturia inaequalis*** *Phytopathology* 1991, **81**(11):1375-1379.
72. Ohtaki S, Maeda H, Takahashi T, Yamagata Y, Hasegawa F, Gomi K, Nakajima T, Abe K: **Novel hydrophobic surface binding protein, HsbA, produced by *Aspergillus oryzae***. *Applied and Environmental Microbiology* 2006, **72**(4):2407-2413.
73. Takahashi T, Maeda H, Yoneda S, Ohtaki S, Yamagata Y, Hasegawa F, Gomi K, Nakajima T, Abe K: **The fungal hydrophobin RolA recruits polyesterase and laterally moves on hydrophobic surfaces**. *Molecular Microbiology* 2005, **57**(6):1780-1796.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

74. 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**. *Molecular Plant-Microbe Interactions* 1996, **9**(4):233-242.
75. Kimura M, Yamamoto YY, Seki M, Sakurai T, Sato M, Abe T, Yoshida S, Manabe K, Shinozaki K, Matsui M: **Identification of Arabidopsis genes regulated by high light-stress using cDNA microarray**. *Photochemistry and Photobiology* 2003, **77**(2):226-233.
76. Doss RP: **Treatment of pea pods with Bruchin B results in up-regulation of a gene similar to *MtN19***. *Plant Physiology Biochemistry* 2005, **43**(3):225-231.
77. Naya L, Paul S, Valdés-López O, Mendoza-Soto AB, Nova-Franco B, Sosa-Valencia G, Reyes JL, Hernández G: **Regulation of copper homeostasis and biotic interactions by MicroRNA 398b in common bean**. *PLOS ONE* 2014, **9**(1):e84416.
78. Zhang Z-N, Wu Q-Y, Zhang G-Z, Zhu Y-Y, Murphy RW, Liu Z, Zou C-G: **Systematic analyses reveal uniqueness and origin of the CFEM domain in fungi**. *Scientific Reports* 2015, **5**(1):13032.
79. Zhao S, Shang X, Bi W, Yu X, Liu D, Kang Z, Wang X, Wang X: **Genome-wide identification of effector candidates with conserved motifs from the wheat leaf rust fungus *Puccinia triticina***. *Frontiers in Microbiology* 2020, **11**(1188).
80. Zhu W, Wei W, Wu Y, Zhou Y, Peng F, Zhang S, Chen P, Xu X: **BcCFEM1, a CFEM domain-containing protein with putative GPI-anchored site, is involved in pathogenicity, conidial production, and stress tolerance in *Botrytis cinerea***. *Frontiers in Microbiology* 2017, **8**(1807).
81. Wang J-x, Long F, Zhu H, Zhang Y, Wu J-y, Shen S, Dong J-g, Hao Z-m: **Bioinformatic analysis and functional characterization of CFEM proteins in *Setosphaeria turcica***. *Journal of Integrative Agriculture* 2021, **20**(9):2438-2449.
82. Choi W, Dean RA: **The adenylate cyclase gene *MAC1* of *Magnaporthe grisea* controls appressorium formation and other aspects of growth and development**. *The Plant Cell* 1997, **9**(11):1973-1983.
83. Martínez-Cruz J, Romero D, Hierrezuelo J, Thon M, de Vicente A, Pérez-García A: **Effectors with chitinase activity (EWCAs), a family of conserved, secreted fungal chitinases that suppress chitin-triggered immunity**. *The Plant Cell* 2021, **33**(4):1319-1340.
84. Skrzydeł J, Borowska-Wykręt D, Kwiatkowska D: **Structure, assembly and function of cuticle from mechanical perspective with special focus on perianth**. *International Journal of Molecular Sciences* 2021, **22**(8):4160.
85. Rafiei V, Vélëz H, Tzelepis G: **The role of glycoside hydrolases in phytopathogenic fungi and oomycetes virulence**. *International journal of molecular sciences* 2021, **22**(17):9359.
86. Nguyen QB, Itoh K, Van Vu B, Tosa Y, Nakayashiki H: **Simultaneous silencing of endo-β-1,4 xylanase genes reveals their roles in the virulence of *Magnaporthe oryzae***. *Molecular Microbiology* 2011, **81**(4):1008-1019.
87. Noda J, Brito N, González C: **The *Botrytis cinerea* xylanase Xyn11A contributes to virulence with its necrotizing activity, not with its catalytic activity**. *BMC Plant Biology* 2010, **10**:38.
88. Brito N, Espino JJ, González C: **The endo-beta-1,4-xylanase xyn11A is required for virulence in *Botrytis cinerea***. *Molecular Plant-Microbe Interactions* 2006, **19**(1):25-32.
89. Yu C, Li T, Shi X, Saleem M, Li B, Liang W, Wang C: **Deletion of Endo-β-1,4-Xylanase *VmXyl1* impacts the virulence of *Valsa mali* in apple tree**. *Frontiers in Plant Science* 2018, **9**(663).
90. Solomon PS, Tan K-C, Oliver RP: **The nutrient supply of pathogenic fungi; a fertile field for study**. *Molecular Plant Pathology* 2003, **4**(3):203-210.
91. Martínez P, Ljungdahl PO: **Divergence of Stp1 and Stp2 transcription factors in *Candida albicans* places virulence factors required for proper nutrient acquisition under amino acid control**. *Molecular Cell Biology* 2005, **25**(21):9435-9446.
92. Vogt E, Künzler M: **Discovery of novel fungal RiPP biosynthetic pathways and their application for the development of peptide therapeutics**. *Applied Microbiology and Biotechnology* 2019, **103**(14):5567-5581.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

93. Kessler SC, Zhang X, McDonald MC, Gilchrist CLM, Lin Z, Rightmyer A, Solomon PS, Turgeon BG, Chooi Y-H: **Victorin, the host-selective cyclic peptide toxin from the oat pathogen *Cochliobolus victoriae* is ribosomally encoded**. *Proceedings of the National Academy of Sciences* 2020, **117**(39):24243.

94. Hassing B, Winter D, Becker Y, Mesarich CH, Eaton CJ, Scott B: **Analysis of *Epichloë festucae* small secreted proteins in the interaction with *Lolium perenne***. *PLOS ONE* 2019, **14**(2):e0209463.

95. Johnson RD, Lane GA, Koulman A, Cao M, Fraser K, Fleetwood DJ, Voisey CR, Dyer JM, Pratt J, Christensen M *et al*: **A novel family of cyclic oligopeptides derived from ribosomal peptide synthesis of an *in planta*-induced gene, *gigA*, in *Epichloë* endophytes of grasses**. *Fungal Genetics and Biology* 2015, **85**:14-24.

96. Slazak B, Kapusta M, Strömstedt AA, Słomka A, Krychowiak M, Shariatgorji M, Andrén PE, Bohdanowicz J, Kuta E, Göransson U: **How does the sweet violet (*Viola odorata* L.) fight pathogens and pests – cyclotides as a comprehensive plant host defense system**. *Frontiers in Plant Science* 2018, **9**(1296).

97. 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**. *Proceedings of the National Academy of Sciences* 2021, **118**(49):e2110968118.

98. Sacristán S, Vigouroux M, Pedersen C, Skamnioti P, Thordal-Christensen H, Micali C, Brown JKM, Ridout CJ: **Coevolution between a family of parasite virulence effectors and a class of LINE-1 retrotransposons**. *PLOS ONE* 2009, **4**(10):e7463.

99. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, Ver Loren van Themaat E, Brown JK, Butcher SA, Gurr SJ *et al*: **Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism**. *Science* 2010, **330**(6010):1543-1546.

100. Cesari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, Rivas S, Alaux L, Kanzaki H, Okuyama Y *et al*: **The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding**. *The Plant Cell* 2013, **25**(4):1463-1481.

101. Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, Tosa Y, Chuma I, Takano Y, Win J *et al*: **Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*.** *The Plant Cell* 2009, **21**(5):1573-1591.

102. Ashikawa I, Hayashi N, Yamane H, Kanamori H, Wu J, Matsumoto T, Ono K, Yano M: **Two adjacent nucleotide-binding site–leucine-rich repeat class genes are required to confer *Pikm*-specific rice blast resistance**. *Genetics* 2008, **180**(4):2267.

103. 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***. *Scientific Reports* 2015, **5**:11642.

104. Oikawa K, Fujisaki K, Shimizu M, Takeda T, Saitoh H, Hirabuchi A, Hiraka Y, Białas A, Langner T, Kellner R *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* 2020:2020.2012.2001.406389.

105. Guo L, Cesari S, de Guillen K, Chalvon V, Mammri L, Ma M, Meusnier I, Bonnot F, Padilla A, Peng Y-L *et al*: **Specific recognition of two MAX effectors by integrated HMA domains in plant immune receptors involves distinct binding surfaces**. *Proceedings of the National Academy of Sciences* 2018, **115**(45):11637.

106. Franceschetti M, Maqbool A, Jiménez-Dalmaroni Maximiliano J, Pennington Helen G, Kamoun S, Banfield Mark J: **Effectors of filamentous plant pathogens: commonalities amid diversity**. *Microbiology and Molecular Biology Reviews* 2017, **81**(2):e00066-00016.

107. 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**. *The Plant Journal* 2022, **n/a**(n/a).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

108. Ma L, Cornelissen BJC, Takken FLW: **A nuclear localization for Avr2 from *Fusarium oxysporum* is required to activate the tomato resistance protein I-2**. *Frontiers in Plant Science* 2013, **4**:94-94.
109. Dodds PN, Lawrence GJ, Catanzariti A-M, Ayliffe MA, Ellis JG: **The *Melampsora lini AvrL567* avirulence genes are expressed in haustoria and their products are eecognized inside plant cells**. *The Plant Cell* 2004, **16**(3):755-768.
110. Wang C-IA, Gunčar G, Forwood JK, Teh T, Catanzariti A-M, Lawrence GJ, Loughlin FE, Mackay JP, Schirra HJ, Anderson PA *et al*: **Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity**. *The Plant Cell* 2007, **19**(9):2898-2912.
111. Rafiqi M, Gan PH, Ravensdale M, Lawrence GJ, Ellis JG, Jones DA, Hardham AR, Dodds PN: **Internalization of flax rust avirulence proteins into flax and tobacco cells can occur in the absence of the pathogen**. *The Plant Cell* 2010, **22**(6):2017-2032.
112. Manning VA, Ciuffetti LM: **Localization of Ptr ToxA produced by *Pyrenophora tritici-repentis* reveals protein import into wheat mesophyll cells**. *The Plant Cell* 2005, **17**(11):3203-3212.
113. Manning VA, Hamilton SM, Karplus PA, Ciuffetti LM: **The Arg-Gly-Asp-containing, solvent-exposed loop of Ptr ToxA is required for internalization**. *Molecular Plant-Microbe Interactions* 2008, **21**(3):315-325.
114. Belfanti E, Silfverberg-Dilworth E, Tartarini S, Patocchi A, Barbieri M, Zhu J, Vinatzer BA, Gianfranceschi L, Gessler C, Sansavini S: **The *HcrVf2* gene from a wild apple confers scab resistance to a transgenic cultivated variety**. *Proceedings of the National Academy of Sciences* 2004, **101**(3):886-890.
115. Schouten HJ, Brinkhuis J, Burgh AMvd, Schaart JG, Groenwold R, Broggini GAL, Gessler C: **Cloning and functional characterization of the *Rvi15* (*Vr2*) gene for apple scab resistance**. *Tree Genetics & Genomes* 2013, **10**:251-260.
116. Ebert MK, Rangel LI, Spanner RE, Taliadoros D, Wang X, Friesen TL, de Jonge R, Neubauer JD, Secor GA, Thomma BPHJ *et al*: **Identification and characterization of *Cercospora beticola* necrosis-inducing effector CbNip1**. *Molecular Plant Pathology* 2021, **22**(3):301-316.
117. Seong K, Krasileva KV: **Computational structural genomics unravels common folds and novel families in the secretome of fungal phytopathogen *Magnaporthe oryzae***. *Molecular Plant-Microbe Interactions* 2021:MPMI-03-21-0071-R.
118. 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 Research* 2018, **46**(D1):D454-D458.
119. de Guillen K, Lorrain C, Tsan P, Barthe P, Petre B, Saveleva N, Rouhier N, Duplessis S, Padilla A, Hecker A: **Structural genomics applied to the rust fungus *Melampsora larici-populina* reveals two candidate effector proteins adopting cystine knot and NTF2-like protein folds**. *Scientific Reports* 2019, **9**(1):18084.
120. Spanu PD: **Cereal immunity against powdery mildews targets RNase-Like Proteins associated with Haustoria (RALPH) effectors evolved from a common ancestral gene**. *New Phytologist* 2017, **213**(3):969-971.
121. Rocafort M, Arshed S, Hudson D, Sidhu JS, Bowen JK, Plummer KM, Bradshaw RE, Johnson RD, Johnson LJ, Mesarich CH: **CRISPR-Cas9 gene editing and rapid detection of gene-edited mutants using high-resolution melting in the apple scab fungus, *Venturia inaequalis***. *Fungal Biology* 2021, **126**(1):35-46.
122. 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):633-641.
123. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

124. Win J, Greenwood DR, Plummer KM: **Characterisation of a protein from *Venturia inaequalis* that induces necrosis in Malus carrying the *Vm* resistance gene**. *Physiological and Molecular Plant Pathology* 2003, **62**(4):193-202.
125. Bruzzese E, Hasan S: **A whole leaf clearing and staining technique for host specificity studies of rust fungi**. *Plant Pathology* 1983, **32**(3):335-338.
126. Wu TD, Watanabe CK: **GMAP: a genomic mapping and alignment program for mRNA and EST sequences**. *Bioinformatics* 2005, **21**(9):1859-1875.
127. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements**. *Nature Methods* 2015, **12**(4):357-360.
128. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL: **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype**. *Nature Biotechnology* 2019, **37**(8):907-915.
129. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM *et al*: **Twelve years of SAMtools and BCFtools**. *Gigascience* 2021, **10**(2).
130. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome**. *Nature Biotechnology* 2011, **29**(7):644-652.
131. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al*: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data**. *Bioinformatics* 2012, **28**(12):1647-1649.
132. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al*: **InterProScan 5: genome-scale protein function classification**. *Bioinformatics* 2014, **30**(9):1236-1240.
133. Törönen P, Medlar A, Holm L: **PANNZER2: a rapid functional annotation web server**. *Nucleic Acids Research* 2018, **46**(W1):W84-W88.
134. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H: **SignalP 5.0 improves signal peptide predictions using deep neural networks**. *Nature Biotechnology* 2019, **37**(4):420-423.
135. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes**. *Journal of Molecular Biology* 2001, **305**(3):567-580.
136. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y: **dbCAN2: a meta server for automated carbohydrate-active enzyme annotation**. *Nucleic Acids Research* 2018, **46**(W1):W95-W101.
137. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R: **antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences**. *Nucleic Acids Research* 2011, **39**(Web Server issue):W339-346.
138. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T: **antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers**. *Nucleic Acids Research* 2013, **41**(Web Server issue):W204-212.
139. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W *et al*: **antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters**. *Nucleic Acids Research* 2015, **43**(W1):W237-243.
140. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M *et al*: **antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification**. *Nucleic Acids Research* 2017, **45**(W1):W36-w41.
141. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T: **antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline**. *Nucleic Acids Research* 2019, **47**(W1):W81-W87.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

142. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T: **antiSMASH 6.0: improving cluster detection and comparison capabilities**. *Nucleic Acids Research* 2021, **49**(W1):W29-w35.
143. Sperschneider J, Dodds PN: **EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes**. *Molecular Plant-Microbe Interactions* 2022, 35(2): 146-156.
144. Nepusz T, Sasidharan R, Paccanaro A: **SCPS: a fast implementation of a spectral method for detecting protein families on a genome-wide scale**. *BMC Bioinformatics* 2010, **11**(1):120.
145. Wingett SW, Andrews S: **FastQ Screen: A tool for multi-genome mapping and quality control**. *F1000Research* 2018, **7**:1338.
146. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features**. *Bioinformatics* 2014, **30**(7):923-930.
147. 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):3047-3048.
148. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome Biology* 2014, **15**(12):550.
149. Wickham H: **ggplot2: Elegant graphics for data analysis**. *Springer-Verlag New York* 2016.
150. Gu Z, Eils R, Schlesner M: **Complex heatmaps reveal patterns and correlations in multidimensional genomic data**. *Bioinformatics* 2016, **32**(18):2847-2849.
151. Alexa A, Rahnenführer J, Lengauer T: **Improved scoring of functional groups from gene expression data by decorrelating GO graph structure**. *Bioinformatics* 2006, **22**(13):1600-1607.
152. Ruff KM, Pappu RV: **AlphaFold and implications for intrinsically disordered proteins**. *Journal of Molecular Biology* 2021, **433**(20):167208.
153. Mirdita M, Ovchinnikov S, Steinegger M: **ColabFold - Making protein folding accessible to all**. *bioRxiv* 2021:2021.2008.2015.456425.
154. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J *et al*: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega**. *Molecular Systems Biology* 2011, **7**(1):539.
155. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R: **A new bioinformatics analysis tools framework at EMBL–EBI**. *Nucleic Acids Research* 2010, **38**(suppl_2):W695-W699.
156. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V: **A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core**. *Journal of Molecular Biology* 2018, **430**(15):2237-2243.
157. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V: **Protein Sequence Analysis Using the MPI Bioinformatics Toolkit**. *Current Protocol Bioinformatics* 2020, **72**(1):e108.
158. 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):1402-1404.
159. Ishida T, Kinoshita K: **PrDOS: prediction of disordered protein regions from amino acid sequence**. *Nucleic Acids Research* 2007, **35**(Web Server issue):W460-464.
160. Shindyalov IN, Bourne PE: **Protein structure alignment by incremental combinatorial extension (CE) of the optimal path**. *Protein Engineering* 1998, **11**(9):739-747.
161. Zhang Y, Skolnick J: **TM-align: a protein structure alignment algorithm based on the TM-score**. *Nucleic Acids Research* 2005, **33**(7):2302-2309.
162. Ayoub R, Lee Y: **RUPEE: A fast and accurate purely geometric protein structure search**. *PLOS ONE* 2019, **14**(3):e0213712.
163. Ayoub R, Lee Y: **Protein structure search to support the development of protein structure prediction methods**. *Proteins: Structure, Function, and Bioinformatics* 2021, **89**(6):648-658.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.22.482717; this version posted March 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

164. Chandonia J-M, Fox NK, Brenner SE: **SCOPe: classification of large macromolecular structures in the structural classification of proteins—extended database**. *Nucleic Acids Research* 2019, **47**(D1):D475-D481.