The *Venturia inaequalis* effector repertoire is expressed in waves and is dominated by expanded families with predicted structural similarity to avirulence proteins from other plant-pathogenic 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. Consistent with other plant-pathogenic fungi, many of these effectors are expected to be non-enzymatic proteins, some of which can be recognized by corresponding host resistance proteins to activate plant defences, thus acting as avirulence determinants. To develop durable control strategies

against scab, a better understanding of the roles that these effector proteins play in promoting subcuticular growth by *V. inaequalis*, as well as in activating, suppressing or circumventing resistance protein-mediated defences in apple, is required.

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

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

Keywords

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

Background

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

Scab (or black spot), caused by *Venturia inaequalis*, is the most economically important disease of apple (*Malus* x *domestica*) worldwide [9, 10]. Under favourable conditions, this disease can render the fruit unmarketable and cause a yield reduction of up to 70% [10]. During biotrophic host colonization, *V. inaequalis* exclusively colonizes the subcuticular environment without penetrating the underlying plant epidermal cells [9, 11, 12]. It is here that the fungus develops specialized infection

77 structures, known as stromata and runner hyphae [9, 11]. Stromata give rise to asexual conidia, but
78 are also likely required for nutrient acquisition and effector secretion [9, 11, 13]. Runner hyphae, on
79 the other hand, enable the fungus to radiate out from the initial site of host penetration, acting as a
80 base from which additional stromata can be differentiated [11]. At the end of the season, in autumn,
81 *V. inaequalis* switches to saprobic growth inside fallen leaves, where it undergoes sexual reproduction
82 [9].

83 Scab disease is largely controlled through fungicides, which greatly accelerate the
84 development of fungicide resistance in *V. inaequalis* [14, 15]. Scab-resistant apple cultivars, developed
85 through the incorporation of *R* genes, represent a more sustainable disease control option [16].
86 However, races of *V. inaequalis* that can overcome resistance in apple mediated by single *R* genes
87 have been identified [16, 17]. Therefore, it is likely that multiple 'unbroken' *R* genes, perhaps together
88 with quantitative trait loci (QTL), will need to be stacked in each of these cultivars to provide durable
89 disease resistance [16, 17]. For this to be successful, prior knowledge of the molecular mechanisms
90 used by the fungus to overcome *R* gene-mediated resistance, including an understanding of how these
91 mechanisms impact effector function and pathogen fitness, will be required. So far, however, there
92 have been no publications describing the cloning of *Avr* effector genes from *V. inaequalis.* Crucially,
93 the genomes of multiple *V. inaequalis* isolates have been sequenced [18-23] and bioinformatic studies
94 have identified a large catalogue of secreted, non-enzymatic proteinaceous effector candidates (ECs)
95 from which candidate Avr effectors can be identified [19]. Most of these ECs belong to expanded
96 (sequence-related) families [19]; however, the driving force behind the expansion of these families is
97 not yet known. Nevertheless, many *EC* genes are located next to repetitive sequences in the genome
98 of *V. inaequalis* and, therefore, it is possible that transposition is one of the mechanisms driving this
99 expansion [19]. Examples where *EC* genes of *V. inaequalis* are located next to repetitive sequences
100 include members of the *AvrLm6-like* family, which encode proteins with sequence similarity to AvrLm6
101 [13], an Avr effector from the fungal pathogen *Leptosphaeria maculans* (blackleg of canola) [24], as
102 well as members of the *Ave1-like* family [19], which encode proteins with sequence similarity to Ave1,

103 an antimicrobial Avr effector from the fungal pathogen *Verticillium dahliae* (Verticillium wilt disease)
104 [25, 26].

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

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

122

123 **Results**

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

126 To investigate changes in *V. inaequalis* gene expression during host colonization and relative to growth
127 in culture, we set up an infection time course involving detached leaves of susceptible apple cultivar
128 *M.* x *domestica* 'Royal Gala' and compared it to growth of the fungus on cellophane membranes

129 overlying potato dextrose agar (PDA). Here, six *in planta* time points (12 and 24 hours post-inoculation
130 [hpi], as well as 2, 3, 5 and 7 days post-inoculation [dpi]) and one in culture time point (7 dpi) were
131 used.

132 Analysis of leaf material from the infection time course by bright-field microscopy revealed
133 that, at 12 hpi, conidia of *V. inaequalis* had germinated and formed appressoria on the leaf surface
134 (**Fig. 1A**). At 24 hpi, primary hyphae had developed, indicating that colonization of the subcuticular
135 environment was underway. Then, by 2 and 3 dpi, stromata had differentiated from primary hyphae
136 and, in many cases, these stromata had undergone a rapid expansion in size through non-polar
137 division (**Fig. 1A**). Subcuticular runner hyphae had also started to radiate out from stromata (**Fig. 1A**).
138 By 5 dpi, fungal biomass had accumulated extensively in the subcuticular environment and additional
139 stromata had started to develop from runner hyphae (**Fig. 1A**). Often, these stromata had formed
140 conidiophores, from which conidia had developed (**Fig. 1A**). Finally, at 7 dpi, conidia of *V. inaequalis*
141 had started to rupture through the plant cuticle (**Fig. 1A**) and the first macroscopic olive-brown lesions,
142 indicative of heavy sporulation, were apparent. This represented the end of the biotrophic infection
143 stage, as detached apple leaves had started to decay after this time point.

144 Inspection of the RNA-seq data revealed that the percentage of reads mapping to the
145 *V. inaequalis* genome [19] increased as the infection time course progressed (**Fig. 1B**, **Additional file**
146 **1: Table S1**). Furthermore, the biological replicates clustered robustly within time points and a clear
147 distinction between the early and mid-late infection stages, as well as between the *in planta* and in
148 culture growth conditions, was observed (**Fig. 1C**).

149

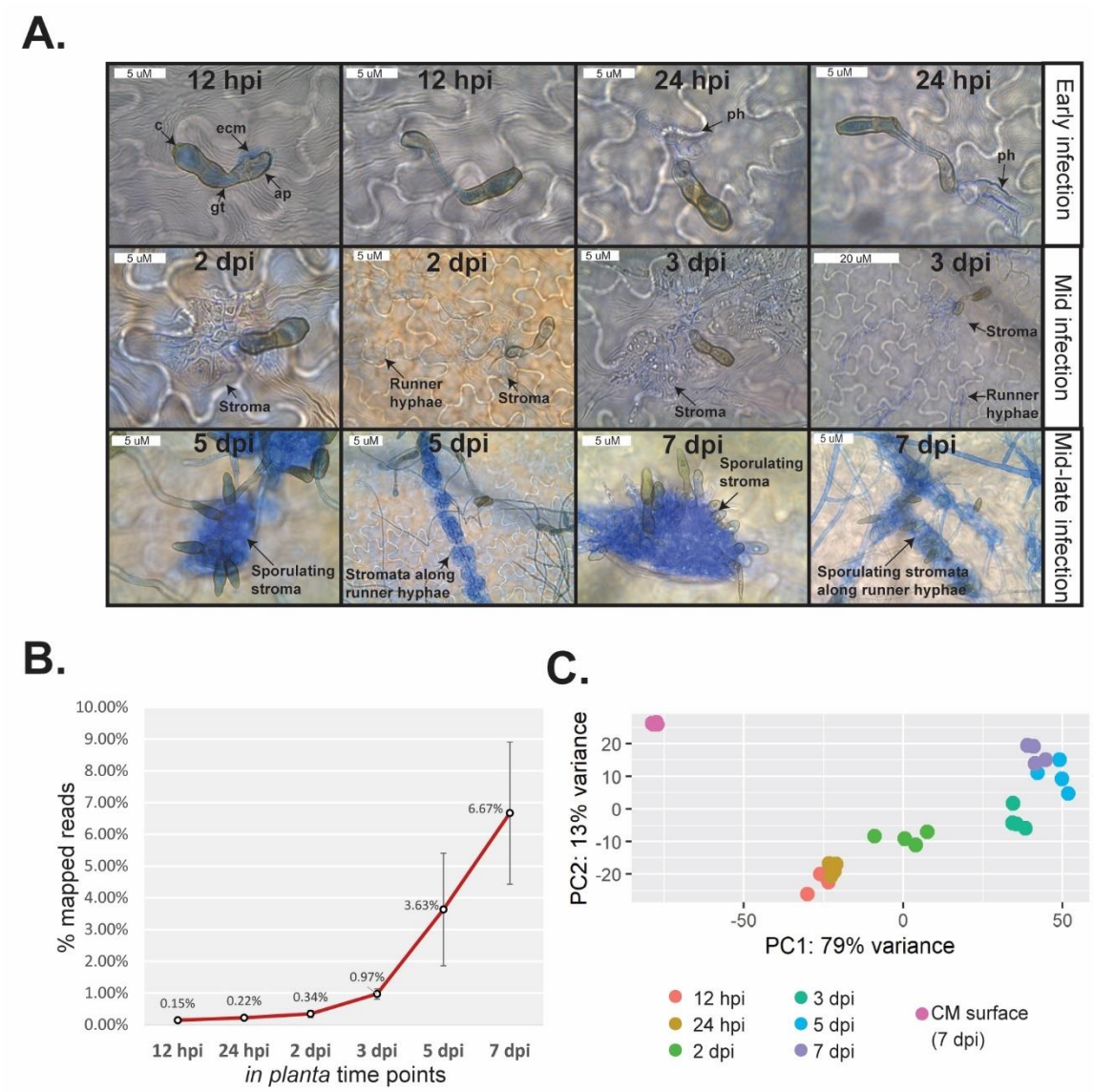


Fig. 1 Microscopic and transcriptomic profile of *Venturia inaequalis* during infection of detached leaves from susceptible apple cultivar 'Royal Gala'. **A.** Microscopic evaluation of *V. inaequalis* during colonization of apple leaves. Infection structures observed by bright-field microscopy were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

161

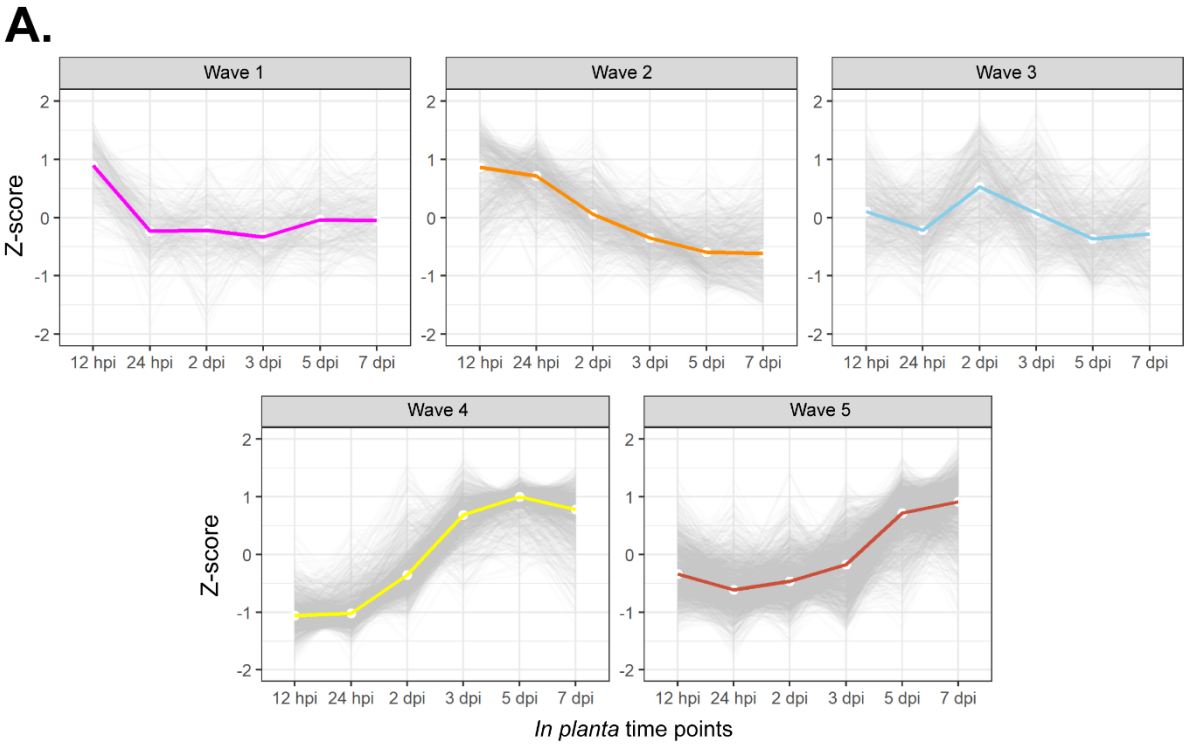
162 **Genes of *V. inaequalis* are expressed in temporal waves during infection of apple leaves**

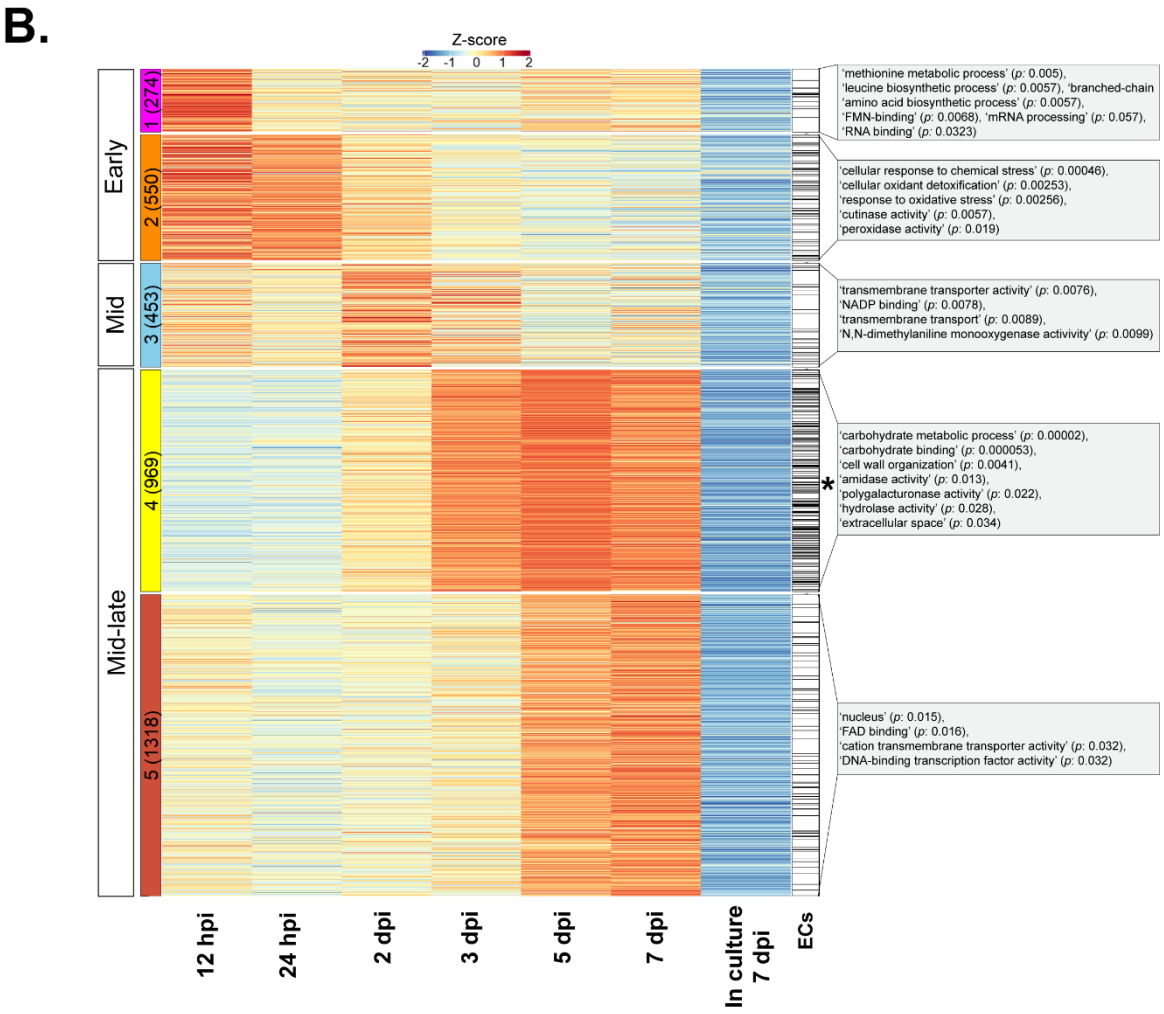
163 We set out to identify which genes of *V. inaequalis* are up-regulated during infection of apple, when
164 compared to growth in culture, as these genes are most likely to be required for promoting host
165 colonization. For this purpose, we updated the current gene catalogue for isolate MNH120 [19] to
166 increase the total number of annotated genes, including those that encode ECs, which are notoriously
167 difficult to predict in fungi (**Additional file 2: Fig. S1**). In total, 24,502 genes, excluding splice variants,
168 were predicted and, of these, 3,563 were up-regulated and 1,462 were down-regulated at one or more
169 *in planta* time points (*p* value of 0.01 and $\log_2$-fold change of 1.5) (**Additional file 3**). It must be pointed
170 out here that our approach was to predict as many genes as possible and, consequently, it is expected
171 that some spurious genes were included in the annotation. However, as many of these spurious genes
172 were anticipated to show a negligible level of expression, most would not have featured in our list of
173 differentially expressed genes, which formed the central focus of our study. For example, of the 24,502
174 predicted genes, 9,284 had a maximum DESeq2-normalized gene expression count across growth
175 conditions of <1, indicating that they were neither expressed under the conditions tested nor up-
176 regulated *in planta*.

177 The total set of *in planta* up-regulated genes was used to identify temporal host infection-
178 specific gene expression clusters, henceforth referred to as waves. Here, all expression data were
179 scaled across all samples (Z-score) to visualize the gene expression deviation from the overall mean.
180 For hierarchical clustering, the parameters were set to specifically identify the minimum number of
181 waves for which a distinct gene expression profile could be observed (**Fig. 2A**). In total, five distinct
182 gene expression waves (**Fig. 2A**), representing three separate infection stages (**Fig. 2B**), were
183 identified. More specifically, genes of waves 1 and 2 peaked in expression during early infection at 12
184 hpi, with expression largely plateauing (wave 1) or trending downwards (wave 2) throughout the
185 remaining infection time points (**Figs. 2A and 2B**). Wave 3 contained genes that peaked in expression
186 during mid infection at 2 dpi (**Figs. 2A and 2B**). Genes of wave 4 displayed their lowest level of

187 expression at 12 and 24 hpi, with expression strongly increasing through 2 and 3 dpi, peaking at 5 dpi
188 during mid-late infection (**Figs. 2A and 2B**). Finally, for genes of wave 5, a similar profile was observed
189 to genes of wave 4, but with expression strongly increasing from 3 dpi and peaking during mid-late
190 infection at 7 dpi (**Figs. 2A and 2B**).

191 To determine which biological processes are overrepresented in the five temporal gene
192 expression waves, gene ontology (GO) (**Fig. 2B**) and protein family (Pfam) enrichment (**Additional file**
193 **4**) analyses were performed. Genes from waves 1 and 2 were mostly characterized by GO terms
194 associated with high metabolic activity, responses to oxidative stress and cutinase activity. Here,
195 cutinases of carbohydrate esterase family 5 (CE5) were abundant (**Additional file 5: Fig. S2**). In
196 contrast, genes of wave 3 were mostly characterized by GO terms associated with transmembrane
197 transport (**Fig. 2B**). Finally, genes of waves 4 and 5 were mostly characterized by GO terms associated
198 with carbohydrate metabolism and transcription (**Fig. 2B**). In wave 4, for example, a GO enrichment
199 for polygalacturonase activity was observed. This, together with the more general enrichment for
200 carbohydrate metabolism, was supported by the high number of plant cell wall-degrading enzyme
201 (PCWDE)-encoding genes in wave 4, most of which were predicted to encode polygalacturonase
202 enzymes of glycoside hydrolase family 28 (GH28) (**Additional file 5: Fig. S2**).
203


A.
Wave 1
Wave 2
Wave 3
Wave 4
Wave 5
Z-score
12 hpi
24 hpi
2 dpi
3 dpi
5 dpi
7 dpi
In planta time points


B.
Z-score
Early
Mid
Mid-late
1 (274)
2 (550)
3 (453)
4 (969)
5 (1318)
12 hpi
24 hpi
2 dpi
3 dpi
5 dpi
7 dpi
In culture 7 dpi
ECs
'methionine metabolic process' (p: 0.005), 'leucine biosynthetic process' (p: 0.0057), 'branched-chain 'amino acid biosynthetic process' (p: 0.0057), 'FMN-binding' (p: 0.0068), 'mRNA processing' (p: 0.057), 'RNA binding' (p: 0.0323)
'cellular response to chemical stress' (p: 0.00046), 'cellular oxidant detoxification' (p: 0.00253), 'response to oxidative stress' (p: 0.00256), 'cutinase activity' (p: 0.0057), 'peroxidase activity' (p: 0.019)
'transmembrane transporter activity' (p: 0.0076), 'NADP binding' (p: 0.0078), 'transmembrane transport' (p: 0.0089), 'N,N-dimethylaniline monooxygenase activivity' (p: 0.0099)
'carbohydrate metabolic process' (p: 0.00002), 'carbohydrate binding' (p: 0.000053), 'cell wall organization' (p: 0.0041), 'amidase activity' (p: 0.013), 'polygalacturonase activity' (p: 0.022), 'hydrolase activity' (p: 0.028), 'extracellular space' (p: 0.034)
*
'nucleus' (p: 0.015), 'FAD binding' (p: 0.016), 'cation transmembrane transporter activity' (p: 0.032), 'DNA-binding transcription factor activity' (p: 0.032)

204

205 **Fig. 2** Genes of *Venturia inaequalis* up-regulated during infection of susceptible apple cultivar 'Royal Gala',
206 relative to growth of the fungus in culture on the surface of cellophane membranes overlying potato dextrose
207 agar, belong to one of five distinct temporal waves of expression. **A.** Expression profile of the five distinct
208 temporal expression waves at 12 and 24 hours post-inoculation (hpi), as well as 2, 3, 5 and 7 days post-
209 inoculation (dpi), relative to growth in culture (7 dpi). **B.** Heatmap of all *V. inaequalis* genes up-regulated *in*
210 *planta* when compared with growth in culture. Gene expression data are scaled rlog-normalized counts across
211 all samples (Z-score), averaged from four biological replicates. Genes up-regulated *in planta* were clustered using
212 hclust according to the Ward.D2 and Euclidean distance methods. Early (12 and 24 hpi), mid (2 and 3 dpi) and
213 mid-late infection (5 and 7 dpi) refer to the stage of infection where peak gene expression was observed.
214 Coloured block labels on the left indicate gene expression waves. Numbers in brackets indicate number of genes
215 per wave. The black asterisk indicates the wave significantly enriched for genes encoding effector candidates
216 (ECs) (*p* value: 3.644e-14). Grey boxes indicate enriched gene ontology (GO) terms. *p*: *p* value.
217

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

220 From the 24,502 predicted genes of *V. inaequalis*, 1,955 genes were predicted to encode a secreted
221 protein without a transmembrane domain and, of these, 1,369 were predicted to encode a non-
222 enzymatic effector candidate (EC). Here, only small, secreted proteins of ≤400 amino acid residues in
223 length, as well as larger secreted proteins with an effector prediction by EffectorP v3.0, were
224 considered to be ECs (**Additional file 6: Fig. S3A**). Of the 1,369 EC proteins, 518 were identical to
225 proteins predicted by Deng et al. [19] (**Additional file 6: Fig. S3B**). Sequence similarities within our
226 predicted set of ECs were investigated using BLASTp, with the ECs then grouped into families using
227 spectral clustering. Based on this analysis, 759 of the ECs were grouped into 118 families ranging in
228 size from two to 75 members. Of these, 32 families were expanded with five or more members. Due
229 to the observed similarity in sequence between family members, and the fact that expanded families
230 had previously been predicted by Deng et al. [19], we do not expect the gene models encoding these
231 ECs to be spurious. In contrast to the 759 ECs mentioned above, 610 of the ECs were predicted to be

232 singletons that did not belong to any family (**Additional file 7**). Interestingly, ~22% of the ECs that
233 belonged to families were encoded by genes that either clustered together or were located in close
234 proximity to each other (i.e. within 10 genes) in the genome of isolate MNH120 (**Additional file 7**).

235 Based on the differential gene expression analysis presented above, 686 of the 1,369
236 predicted *EC* genes from *V. inaequalis* were up-regulated *in planta*. To determine when these genes
237 peaked in expression, their expression profile across the five temporal waves during host colonization
238 was investigated. In total, ~20% of the up-regulated genes peaked in expression during early infection,
239 ~9% during mid infection, and ~71% during mid-late infection (**Fig. 3**). In all cases, families made up
240 the majority of ECs in each wave (**Fig. 3**). Most of the *EC* genes that peaked in expression during early
241 infection (121) encoded proteins that lacked predicted functional domains. Exceptions included: (1) a
242 family that encoded proteins with an 'Egh16-like virulence factor' domain (PF11327) and had
243 sequence similarity to the appressorium-specific Gas1 effector from the rice blast fungus
244 *Magnaporthe oryzae* [38], hereafter named the Gas1-like family, (2) a family that encoded proteins
245 with a 'stress-up regulated Nod19' domain (PF07712), hereafter named the Nod19 family, (3) a family
246 that encoded proteins with a 'hydrophobic surface-binding protein A' (HsbA) domain (PF12296),
247 hereafter named the HsbA family, and (4) a family that encoded proteins with a 'common fold in
248 several fungal extracellular membrane proteins' (CFEM) domain (PF05730) [39], hereafter named the
249 CFEM family.

250 While the *HsbA* and *CFEM* families possessed genes that mostly peaked in expression during
251 early host colonization (**Fig. 3**), each of these families also had a smaller number of genes that peaked
252 in expression during waves 4 and 5 of mid-late infection (**Additional file 8**: **Fig. S4**). Given that *HsbA*
253 and *cutinase* genes have been shown to be regulated by the same transcription factor in *Aspergillus*
254 *nidulans* [40, 41], and that most *cutinase* and *HsbA* genes of *V. inaequalis* peaked in expression during
255 wave 2 of the early infection stage (**Fig. 3, Additional file 5: Fig. S2**), we set out to determine whether
256 the *cutinase* and *HsbA* genes of *V. inaequalis* were co-expressed. Based on the Pearson correlation
257 coefficient, which was calculated between the *cutinases* and *HsbA* gene expression profiles during the

258 early infection stage, the *cutinase* and *HsbA* genes were indeed found to be co-expressed (R > 0.8, $p <$

259 0.01). Another family with members exhibiting different expression profiles during host colonization

260 was the *Cin1* family (**Additional file 8**: **Fig. S4**), which is specific to the *Venturia* genus [42]. This family

261 contains the *Cin1* gene (*g8385*), which encodes a cysteine-rich protein with eight repeats, and two

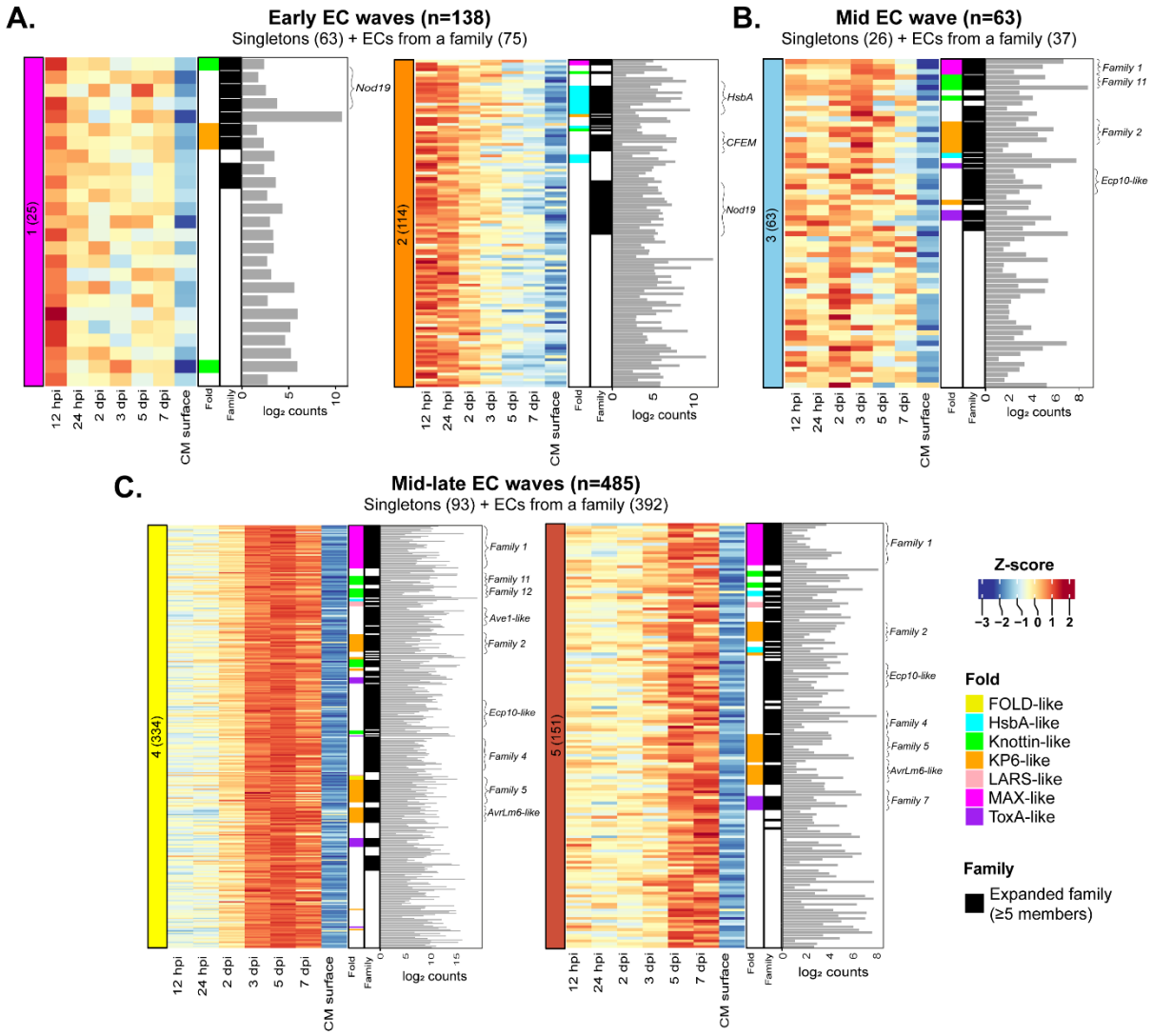
262 *Cin1-like* genes, *Cin1-like 1* (*g10529*) and *Cin1-like 2* (*g13013*), which encode smaller proteins with only

263 one repeat. *Cin1* peaked in expression during wave 4, and was the most highly expressed gene during

264 mid-late host colonization. In contrast, the *Cin1-like* genes peaked in expression during wave 2 of early

265 infection (**Additional file 8**: **Fig. S4**).

266



267

268 **Fig. 3** Genes encoding secreted, non-enzymatic proteinaceous effector candidates (ECs) of *Venturia inaequalis*

269 are expressed in temporal waves during colonization of susceptible apple cultivar 'Royal Gala'. **A.** Heatmap of
270 genes demonstrating peak expression during waves 1 and 2 of the early infection stage at 12 and 24 hours post-
271 inoculation (hpi). **B.** Heatmap of genes demonstrating peak expression during wave 3 of the mid infection stage
272 at 2 and 3 days post-inoculation (dpi). **C.** Heatmap of genes demonstrating peak expression during waves 4 and
273 5 of the mid-late infection stage at 5 and 7 dpi. Heatmap gene expression data are scaled rlog-normalized counts
274 across all samples (Z-score). Bar plots depict the maximum $\log_2$ DESeq2-normalized count value across all *in*
275 *planta* time points. Black annotations indicate genes encoding proteins that belong to an expanded EC family
276 (≥5 members), with brackets highlighting large families and families with sequence similarity to avirulence (Avr)
277 effector proteins from other plant-pathogenic fungi. CM: cellophane membrane.
278
279 The mid and mid-late waves were characterized by 513 genes that mostly encoded EC proteins
280 without a functional domain. The main exception was for members of the Ave1-like family [19], which
281 had an 'RlpA-like domain superfamily annotation (IPR036908). Notably, most of the expanded *EC*
282 families that encoded proteins with sequence similarity to effectors or Avr effectors from other plant-
283 pathogenic fungi, such as the *AvrLm6-like* [13] and *Ave1-like* [19] families, displayed a peak level of
284 expression during waves 4 and 5 (**Fig. 3, Additional file 9: Table S2**). Other examples included the
285 *Ecp10-like* family, which encoded proteins with sequence similarity to the Ecp10-1 Avr candidate from
286 the tomato leaf mold fungus *Fulvia fulva* (formerly *Cladosporium fulvum*), as well as the *Ecp39-like*
287 family, which encoded proteins with sequence similarity to the *F. fulva* EC Ecp39 [43]. An *Ecp6-like*
288 gene (singleton), which encoded a protein with sequence similarity to the Ecp6 effector from *F. fulva*
289 [19, 44], also peaked in expression during wave 4. Interestingly, most expanded *EC* families that
290 peaked in expression during waves 4 and 5 encoded proteins that lacked sequence similarity to other
291 proteins. Examples included the most expanded *EC* family in *V. inaequalis*, family 1 (**Fig. 4**), as well
292 members from family 2 (**Additional file 15: Fig. S7**). In most cases, only one or a few family members
293 were very highly expressed during host colonization (**Fig. 4**).
294 Finally, as it known that some fungal effectors are cyclic ribosomally-synthesized and post-
295 translationally modified peptides (RiPPs) called dikaritins [45], and that dikaritin precursor peptides

296 are often mistaken for standard secretory proteins, we set out to determine whether any of our 1,369
297 ECs were in fact dikaritin precursor peptides. A hallmark of dikaritin precursor peptides is an N-
298 terminal signal peptide followed by multiple perfect or imperfect tandem repeats [46]. As the genes
299 encoding these precursor peptides form part of a biosynthetic gene cluster that also includes a gene
300 encoding a DUF3328 protein, we screened the *V. inaequalis* genome for *DUF3328* genes (and thus,
301 dikaritin gene clusters) to determine which ECs could be dikaritins. In total, nine *dikaritin* gene clusters
302 were identified, with each cluster containing one or more *dikaritin* precursor genes. Based on this
303 analysis, ten of the ECs were identified as putative dikaritin precursor peptides and, of these, four
304 were encoded by genes that peaked in expression during mid-late infection (waves 4 and 5)
305 (**Additional file 10: Fig S5**). The most highly expressed dikaritin precursor gene (*g7830*, from the
306 dikaritin-2 cluster) corresponded to the previously identified gene *Cin3*, which was formerly
307 considered to encode a repeat-containing EC protein [11, 19].
308
309 **Several expanded effector candidate families of *V. inaequalis* have predicted structural similarity to**
310 **Avr effector proteins from other plant-pathogenic fungi**
311 To gain insights into the putative function of ECs, we predicted their tertiary structures using
312 AlphaFold2 [47], and then investigated these structures for similarity to proteins of characterized
313 tertiary structure (and in some cases, function) using the Dali server [48]. This analysis was specifically
314 performed on the most highly expressed member from each EC family (referred to as the
315 representative family member), as well as each singleton, expressed during the temporal host
316 infection-specific waves. In total, the tertiary structure was confidently predicted for the
317 representative family member of 71 (~76%) EC families and 118 (~65%) EC singletons (**Additional file**
318 **11**).
319 Strikingly, many EC families were predicted to be structurally similar to one or more ECs or
320 Avr effector proteins with solved tertiary structures from other plant-pathogenic fungi (**Fig. 4,**
321 **Additional file 12: Table S3**, **Additional file 13: Fig. S6**). More specifically, 12 EC families were

322 predicted to be structurally analogous to ECs or Avr effector proteins (**Additional file 12: Table S3**).
323 Remarkably, many of these families were among the most expanded families in *V. inaequalis*. In
324 contrast to the families, only three EC singletons had predicted structural similarity to ECs or Avr
325 effector proteins from other plant-pathogenic fungi (**Additional file 12: Table S3**).

326 The representative member of the most expanded EC family in *V. inaequalis*, family 1, was
327 confidently predicted to adopt a six-stranded β-sandwich fold with structural similarity to seven
328 '*Magnaporthe* Avr and ToxB-like' (MAX) effectors. The identified MAX effectors were the recently
329 described MAX effector 6R5J [49] and the Avr effectors AvrPiz-t [50], Avr-Pia [51, 52], Avr-Pib [53],
330 Avr1-CO39 [51] and Avr-Pik [54] from *M. oryzae*, as well as the host-selective toxin effector ToxB from
331 *Pyrenophora tritici-repentis* [55], the fungal pathogen responsible for tan spot of wheat (**Additional**
332 **file 12: Table S3**). A sequence alignment constructed from all family 1 members of *V. inaequalis*, as
333 well as the identified MAX effectors from *M. oryzae*, revealed that these proteins lacked significant
334 sequence similarity, with a maximum pairwise identity of only 13.9% observed between g13386 and
335 Avr1-CO39 (**Additional file 14**). However, these sequence-diverse proteins did share the characteristic
336 conserved disulphide bond between β1 and β5 that has previously been reported for MAX effectors
337 [51] (**Fig. 4A**).

338 Similarly, two expanded families, families 7 and 28, as well as family 38 and a singleton, were
339 predicted to share a common β-sandwich fold with structural similarity to the host-selective toxin
340 effector ToxA from *P. tritici-repentis* [56], the Avr effector Avr2/Six3 from the wilt fungus *Fusarium*
341 *oxysporum* [57, 58], and the AvrL567 (AvrL567-A and/or AvrL567-D) Avr effectors from the flax rust
342 fungus *Melampsora lini* [59, 60] (**Fig. 4A, Additional file 12: Table S3, Additional file 13: Fig. S6**).
343 Notably, despite sharing an overall β-sandwich fold with similar topology, these proteins were very
344 diverse at the amino acid level, with a maximum sequence identity of 12.7% (**Additional file 14**). All of
345 the ToxA-like families had a different number of β-sheets (**Additional file 13: Fig. S6**).

346 Another two families, families 15 and 47, were predicted to have structural similarity to the
347 AvrLm4-7 and AvrLm5-9 Avr effectors from *L. maculans*, as well as the Ecp11-1 Avr candidate from *F.*

348 *fulva* [61, 62] (**Fig. 4A, Additional file 12: Table S3, Additional file 13: Fig. S6**), which belong to the
349 '*Leptosphaeria* AviRulence-Supressing (LARS)' structural effector family [62]. The *V. inaequalis* LARS-
350 like effectors had the same predicted topology as AvrLm4-7 but with a variable number of β-sheets.
351 At the amino acid level, the identified *V. inaequalis* LARS-like proteins shared only 7.8–16.9% sequence
352 identity with AvrLm4-7, AvrLm5-9 and Ecp11-1 (**Additional file 14**). Interestingly, the *V. inaequalis*
353 structures were predicted to be stabilized by a different number of disulfide bonds and lacked both
354 the conserved disulfide bridge between the α-helix and β-strand, as well as the conserved
355 WR(F/L/V)(R/K) motif, previously reported for the LARS effectors [62].

356 Finally, family 49 was predicted to adopt a two-domain fold similar to the Avr1/Six4 and
357 Avr3/Six1 Avr effectors from *F. oxysporum* (**Fig. 4A, Additional file 12: Table S3**), which are the
358 founding members of the recently identified *Fol* dual-domain (FOLD) structural family of effectors [63].
359 Despite the *V. inaequalis* FOLD-like family representative sharing only 17.6% amino acid identity with
360 Avr1/Six4 (**Additional file 14**), the cysteine spacing pattern was conserved.

361 Interestingly, many expanded EC families had a KP6-like fold similar to the KP6 protein of the
362 P6 virus from the corn smut fungus *Ustilago maydis* (**Additional file 15: Fig. S7**) [64, 65] and the Zt-
363 KP6-1 EC from wheat blotch fungus *Zymoseptoria tritici* [66]. This included the AvrLm6-like family,
364 families 2, 5, 23, and 26, as well as two singletons (**Additional file 15: Fig. S7**). All KP6-like proteins
365 were predicted to have between two and four disulphide bridges, two α-helices and a variable number
366 of β-sheets (**Fig. 4B, Additional file 15: Fig. S7**).

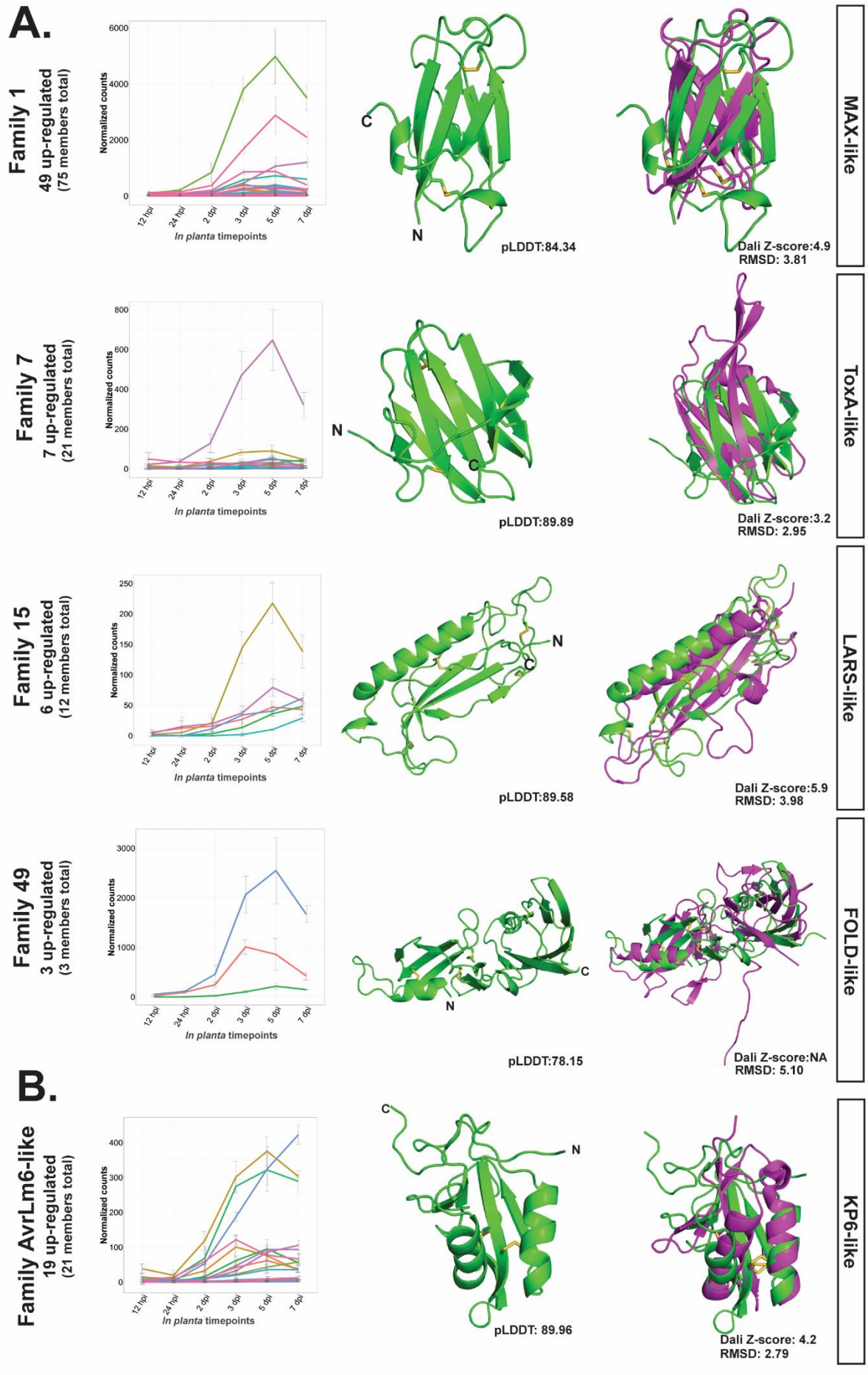
367 Unfortunately, many other EC proteins from *V. inaequalis* were too small to be included in
368 RCSB PDB searches (or did not have a significant hit) using the Dali server. To gain further insights into
369 their function, we investigated their general SCOPe fold classification and identified three families and
370 one singleton that were predicted to adopt a knottin-like fold (**Additional file 11**). The most highly
371 expressed member of family 11 (**Additional file 16: Fig. S8**) and a singleton were predicted to adopt a
372 knottin-like fold with two β-sheets and three disulfide bonds that form an intramolecular knot
373 (**Additional file 16: Fig. S8**). Additionally, families 12, 24, and 35 were predicted to adopt a knottin-

374 like fold using the SCOPe database [67] (**Additional file 16: Fig. S8**). However, these proteins were not
375 predicted to have a true intramolecular knot characteristic of knottin proteins based on our current
376 AlphaFold2 predictions. Instead, these proteins shared a common fold with cysteine-stabilized αβ
377 defensins, which are made up of a single α-helix and three β-strands.

378 In addition to these knottin-like proteins, the Ecp10-like family, along with the candidate Avr
379 effector Ecp10-1 from *F. fulva* [43], were predicted to adopt a small compact β-folded structure that
380 is structurally similar to the PAF protein from *Penicillium chrysogenum* (**Additional file 17: Fig. S9**) [68-
381 70]. Furthermore, the Ecp39-like family, along with the Ecp39 EC from *F. fulva* [43], were predicted to
382 adopt a crambin-like fold (**Additional file 17: Fig. S9**). Unlike that observed for the Ecp10-like family,
383 no analogous protein structures were identified for the Ecp39-like family in the RCSB PDB using the
384 Dali server. Lastly, the Ave1-like family, together with the Ave1 Avr effector from *V. dahliae* [19, 25],
385 were predicted to adopt a double-psi β-barrel fold, stabilized by two conserved disulfide bonds, with
386 high similarity to expansins (**Additional file 17: Fig. S9**).

387 Lastly, we investigated whether a relationship existed between the predicted fold types of EC
388 proteins and the expression profiles of genes that encoded them. Consistent with the results
389 mentioned above, ECs with a predicted HsbA-like fold were predominantly encoded by genes that
390 peaked in expression during wave 2 of early infection (**Fig. 3**). In contrast, ECs with structural similarity
391 to ECs and Avr effector proteins from other plant-pathogenic fungi predominantly peaked in
392 expression during mid and mid-late infection (**Fig. 3**). Remarkably, all members of the FOLD-like and
393 LARS-like families peaked in expression during mid-late infection (wave 4 and wave 5). Similarly, the
394 majority of members from the MAX-like (65%), ToxA-like (70%), KP6-like (~59%) and knottin-like
395 (~66%) families peaked in expression during wave 4.

396


A.
Family 1
49 up-regulated
(75 members total)
Normalized counts
In planta timepoints
12 hpi
24 hpi
2 dpi
3 dpi
5 dpi
7 dpi
C
N
pLDDT:84.34
Dali Z-score:4.9
RMSD: 3.81
MAX-like
Family 7
7 up-regulated
(21 members total)
pLDDT:89.89
Dali Z-score:3.2
RMSD: 2.95
ToxA-like
Family 15
6 up-regulated
(12 members total)
pLDDT:89.58
Dali Z-score:5.9
RMSD: 3.98
LARS-like
Family 49
3 up-regulated
(3 members total)
pLDDT:78.15
Dali Z-score:NA
RMSD: 5.10
FOLD-like
B.
Family AvrLm6-like
19 up-regulated
(21 members total)
pLDDT: 89.96
Dali Z-score: 4.2
RMSD: 2.79
KP6-like

397

398 **Fig. 4** Predicted tertiary structures of secreted non-enzymatic proteinaceous effector candidates (ECs) from
399 *Venturia inaequalis*. **A.** Representative EC family members with structural similarity to avirulence (Avr) effector
400 proteins from other plant-pathogenic fungi. MAX-like: representative predicted family 1 protein structure
401 (green; g13386) aligned to an uncharacterized MAX effector from *Magnaporthe oryzae* (6R5J) (purple). ToxA-
402 like: representative predicted family 7 (green; g4781) protein structure aligned to ToxA from *Pyrenophora tritici-*
403 *repentis* (1ZLE) (purple). LARS-like: representative predicted family 15 (green; g11097) protein structure aligned
404 to AvrLm4-7 from *Leptosphaeria maculans* (7FPR) (purple). FOLD-like: representative predicted family 49 (green;
405 g3787) protein structure aligned to Avr1/Six4 from *Fusarium oxysporum* (7T6A) (purple). **B**. Predicted EC family
406 members with a KP6-like fold. KP6-like: representative predicted AvrLm6-like family (green; g20030) protein
407 structure aligned to EC Zt-KP6-1 from *Zymoseptoria tritici* (6QPK). Expression data of up-regulated *EC* genes
408 during colonization of susceptible apple cultivar 'Royal Gala' are DESeq2-normalized counts, averaged from four
409 biological replicates, with error bars representing standard deviation (hpi: hours post-inoculation; dpi: days post-
410 inoculation). Protein structures predicted by AlphaFold2 represent the most highly expressed member of each
411 EC family from *V. inaequalis*. Disulfide bonds coloured in yellow. N: amino (N) terminus; C: carboxyl (C) terminus.
412 pLDDT: predicted Local Distance Difference Test score (0–100). A pLDDT score of 70–100 is indicative of medium
413 to high confidence. A Dali Z-score above 2 indicates 'significant similarities' between proteins. RMSD: root-mean-
414 square deviation. The protein structure of Avr1/Six4 was not present in the Dali database at the time of writing
415 this manuscript and therefore no Dali Z-score is shown.
416

417 **Discussion**

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

425 temporal gene expression waves were biologically distinct from each other and were enriched for
426 different GO terms.

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

440 Among the *EC* genes that peaked in expression during early colonization, several belonged to
441 the *HsbA* family. HsbA proteins have been suggested to recruit cutinases to hydrophobic surfaces [41]
442 and, in *A. nidulans*, *HsbA* and *cutinase* genes are co-regulated by the same transcription factor [40,
443 41]. Consistent with this previous research, we observed that some *HsbA* and *cutinase* genes of *V.*
444 *inaequalis* were co-expressed, suggesting that, as previously hypothesized, the HsbA proteins of
445 *V. inaequalis* may recruit cutinases to facilitate the degradation and digestion of the hydrophobic
446 apple cuticle during early host infection [19, 41, 71]. In line with this observation, early infection was
447 enriched for the GO term 'cutinase activity' and many CE cutinase-encoding genes were up-regulated.
448 Altogether, these observations support previous reports showing that localized enzymatic hydrolysis
449 is needed for penetration of the apple cuticle by *V. inaequalis* to facilitate access to the subcuticular
450 environment [9, 72].

451 Other *EC* genes that peaked in expression during early host colonization included members of
452 the *Gas1-like* family, which encode proteins with a 'Egh16-like virulence factor' domain, and members
453 of the *Nod19* family, which encode proteins with a 'stress up-regulated Nod19' domain. Recently,
454 Gas1-like proteins have been shown to form part of the widely distributed fungal 'effectors with
455 chitinase activity' (EWCA) family [73]. EWCA proteins are secreted chitinases without a characterized
456 enzymatic domain in the CAZyme database that degrade immunogenic chitin fragments to prevent
457 chitin-triggered immunity in plants [73]. Thus, it is tempting to speculate that members of the Gas1-
458 like family from *V. inaequalis* play a similar role upon access to the subcuticular environment. The
459 function of proteins with a stress up-regulated Nod19 domain is currently unknown. However, it has
460 been suggested that these proteins are associated with responses to abiotic and biotic stress [74-77].
461 Based on these studies and the expression of the *Nod19* genes during early infection, the Nod19 family
462 from *V. inaequalis* could play a role in modulating oxidative stress during early subcuticular
463 colonization of apple. Crucially, other genes associated with oxidative stress tolerance, such as those
464 encoding peroxidases, were also enriched during early infection, suggesting that modulation of
465 oxidative stress is crucial for early host colonization by *V. inaequalis*.
466 Like the *HsbA* family, most members of the *CFEM* family peaked in expression during early
467 host colonization. The CFEM domain is found in several fungal proteins, including those of plant
468 pathogens [78], where it has been shown to confer a diverse range of functions ranging from the
469 promotion or suppression of plant cell death and chlorosis [79-81] to the development of appressoria
470 [82]. Consistent with this functional diversity, and because some members of the CFEM family from
471 *V. inaequalis* instead demonstrated a peak level of expression during mid-late infection, it is likely that
472 the CFEM proteins also play a diverse range of roles in *V. inaequalis* during colonization of apple.
473 The mid infection stage, which was characterized by the large-scale expansion and continued
474 differentiation of subcuticular infection structures (i.e. stromata and runner hyphae), was enriched
475 for GO terms associated with transmembrane transport. Interestingly, while many *EC* genes of *V.*
476 *inaequalis* were highly expressed during this infection stage, very few displayed their peak level of

477 expression here. This may suggest that the expression of most *EC* genes is still increasing during mid
478 infection. However, it is important to point out that the mid infection wave was not well defined,
479 overlapping partially with the mid-late infection waves. This is presumably due to the asynchronous
480 nature of *V. inaequalis* infection.

481 Finally, during the mid-late infection stage, when *V. inaequalis* is heavily colonizing the
482 subcuticular space, most *EC* genes peaked in expression. Based on this expression profile, and the fact
483 that their expression steadily ramped up from the onset of the infection process, we believe that these
484 genes likely play a key role in the establishment and maintenance of biotrophy. Intriguingly, many
485 genes encoding GH enzymes associated with the degradation of the plant cell wall, such as pectin-
486 degrading GH28 proteins, also peaked in expression during this infection stage. As nutrients in the
487 subcuticular environment are likely to be scarce throughout host colonization, it is anticipated that *V.*
488 *inaequalis* meets a portion of its nutritional requirements through the degradation of the pectin-rich
489 layer located between the cuticle and epidermal cells of apple using these enzymes [83]. Related to
490 this, it is well known that fungal GH28 enzymes can be recognized as MAMPs, while host cell wall
491 fragments released as a consequence of GH28 hydrolytic activity can be recognized as damaged-
492 associated molecular patterns (DAMPs), by PRRs, to activate the plant immune system [7]. With this
493 in mind, a subset of the ECs encoded by genes that peaked in expression during mid-late infection may
494 function to suppress plant defences responses initiated by these GH28 enzymes.

495 Remarkably, many of the *EC* genes that peaked in expression during mid-late infection
496 encoded proteins that belonged to expanded families. Such a phenomenon, where ECs are known to
497 form part of expanded families, has also been observed in other lineage-specific pathogens, including
498 *Blumeria graminis* [84, 85]. Although the relevance of EC family expansion to *V. inaequalis* is not yet
499 well understood, it is anticipated that this expansion facilitates the diversification of effector function
500 and enables the avoidance of recognition by cognate host R proteins [19]. In any case, repetitive
501 elements are expected to play a major role in the expansion process, enabling both *EC* gene
502 duplication and subsequent transposition to other regions of the *V. inaequalis* genome [86, 87]. In line

503 with this, it has previously been shown that *ECs* of *V. inaequalis*, including members of the *Ave1-like*
504 and *AvrLm6-like EC* families, tend to be closely associated with repetitive elements [13, 19]. Moreover,
505 *EC* genes belonging to the same expanded families have been shown to cluster together in the *V.*
506 *inaequalis* genome (this study), indicating that tandem duplication events might have occurred. To
507 provide further insights into the process of *EC* family expansion in *V. inaequalis*, a chromosome-level
508 genome assembly of this fungus is now required as, due to its highly fragmented nature (1,012
509 scaffolds) [19], the current Illumina genome provides an incomplete picture of repeat composition
510 and gene clusters.

511 To gain insights into the function of EC proteins from *V. inaequalis*, we used the *de novo*
512 folding algorithm AlphaFold2 to predict their tertiary structures, as AlphaFold2 has been successfully
513 benchmarked against effectors of characterized tertiary structure from other plant-pathogenic fungi
514 [63, 88]. One of the main limitations when using AlphaFold2 is that proteins with a low number of
515 homologous sequences in public databases normally result in predictions with low confidence scores.
516 In an attempt to overcome this, we generated custom multiple sequence alignments (MSAs) that
517 included the amino acid sequences of all EC family members identified in this study, many of which
518 were not available in public sequence databases, which greatly improved prediction scores
519 (**Additional file 18: Fig. S10**).

520 Strikingly, many of the EC families, especially the expanded EC families, demonstrated
521 predicted structural similarity to Avr effector proteins from other plant-pathogenic fungi. The biggest
522 of these was the MAX-like family, which had predicted structural similarity to one of the largest
523 effector/EC families from *M. oryzae*, the MAX family [89, 90]. Intriguingly, a recent computational
524 study of secreted proteins from multiple plant pathogens based on AlphaFold2 concluded that the
525 MAX fold was almost exclusive to *M. oryzae*, with a few members of this family also found in the vanilla
526 black spot fungus *C. orchidophilum* [90]. However, here we show that the MAX-like family has
527 undergone massive expansion and diversification in *V. inaequalis*, highlighting the need for more

528 comprehensive sampling of fungal species using AlphaFold2 to better understand the evolutionary
529 origin and distribution of the MAX structural fold.

530 In *M. oryzae*, Avrs of the MAX effector family are translocated into host cells, where they are
531 recognized by NLR R proteins [50, 53, 91-94]. Of these, Avr-PikD and Avr1-CO39 directly interact with
532 their corresponding NLR R proteins; an interaction mediated through a heavy metal-associated (HMA)
533 domain that is integrated into the R protein itself. Similarly, the MAX effector Avr-Pik binds and
534 stabilizes an independent HMA protein to modulate host immunity [95]. Altogether, these studies
535 suggest that the MAX fold could be well suited to interactions with HMA domains [96, 97]. It is
536 therefore tempting to speculate that ECs of the MAX-like family from *V. inaequalis* are translocated
537 into host cells, where they interact with HMA domain-containing proteins of apple. Certainly, as Rvi15,
538 an R protein of apple that recognizes the AvrRvi15 Avr effector of *V. inaequalis*, is an NLR [98], it seems
539 likely that a subset of Avrs from this fungus are translocated into host cells.

540 EC families and singletons with predicted structural similarity to ECs and Avr proteins from the
541 ToxA-like family were also identified in *V. inaequalis*. Interestingly, like that observed in *V. inaequalis*,
542 the recent computational study by Seong and Krasileva [90] showed that ToxA-like ECs are greatly
543 expanded in the cereal stem rust fungus *Puccinia graminis*. The same study also showed that a further
544 four species, *F. oxysporum*, *C. orchidophilum*, *V. dahliae*, and *U. maydis*, have members of this family
545 [90]. However, in the case of these four species, only a few members of the ToxA family could be
546 identified. Thus, *V. inaequalis* appears to have one of the largest repertoires of ToxA-like ECs in fungal
547 species investigated to date.

548 Like the MAX family, some members of the ToxA-like family (Avr2/Six3 and AvrL567) are
549 translocated into plant cells, where they perform their virulence functions and are recognized by their
550 corresponding R proteins [57, 99]. For example, Avr2/Six3 of *F. oxysporum* functions together with
551 another effector, Six5, to facilitate the cell-to-cell movement of effectors by increasing the size
552 exclusion limit of plasmodesmata [100]. Other members, however, are thought to be apoplastic. For
553 instance, an ortholog of ToxA from the wheat blotch fungus *Parastagonospora nodorum* interacts with

554 an integral membrane protein of wheat from the apoplast to facilitate a cell death reaction that
555 involves the intracellular protein, Tsn1 [101]. As all ToxA-like effectors functionally characterized to
556 date display different virulence functions, no insights into the possible function of ToxA-like proteins
557 from *V. inaequalis* can be made. However, as the Rvi6 R protein of apple, which recognizes the AvrRvi6
558 Avr effector protein of *V. inaequalis*, is an RLP [102], it is possible that these or other ECs of this fungus
559 identified in our study could be recognised as Avr determinants in the subcuticular environment.

560 Two other structural EC families in *V. inaequalis* were the LARS-like family [62] and the two-
561 domain FOLD-like family [63]. Of note, only one small EC family of *V. inaequalis* was predicted to adopt
562 the FOLD-like fold. However, it must be pointed out that proteins with this fold are known to be
563 difficult to predict [90] and, as a consequence, some members may have been missed. Interestingly,
564 while the precise functions of the LARS and FOLD effector families are currently unknown, both
565 contain members that suppress the host immune response related to the recognition of another
566 member. More specifically, in terms of the LARS family, AvrLm4-7 suppresses immune responses
567 triggered by AvrLm3 and AvrLm5-9 [103, 104], while for the FOLD family, Avr1/Six4 suppresses
568 immune responses triggered by Avr3/Six1 [105]. Together, this suggests that these protein structures
569 may play a role in molecular mimicry to prevent detection [62, 63, 106]. It will be interesting to
570 determine whether similar relationships are observed for the LARS-like and FOLD-like ECs of *V.*
571 *inaequalis*.

572 Aside from the folds described above, an intriguing observation was that many of the EC
573 families and singletons from *V. inaequalis* were predicted to adopt a KP6-like or knottin-like fold. The
574 KP6 fold was first described in the antifungal KP6 protein from the P6 virus from *U. maydis* [64, 65]
575 and, in our study, was predicted to be adopted by the AvrLm6-like family as well as AvrLm6 Avr
576 effector protein from *L. maculans*. Notably, this fold is known to be adopted by the EC Zt-KP6-1 (6QPK)
577 from *Z. tritici* [66] and has also been predicted to be adopted by many effectors such as the Avr
578 effector AvrLm10 from *L. maculans* [90, 107], the BAS4 effector from *M. oryzae* [90, 108], the Ecp28
579 EC family and Ecp29 EC from *F. fulva* [43], and the CbNip1 necrosis-inducing effector from the sugar

580 beet leaf spot fungus *Cercospora beticola* [109]. Even more interesting, as observed in *V. inaequalis*,
581 the KP6-like fold has been predicted to be the most abundant fold in the *M. oryzae* secretome [89, 90]
582 and is widely shared among phytopathogens [90]. Altogether, this suggests that this KP6-like fold is a
583 widely conserved structural family of effectors in different phytopathogens.

584 In terms of ECs from *V. inaequalis* that had a predicted knottin-like fold, several families and
585 a singleton were identified. Knottins are small, ultra-stable proteins with at least three disulfide
586 bridges that form an intramolecular knot known to provide stability in hostile conditions such as the
587 plant apoplast [110]. Current evidence suggests that multiple fungal effectors adopt this fold. Indeed,
588 the Avr effector Avr9 from *F. fulva* has previously been suggested to adopt a knottin fold, based on
589 NMR [111] and cysteine bond connectivity [112] data. Furthermore, an EC from the poplar rust fungus
590 *Melampsora larici-populina*, MLP124266, which is a homolog of the AvrPm4 Avr effector from *M. lini*,
591 was recently shown to adopt this fold [113]. Other knottin-like families identified in our study appear
592 to adopt a fold resembling defensins, similar to the VdAMP3 effector of *V. dahliae*, which has
593 antifungal activity and facilitates microsclerotia formation [114].

594 Curiously, the Ecp39-like EC family from *V. inaequalis*, along with the Ecp39 EC from *F. fulva*,
595 were predicted to adopt a crambin fold, which is commonly found in antimicrobial proteins [115],
596 while the Ecp10-like family from *V. inaequalis*, together with the Ecp10-1 Avr candidate from *F. fulva*,
597 were predicted to adopt a fold with similarity to the antimicrobial PAF protein from *P. chrysogenum*
598 [68-70]. The large number of proteins from *V. inaequalis* that are predicted to have structural (KP6-
599 like, knottin-like, Ecp39-like and Ecp10-like) or sequence (Ave1-like) similarity to antimicrobial
600 proteins suggests that a considerable proportion of the ECs from this fungus may be dedicated to
601 antagonistic fungus–microbe interactions to ward off microbial competitors.

602 It should be noted that as part of our EC prediction pipeline, we also identified four putative
603 dikaritin RiPP precursor peptides that were encoded by genes displaying a peak level of expression
604 during late infection. Dikaritins are a class of cyclic bioactive peptides that have recently been
605 discovered in fungi [116] and have been hypothesized to play a role as effectors in promoting host

606 colonization [6]. In line with this, Victorin, a host-selective dikaritin toxin from the necrotrophic oat
607 blight fungus *Cochliobolus victoriae*, has been shown to be essential for pathogenicity in oat cultivars
608 resistant to the biotrophic crown rust fungus *Puccinia coronata* [45]. The late-expression profile of the
609 *V. inaequalis* dikaritins, together with the finding that many RiPPs from plants have potent
610 antimicrobial activity [117], may suggest that these peptides, perhaps in addition to some of the ECs
611 described above, promote host colonization through the eradication of microbial competitors in
612 preparation for saprobic growth inside fallen leaf litter.

613 Taken together, our study on *V. inaequalis*, along with previous studies on effector proteins
614 from *M. oryzae*, *F. oxysporum* and other fungi [51, 62, 63, 89, 90], reinforces the idea that many
615 sequence-diverse fungal effectors share common structural folds. This provides weight to the
616 hypothesis that many fungal effector proteins have in fact originated from ancestral folds and suggests
617 that the genes encoding these effectors have evolved through duplication, followed by sequence
618 diversification, to encode sequence-unrelated but structurally similar proteins. Under this hypothesis,
619 the effector proteins have evolved rapidly to a point where almost all sequence similarity, with the
620 exception of residues involved in the maintenance of the overall structural fold, has been lost [51]. It
621 is of course possible, however, that the appearance of common folds, at least in some instances, could
622 be the result of convergent evolution, whereby certain similar folds have evolved independently in
623 different fungi [51].

624 Specific protein folds may be common across fungal effector proteins as they provide a stable
625 structural scaffold on which surface or loop features can be altered to enable functional diversification
626 [106]. For those Avr effectors that are directly recognized by their corresponding R proteins, it may
627 also be possible that these alterations extend to the evasion of host recognition. Another possibility is
628 that particular structural folds are well suited to certain functions or to interactions with specific host
629 components [106]. Most likely, though, both above-mentioned scenarios are possible [106]. Future
630 research focussing on the finer details of the distribution of structural effector families among both

631 pathogenic and non-pathogenic fungi, and on the functional characterization of members within these
632 families, will shed more light on this intriguing topic.
633
634 **Conclusions**
635 In conclusion, we have performed the first comprehensive gene expression analysis of a subcuticular
636 pathogen, with a specific focus on genes encoding non-enzymatic proteinaceous ECs, during host
637 colonization. In doing so, our study provides valuable new insights into the molecular mechanisms
638 underpinning subcuticular host colonization by this largely understudied class of fungi, including
639 *V. inaequalis*. Notably, in conjunction with structural modelling, we have also provided an enriched
640 list of ECs from which effectors and Avr effectors of *V. inaequalis* can be identified and functionally
641 characterized. Such a resource is desperately needed as, to date, there have been no publications
642 reporting the cloning of *Avr* effector genes from this fungus. Once identified, these *Avr* effector genes
643 will enable the real-time detection of resistance-breaking strains in the orchard. Should Avr effectors
644 of *V. inaequalis* belong to expanded protein families, it may then be possible to engineer their cognate
645 R proteins to recognize features common to the structural fold (direct recognition) or to monitor
646 specific host components targeted by multiple members of the Avr family (indirect recognition).
647 Certainly, with the recent development of CRISPR-Cas9 technology in *V. inaequalis* [118], the
648 functional characterization of ECs, and in particular those that form part of expanded EC families, is
649 now possible. Finally, our study has provided further evidence that many sequence-diverse fungal
650 effectors share common structural folds. Given that the genomes of many other *Venturia* species have
651 now been sequenced [19, 22, 119-125], it will be interesting to determine whether specific effector
652 folds are associated with subcuticular growth or the infection of specific host species.
653
654 **Materials and methods**
655 ***V. inaequalis* isolates**

656 *V. inaequalis* isolate MNH120, also known as ICMP 13258 and Vi1 [19, 126], was used for all bright-
657 field microscopy, transcriptome sequencing, gene predictions and AlphaFold2 protein structure
658 predictions in this study. MNH120 is a race (1) isolate, meaning that it can only overcome resistance
659 mediated by the *Rvi1* gene in apple [19]. This is presumably due to a mutated, absent, or non-
660 expressed copy of the corresponding *AvrRvi1* gene (a functional copy of which is rare among isolates
661 of *V. inaequalis* [127]). MNH120 is, however, anticipated to possess a functional copy of all other
662 known *V. inaequalis Avr* effector genes (*AvrRvi2–20*), corresponding to all other known apple *R* genes
663 (*Rvi2–20*) [19]. Gene predictions from *V. inaequalis* isolate 05/172 [21], which is of unknown race
664 status [128], were used to assist the gene predictions in isolate MNH120.
665
666 **Growth in culture for RNA sequencing**
667 *V. inaequalis* isolate MNH120 was grown as a lawn culture from conidia on cellophane membranes
668 (Waugh Rubber Bands, Wellington, New Zealand) [129] overlaying PDA (Difco™, NJ, USA) at 20°C for
669 7 days under white fluorescent lights (4,300 K) with a 16 h light/8 h dark photoperiod. Four culture
670 plates, representing four independent biological replicates, were then flooded with 1 mL sterile
671 distilled water and the fungal biomass was scraped from the membrane surface using a cell spreader.
672 Following this step, fungal suspensions were transferred to independent microcentrifuge tubes,
673 pelleted by centrifugation at 21,000 x g for 1 min, snap frozen in liquid nitrogen, and then ground to
674 a powder in preparation for RNA extraction. The 7 dpi time point was chosen as *V. inaequalis* had
675 produced enough biomass on the surface of cellophane membranes for adequate RNA extraction.
676
677 **Plant infection assays for RNA sequencing and microscopy**
678 Seeds from open-pollinated *M.* x *domestica* cultivar 'Royal Gala' (Hawke's Bay, New Zealand), which
679 is a cultivar susceptible to scab disease caused by *V. inaequalis*, were germinated at 4°C in moist
680 vermiculite with 100 mg/ml Thiram fungicide (Kiwicare Corporation Limited; Christchurch, New
681 Zealand) for approximately two months in the dark. Germinated seedlings were planted in potting mix

682 (Daltons™ premium potting mix; Daltons, Matamata, New Zealand) and grown under a 16 h light/8 h
683 dark cycle with a Philips SON-T AGRO 400 Sodium lamp, at 20°C with ambient humidity. Inoculations
684 with *V. inaequalis* isolate MNH120 were performed on freshly un-furled detached leaves from 4- to 6-
685 week old apple seedlings, as described previously [130], with the exception that 5 µl droplets of
686 conidial suspension (1 x 10^5 ml^{-1}) were used to cover the entire leaf surface. At 12 and 24 hpi, as well
687 as 2, 3, 5 and 7 dpi, four infected leaves, each from an independent seedling, were sampled to give
688 four biological replicates. A microscopic evaluation of infection was then performed on harvested tips
689 from these leaves. Here, leaf tips were cleared and stained according to a previously established
690 protocol [131] and then visualised by bright-field microscopy, with images captured using a Leica DFC
691 295 digital camera and the Leica Application Suite X (LAS X). Immediately following tip harvesting,
692 leaves were snap frozen in liquid nitrogen and then ground to a powder in preparation for RNA
693 extraction.
694
695 **RNA extraction and sequencing**
696 Total RNA was extracted from samples of *V. inaequalis* isolate MNH120 grown in culture, as well as
697 infected leaves, using a Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA), with DNA
698 subsequently removed using DNase I (Invitrogen™, Thermo Fisher Scientific, MA, USA). RNA
699 concentration and purity were quantified using a Nanodrop ND-1000 Spectrophotometer (NanoDrop
700 Technologies, Rockland, DE, USA), while RNA integrity was assessed on the Agilent 2100 Bioanalyser
701 (Agilent Technologies, Waldbronn, Germany) using an Agilent RNA 6000 Nano Kit in conjunction with
702 Agilent 2100 Bioanalyzer software. Genomic DNA contamination was excluded by visualisation of RNA
703 on a 0.8% agarose gel and absence of polymerase chain reaction (PCR) amplification products specific
704 to the *actin* gene of apple (GenBank Accession OU745002.1 [location 20,713,063–20,714,807];
705 primers RE45 [5´–TGACCGAATGAGCAAGGAAATTACT–3´] and RE64 [5´–
706 TACTCAGCTTTGGCAATCCACATC–3´]) [132]. Following these quality control checks, total RNA from
707 each of the samples was sequenced on a HiSeq X platform at Novogene (Beijing, China) via the Massey

708 Genome Service facility (Palmerston North, New Zealand; project number MGS00286). Here, only
709 those RNA samples with an RNA Integrity Number (RIN) value of ≥3.5 were sequenced.
710
711 **Gene prediction**
712 The genome sequence of *V. inaequalis* isolate MNH120 [19] was downloaded from the Joint Genome
713 Institute (JGI) MycoCosm portal (https://mycocosm.jgi.doe.gov/Venin1/Venin1.home.html) and a
714 new gene catalogue predicted to accommodate those genes that may have been missed in the initial
715 annotation by Deng et al. [19] (summarized in **Additional file 2: Fig. S1**). Here, we combined three
716 sources of information to generate a more complete gene catalogue. More specifically, for the first
717 source of information, the complete set of predicted genes (coding sequences; CDSs) for *V. inaequalis*
718 isolate 05/172 was downloaded from the National Center for Biotechnology Information (NCBI;
719 https://www.ncbi.nlm.nih.gov/nuccore/QFBF00000000.1/) and mapped to the MNH120 genome
720 using GMAP v2021-02-22 [133] to generate a genome annotation file. Isolate 05/172 was used as the
721 source of information for homology-based gene prediction, as it was deemed to have a more complete
722 set of predicted *EC* genes than the previously predicted gene catalogue for isolate MNH120 [19], based
723 on a crude assessment of how many EC family members were present in the non-redundant (nr)
724 protein database using a tBLASTn search. For the second source of information, we used the RNA-seq
725 data generated from isolate MNH120 in this study to predict open reading frames (ORFs), and thus
726 CDSs, from transcript sequences. More specifically, high quality RNA-seq reads (see the RNA-seq read
727 analysis section below) from one biological replicate of each time point of *V. inaequalis* grown *in*
728 *planta* and in culture were mapped to the MNH120 genome using HISAT2 v2.2.1 [134, 135], with
729 unmapped reads filtered out using SAMtools v9.2.0 [136]. Then, a genome-guided *de novo*
730 transcriptome assembly was performed with the mapped reads using Trinity v2.12.0 [137]. Likely CDSs
731 were identified using Transdecoder v5.5.0 (https://github.com/TransDecoder/TransDecoder), a
732 program that predicts genes from transcript sequences, in conjunction with a minimum ORF length of
733 50 amino acids. To support gene predictions, a Pfam domain search was performed to ensure that

734 translated ORFs with one or more characterized functional domains were retained. Additionally, to
735 capture as many EC-encoding ORFs as possible, translated ORFs were scanned against a list of EC
736 proteins identified in the previous gene annotation by Deng et al. [19], supplemented with an in-house
737 database of putative MNH120 effector proteins (de la Rosa and Mesarich, unpublished). Here, a
738 BLASTp e-value threshold of ≤0.05 was employed, with proteins identified using this analysis retained
739 as likely CDSs. Finally, for the third source of information, the original gene catalogue predicted for
740 isolate MNH120 by Deng et al. [19] was downloaded from the JGI MycoCosm portal as above.
741 Together, the three sets of CDSs, derived from the three sources of information described above, were
742 then loaded onto the MHN120 genome as different tracks in Geneious v9.05 [138], and manual
743 curation was performed to create an updated gene catalogue based on a consensus prediction for
744 each gene. More specifically, predictions that were identical across at least two of the three sets of
745 CDSs for a given gene were accepted as correct. Likewise, if for a given gene predictions were different
746 across all CDSs, only the CDS supported by mapped RNA-seq reads (intron–exon boundaries) was
747 selected. Finally, when a CDS was only predicted with one method it was accepted as correct.
748
749 **Prediction of protein functions**
750 Protein functional domains were predicted using InterProScan v5.51-85.0 in conjunction with the
751 Pfam, HAMAP, MOBIDB, PIRSF, PROSITE and SUPERFAMILY tools [139], while GO class predictions
752 were carried out using Pannzer2 [140]. N-terminal signal peptides were predicted using SignalP v5.0
753 [141] and transmembrane (TM) domains were predicted using TMHMM v2.0 [142]. CAZymes were
754 predicted using dbCAN2 in conjunction with the Hotpep, HMMER and DIAMOND tools [143]. Only
755 CAZymes predicted with at least two of the three tools were retained for further analysis. Putative
756 PCWDEs were manually identified based on their CAZy classification (http://www.cazy.org) [144],
757 KEGG description and InterPro annotation. *RiPP* gene clusters were manually identified in the *V.*
758 *inaequalis* MNH120 genome (**Additional file 19: S1 Text**) through the presence of a gene encoding a
759 protein with a DUF3328 domain in close proximity to a gene encoding a dikaritin precursor peptide

760 with an N-terminal signal peptide, followed by one or more perfect or imperfect tandem sequence
761 repeats of at least 10 amino acid residues in length separated by putative kexin protease cleavage
762 sites.
763
764 **Prediction of effector candidates and effector candidate families**
765 Small proteins of ≤400 amino acid residues in length with a predicted N-terminal signal peptide, but
766 without a predicted TM domain or endoplasmic reticulum (ER) retention motif (HDEL/KDEL), were
767 annotated as ECs. This list of ECs was supplemented with proteins of >400 amino acids in length with
768 a predicted N-terminal signal peptide, but no predicted TM domain or ER retention motif, provided
769 that they were predicted to be an effector using EffectorP v3.0 [145]. ECs were grouped into protein
770 families using spectral clustering SCPS v0.9.8 [146]. The identified protein families were then manually
771 curated by eye, taking into account conservation of the N-terminal signal peptide sequence, cysteine
772 spacing, as well as conserved functional domains identified with InterProScan v5.51-85.0. To further
773 refine the list of ECs, proteins with an enzymatic annotation by InterProScan v5.51-85.0 were
774 discarded. In cases where only one or two ECs from a family were predicted to have an enzymatic
775 domain, the EC was retained. To determine whether the ECs had sequence similarity to other proteins,
776 a BLASTp analysis using an e-value threshold of 0.05 was performed against the nr protein database
777 at NCBI.
778
779 **RNA-seq read analysis**
780 Methods associated with this section are summarized in **Additional file 2: Fig. S1**. As a starting point
781 for the analysis of RNA-seq reads, all unique genes of *V. inaequalis* isolate MNH120 were identified,
782 and any additional genes that were identical in sequence, representing paralogs (or false gene
783 duplications generated as a consequence of incorrect genome assembly by Deng et al. [19]), were
784 masked to avoid the multimapping of RNA-seq reads. Here, a gff3 file containing only the duplicated
785 genes was generated using AGAT [147], with the sequences subsequently masked using BEDTools

786 v2.30.0 (maskfasta) [148]. Finally, a mappability mask was applied to the MNH120 genome to prevent
787 the multimapping of reads to repetitive genomic regions, including those genes of *EC* families that are
788 known to be highly similar in sequence. To generate the mappability mask, the MNH120 genome was
789 fragmented into all possible 150-bp stretches and mapped to the MNH120 genome using the Burrows-
790 Wheeler Aligner (bwa) [149] with a gap open penalty of 3, gap extension penalty of 3 and max
791 mismatch penalty of 3. Finally, the resulting SAM file from was used to generate the mappability mask
792 (http://lh3lh3.users.sourceforge.net/download/seqbility-20091110.tar.bz2). Raw RNA-seq reads
793 were then filtered, in which adapter sequences, as well as reads with >10% Ns or ≥50% low quality
794 (Qscore: 5) bases, were removed, and the final quality of reads was checked using fastQC v0.11.9
795 [150]. Next, filtered RNA-seq reads from all samples were mapped to the masked MNH120 genome
796 using HISAT2 v2.2.1 [134, 135] and SAMtools v9.2.0 [136] was used to only keep those reads that
797 mapped to the fungal genome. Uniquely mapped reads were counted using featureCounts from
798 SubRead package v2.0.0 to generate a count matrix [151]. Results from all steps of the RNA-seq
799 analysis were aggregated for quality control assessment using MultiQC v1.11 [152].
800
801 **Differential gene expression and clustering**
802 The count matrix (see the RNA-seq read analysis section) was imported to R and a differential gene
803 expression analysis was performed with DESeq2 package v1.32.0 [153]. Pairwise comparisons from all
804 samples were performed and genes with a $\log_2$fold change in expression of >1.5 and a *p*adj value of
805 <0.01 during at least one *in planta* infection time point, relative to growth in culture, were considered
806 significantly differentially expressed. Here, multiple testing correction was applied using the
807 Benjamini-Hochberg (BH) method of the DESq2 package, in conjunction with a *p*adj value threshold of
808 <0.01. A PCA plot was then generated using the PCA function of the DESeq2 package. Genes that were
809 up-regulated at one or more *in planta* infection time points were selected for hierarchical clustering.
810 For hierarchical clustering, RNA-seq read counts were first normalized using the rlog method from the
811 DESeq2 package and scaled. Hierarchical clustering of the genes was performed using the hclust

812 function according to the Ward.D2 and Euclidean distance methods, with the minimum number of
813 clusters displaying a distinct expression profile during host infection identified in the resulting
814 clustering dendrogram. Using quantitative guidance from the cutree function, the number of clusters
815 was initially set to 10, then systematically reduced due to observed similarity between clusters, to give
816 five distinct clusters with unique expression trends. Visualization of gene expression clusters (waves),
817 with the expression trends plotted, was performed using ggplot2 v3.3.5 [154], while gene expression
818 heatmaps were generated using Complexheatmap v2.9.1 [155]. Pearson correlation coefficients were
819 calculated to investigate the co-expression of genes inside specific clusters.
820
821 **GO and Pfam term enrichment analysis**
822 GO predictions from Pannzer2 [140] with a predictive positive value (PPV) >5 were used for a GO term
823 enrichment analysis across the five distinct gene expression waves. The GO term enrichment analysis
824 was performed with topGO package v2.44.0 [156], using the total set of genes employed for clustering
825 as background and a Fisher's exact test for all GO terms: Biological Process (BP), Cellular Component
826 (CC) and Molecular Function (MF). GO enrichment analysis results were visualized using ggplots2
827 v3.3.5 [154]. An enrichment test for Pfam domains was performed using a Fisher's exact test, with all
828 genes targeted in the clustering analysis used as background.
829
830 **Structural modelling of protein tertiary structures**
831 AlphaFold2 [47], in conjunction with the ColabFold notebook [157], was used to predict the protein
832 tertiary structures of EC family members from *V. inaequalis* isolate MNH120. Here, only those family
833 members that were encoded by genes up-regulated *in planta*, relative to growth in culture, were used
834 in this analysis, with only the most highly expressed member targeted for prediction (i.e. as a family
835 representative). In each case, the mature amino acid sequence of the EC (i.e. without its predicted N-
836 terminal signal peptide) was used as input. For the Avr1/Six4-like family, the published pro-domain
837 [63] was also removed. For those ECs that had <30 proteins with sequence similarity in the NCBI

838 database, as identified by a BLASTp analysis in conjunction with an e-value threshold of ≤0.05, a
839 custom MSA was generated and used as input. These custom MSAs, which were built with up to 100
840 protein sequences, depending on how many protein sequences were available, included all mature EC
841 family members that were unique to the updated MNH120 annotation (i.e. that could not be accessed
842 by the AlphaFold2 search algorithms), as well as similar protein sequences identified through the
843 BLASTp analysis at NCBI. To build the custom MSAs, all sequences were aligned using Clustal Omega
844 [158, 159], with the alignments subsequently converted to the a3m format using ToolSeq [160, 161].
845 The only exception was for members of the FOLD-like family. Here, in an attempt to improve the
846 structural prediction, the *F. oxysporum* Avr1/Six4 and Avr2/Six3 proteins were manually added to the
847 input sequences to generate a custom MSA, even though these proteins were not identified in the
848 initial BLASTp similarity search. For EC singletons, protein tertiary structures were predicted using
849 AlphaFold2 open source code v2.0.1 and v2.1.0 [47], with preset casp14, max_template_date: 2020-
850 05-14, using mature protein sequences as input. Again, only those ECs that were encoded by genes
851 up-regulated *in planta*, relative to growth in culture, were used in this analysis. All predicted protein
852 tertiary structures with a pLDDT score of ≥70 were considered confident predictions. Protein
853 structures with an pLDDT score of 50–60 that also had an intrinsically disordered region predicted with
854 MobiDB-lite [162] or PrDos [163] were also considered confident predictions.
855 Predicted EC protein tertiary structures were screened against the Research Collaboratory for
856 Structural Bioinformatics (RCSB) PDB database to identify proteins with similar folds using the Dali
857 server [48]. Here, all hits with a Z-score of ≥2 were considered similar. Protein tertiary structures were
858 visualized and aligned using PyMol v2.5, in conjunction with the alignment plugin tool CEalign [164].
859 To further investigate similarities between protein tertiary structures, TM-align [165] was used to
860 calculate a root-mean-square deviation (RMSD) value. Finally, the general fold of confidently
861 predicted protein tertiary structures was investigated using RUPEE [166, 167] against the SCOPe v2.08
862 database [67, 168]. Proteins predicted to have a knottin fold in the SCOPe database were assessed
863 using Knotter 3D to determine whether they had a true knottin structure [110].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

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).

Competing interests

The authors declare that they have no competing interests.

Funding

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

Authors' contributions

890 CHM, MR, JKB and KMP conceived the project. JKB, CHM and BM conducted plant infections and
891 prepared samples for sequencing. MR, BH, BM and SdlR performed the bioinformatic analyses. CHM,
892 JKB, MPC and REB provided critical input in experimental design and data analysis. MR, CHM, BH and
893 JKB wrote the manuscript. All authors read, revised, and approved the final manuscript.
894

895 **Acknowledgements**

896 We acknowledge use of the New Zealand eScience Infrastructure (NeSI) high-performance computing
897 facilities, as well as their technical support and training services. In particular, we thank Dinindu
898 Senanayake for his consulting support on the high-throughput prediction of protein tertiary
899 structures. New Zealand's national computing facilities are provided by NeSI and are funded jointly by
900 NeSI's collaborator institutions and through the Ministry of Science & Innovation's Research
901 Infrastructure programme. URL: http://www.nesi.org.nz. We thank Drs Erik Rikkerink and Jay
902 Jayaraman for critically reviewing the manuscript, and Dr Simon Williams for providing the *F.*
903 *oxysporum* Avr1/Six4 and Avr3/Six1 protein structure files ahead of public release.
904

905 **References**

906 1. Ristaino JB, Anderson PK, Bebber DP, Brauman KA, Cunniffe NJ, Fedoroff NV, Finegold C,
907 Garrett KA, Gilligan CA, Jones CM *et al*: **The persistent threat of emerging plant disease**
908 **pandemics to global food security**. *Proceedings of the National Academy of Sciences* 2021,
909 **118**(23):e2022239118.
910 2. Cook DE, Mesarich CH, Thomma BPHJ: **Understanding plant immunity as a surveillance**
911 **system to detect invasion**. *Annual Review of Phytopathology* 2015, **53**(1):541-563.
912 3. Saijo Y, Loo EPi, Yasuda S: **Pattern recognition receptors and signaling in plant–microbe**
913 **interactions**. *The Plant Journal* 2018, **93**(4):592-613.

914 4. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S,
915 Kahmann R: **Fungal effectors and plant susceptibility**. *Annual Review of Plant Biology* 2015,
916 **66**(1):513-545.
917 5. Rovenich H, Boshoven JC, Thomma BPHJ: **Filamentous pathogen effector functions: of**
918 **pathogens, hosts and microbiomes**. *Current Opinion in Plant Biology* 2014, **20**:96-103.
919 6. Rocafort M, Fudal I, Mesarich CH: **Apoplastic effector proteins of plant-associated fungi and**
920 **oomycetes**. *Current Opinion in Plant Biology* 2020, **56**:9-19.
921 7. Bradley EL, Ökmen B, Doehlemann G, Henrissat B, Bradshaw RE, Mesarich CH: **Secreted**
922 **glycoside hydrolase proteins as effectors and invasion patterns of plant-associated fungi and**
923 **oomycetes**. *Frontiers in Plant Science* 2022, **13**:853106.
924 8. Lolle S, Stevens D, Coaker G: **Plant NLR-triggered immunity: from receptor activation to**
925 **downstream signaling**. *Current Opinion in Immunology* 2020, **62**:99-105.
926 9. Bowen JK, Mesarich CH, Bus VG, Beresford RM, Plummer KM, Templeton MD: ***Venturia***
927 ***inaequalis*: the causal agent of apple scab**. *Molecular Plant Pathology* 2011, **12**(2):105-122.
928 10. Jha G, Thakur K, Thakur P: **The *Venturia* apple pathosystem: pathogenicity mechanisms and**
929 **plant defense responses**. *Journal of Biomedicine and Biotechnology* 2009, **2009**:680160.
930 11. Kucheryava N, Bowen JK, Sutherland PW, Conolly JJ, Mesarich CH, Rikkerink EH, Kemen E,
931 Plummer KM, Hahn M, Templeton MD: **Two novel *Venturia inaequalis* genes induced upon**
932 **morphogenetic differentiation during infection and *in vitro* growth on cellophane**. *Fungal*
933 *Genetics and Biology* 2008, **45**(10):1329-1339.
934 12. Nusbaum CJ, Keitt GW: **A cytological study of host-parasite relations of *Venturia inaequalis***
935 **on apple leaves**. *Journal of Agricultural Research* 1938, **56**(8):595-618.
936 13. Shiller J, Van de Wouw AP, Taranto AP, Bowen JK, Dubois D, Robinson A, Deng CH, Plummer
937 KM: **A large family of *AvrLm6-like* genes in the apple and pear scab pathogens, *Venturia***
938 ***inaequalis* and *Venturia pirina***. *Frontiers in Plant Science* 2015, **6**(980).

939 14. Manktelow D, Beresford R, Batchelor T, Walker J: **Use patterns and economics of fungicides**
940 **for disease control in New Zealand apples**. In: *International Conference on Integrated Fruit*
941 *Production: 1995*. 187-192.
942 15. Cox KD: **Fungicide resistance in *Venturia inaequalis*, the causal agent of apple scab, in the**
943 **United States**. In: *Fungicide resistance in plant pathogens: principles and a guide to practical*
944 *management.* Edited by Ishii H, Hollomon DW. Tokyo: Springer Japan; 2015: 433-447.
945 16. Khajuria YP, Kaul S, Wani AA, Dhar MK: **Genetics of resistance in apple against *Venturia***
946 ***inaequalis* (Wint.) Cke**. *Tree Genetics & Genomes* 2018, **14**(2):16.
947 17. Patocchi A, Wehrli A, Dubuis PH, Auwerkerken A, Leida C, Cipriani G, Passey T, Staples M,
948 Didelot F, Philion V *et al*: **Ten years of VINQUEST: first insight for breeding new apple cultivars**
949 **with durable apple scab resistance**. *Plant Disease* 2020, **104**(8):2074-2081.
950 18. Papp D, Singh J, Gadoury D, Khan A: **New North American isolates of *Venturia inaequalis* can**
951 **overcome apple scab resistance of *Malus floribunda* 821**. *Plant Disease* 2020, **104**(3):649-
952 655.
953 19. Deng CH, Plummer KM, Jones DAB, Mesarich CH, Shiller J, Taranto AP, Robinson AJ, Kastner P,
954 Hall NE, Templeton MD *et al*: **Comparative analysis of the predicted secretomes of Rosaceae**
955 **scab pathogens *Venturia inaequalis* and *V. pirina* reveals expanded effector families and**
956 **putative determinants of host range**. *BMC Genomics* 2017, **18**(1):339.
957 20. Passey TAJ, Armitage AD, Sobczyk MK, Shaw MW, Xu X: **Genomic sequencing indicates non-**
958 **random mating of *Venturia inaequalis* in a mixed cultivar orchard**. *Plant Pathology* 2020,
959 **69**(4):669-676.
960 21. Passey TAJ, Armitage AD, Xu X: **Annotated draft genome sequence of the apple scab**
961 **pathogen *Venturia inaequalis***. *Microbiology Resource Announcements* 2018, **7**(12):e01062-
962 01018.
963 22. Le Cam B, Sargent D, Gouzy J, Amselem J, Bellanger M-N, Bouchez O, Brown S, Caffier V, De
964 Gracia M, Debuchy R *et al*: **Population genome gequencing of the scab fungal species**

965 ***Venturia inaequalis*, *Venturia pirina*, *Venturia aucupariae* and *Venturia asperata***. *G3: Genes,*
966 *Genomes, Genetics* 2019, **9**(8):2405-2414.
967 23. Lichtner FJ, Jurick WM, Ayer KM, Gaskins VL, Villani SM, Cox KD: **A genome resource for**
968 **several North American *Venturia inaequalis* isolates with multiple fungicide resistance**
969 **phenotypes**. *Phytopathology* 2020, **110**(3):544-546.
970 24. Fudal I, Ross S, Gout L, Blaise F, Kuhn ML, Eckert MR, Cattolico L, Bernard-Samain S, Balesdent
971 MH, Rouxel T: **Heterochromatin-like regions as ecological niches for avirulence genes in the**
972 ***Leptosphaeria maculans* genome: map-based cloning of *AvrLm6***. *Molecular Plant-Microbe*
973 *Interactions* 2007, **20**(4):459-470.
974 25. de Jonge R, Peter van Esse H, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, Zhang
975 Z, Usami T, Lievens B, Subbarao KV *et al*: **Tomato immune receptor Ve1 recognizes effector**
976 **of multiple fungal pathogens uncovered by genome and RNA sequencing**. *Proceedings of the*
977 *National Academy of Sciences* 2012, **109**(13):5110.
978 26. Snelders NC, Rovenich H, Petti GC, Rocafort M, van den Berg GCM, Vorholt JA, Mesters JR,
979 Seidl MF, Nijland R, Thomma BPHJ: **Microbiome manipulation by a soil-borne fungal plant**
980 **pathogen using effector proteins**. *Nature Plants* 2020, **6**(11):1365-1374.
981 27. Caffier V, Le Cam B, Expert P, Tellier M, Devaux M, Giraud M, Chevalier M: **A new scab-like**
982 **disease on apple caused by the formerly saprotrophic fungus *Venturia asperata***. *Plant*
983 *Pathology* 2012, **61**(5):915-924.
984 28. Latham AJ, Rushing AE: **Development of *Cladosporium caryigenum* in pecan leaves**.
985 *Phytopathology* 1988, **78**(8):1104-1108.
986 29. Lanza B, Ragnelli AM, Priore M, Aimola P: **Morphological and histochemical investigation of**
987 **the response of *Olea europaea* leaves to fungal attack by *Spilocaea oleagina***. *Plant*
988 *Pathology* 2017, **66**(8):1239-1247.
989 30. Avrova A, Knogge W: ***Rhynchosporium commune*: a persistent threat to barley cultivation**.
990 *Molecular Plant Pathology* 2012, **13**(9):986-997.

991 31. Jones P, Ayres PG: ***Rhynchosporium* leaf blotch of barley studied during the subcuticular**
992 **phase by electron microscopy**. *Physiological Plant Pathology* 1974, **4**(2):229-233.
993 32. Blechert O, Debener T: **Morphological characterization of the interaction between**
994 ***Diplocarpon rosae* and various rose species**. *Plant Pathology* 2005, **54**(1):82-90.
995 33. Zhao H, Han Q, Wang J, Gao X, Xiao C-L, Liu J, Huang L: **Cytology of infection of apple leaves**
996 **by *Diplocarpon mali***. *European Journal of Plant Pathology* 2013, **136**(1):41-49.
997 34. Lanver D, Müller AN, Happel P, Schweizer G, Haas FB, Franitza M, Pellegrin C, Reissmann S,
998 Altmüller J, Rensing SA *et al*: **The biotrophic development of *Ustilago maydis* studied by RNA-**
999 **Seq analysis**. *The Plant Cell* 2018, **30**(2):300.
1000 35. Gervais J, Plissonneau C, Linglin J, Meyer M, Labadie K, Cruaud C, Fudal I, Rouxel T, Balesdent
1001 M-H: **Different waves of effector genes with contrasted genomic location are expressed by**
1002 ***Leptosphaeria maculans* during cotyledon and stem colonization of oilseed rape**. *Molecular*
1003 *Plant Pathology* 2017, **18**(8):1113-1126.
1004 36. Bradshaw RE, Guo Y, Sim AD, Kabir MS, Chettri P, Ozturk IK, Hunziker L, Ganley RJ, Cox MP:
1005 **Genome-wide gene expression dynamics of the fungal pathogen *Dothistroma septosporum***
1006 **throughout its infection cycle of the gymnosperm host *Pinus radiata***. *Molecular Plant*
1007 *Pathology* 2016, **17**(2):210-224.
1008 37. Thakur K, Chawla V, Bhatti S, Swarnkar MK, Kaur J, Shankar R, Jha G: ***De novo* transcriptome**
1009 **sequencing and analysis for *Venturia inaequalis*, the devastating apple scab pathogen**. *PLOS*
1010 *ONE* 2013, **8**(1):e53937.
1011 38. Xue C, Park G, Choi W, Zheng L, Dean RA, Xu J-R: **Two novel fungal virulence genes specifically**
1012 **expressed in appressoria of the rice blast fungus**. *The Plant Cell* 2002, **14**(9):2107-2119.
1013 39. Kulkarni RD, Kelkar HS, Dean RA: **An eight-cysteine-containing CFEM domain unique to a**
1014 **group of fungal membrane proteins**. *Trends in Biochemical Sciences* 2003, **28**(3):118-121.
1015 40. Garrido SM, Kitamoto N, Watanabe A, Shintani T, Gomi K: **Functional analysis of FarA**
1016 **transcription factor in the regulation of the genes encoding lipolytic enzymes and**

1017 **hydrophobic surface binding protein for the degradation of biodegradable plastics in**
1018 ***Aspergillus oryzae***. *Journal of Bioscience and Bioengineering* 2012, **113**(5):549-555.
1019 41. Ohtaki S, Maeda H, Takahashi T, Yamagata Y, Hasegawa F, Gomi K, Nakajima T, Abe K: **Novel**
1020 **hydrophobic surface binding protein, HsbA, produced by *Aspergillus oryzae***. *Applied and*
1021 *Environmental Microbiology* 2006, **72**(4):2407-2413.
1022 42. Mesarich CH, Schmitz M, Tremouilhac P, McGillivray DJ, Templeton MD, Dingley AJ: **Structure,**
1023 **dynamics and domain organization of the repeat protein Cin1 from the apple scab fungus**.
1024 *Biochimica et Biophysica Acta-Proteins and Proteomics* 2012, **1824**(10):1118-1128.
1025 43. Mesarich CH, Ökmen B, Rovenich H, Griffiths SA, Wang C, Karimi Jashni M, Mihajlovski A,
1026 Collemare J, Hunziker L, Deng CH *et al*: **Specific hypersensitive response-associated**
1027 **recognition of new apoplastic effectors from *Cladosporium fulvum* in wild tomato**.
1028 *Molecular Plant-Microbe Interactions* 2018, **31**(1):145-162.
1029 44. Bolton MD, Van Esse HP, Vossen JH, De Jonge R, Stergiopoulos I, Stulemeijer IJE, Van Den Berg
1030 GCM, Borrás-Hidalgo O, Dekker HL, De Koster CG *et al*: **The novel *Cladosporium fulvum* lysin**
1031 **motif effector Ecp6 is a virulence factor with orthologues in other fungal species**. *Molecular*
1032 *Microbiology* 2008, **69**(1):119-136.
1033 45. Kessler SC, Zhang X, McDonald MC, Gilchrist CLM, Lin Z, Rightmyer A, Solomon PS, Turgeon
1034 BG, Chooi Y-H: **Victorin, the host-selective cyclic peptide toxin from the oat pathogen**
1035 ***Cochliobolus victoriae* is ribosomally encoded**. *Proceedings of the National Academy of*
1036 *Sciences* 2020, **117**(39):24243.
1037 46. Kessler SC, Chooi Y-H: **Out for a RiPP: challenges and advances in genome mining of**
1038 **ribosomal peptides from fungi**. *Natural Product Reports* 2022, **39**(2):222-230.
1039 47. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates
1040 R, Žídek A, Potapenko A *et al*: **Highly accurate protein structure prediction with AlphaFold**.
1041 *Nature* 2021, **596**(7873):583-589.

1042 48. Holm L: **Using Dali for protein structure comparison**. *Methods in Molecular Biology* 2020,
1043 **2112**:29-42.
1044 49. Hoh F, Padilla, A, De Guillen, K: **New MAX effector from *Magnaporthe oryzae***. 2019. PDB DOI:
1045 10.2210/pdb6R5J/pdb.
1046 50. Li W, Wang B, Wu J, Lu G, Hu Y, Zhang X, Zhang Z, Zhao Q, Feng Q, Zhang H *et al*: **The**
1047 ***Magnaporthe oryzae* avirulence gene *AvrPiz-t* encodes a predicted secreted protein that**
1048 **triggers the immunity in rice mediated by the blast resistance gene *Piz-t***. *Molecular Plant-*
1049 *Microbe Interactions* 2009, **22**(4):411-420.
1050 51. de Guillen K, Ortiz-Vallejo D, Gracy J, Fournier E, Kroj T, Padilla A: **Structure analysis uncovers**
1051 **a highly diverse but structurally conserved effector family in phytopathogenic fungi**. *PLOS*
1052 *Pathogens* 2015, **11**(10):e1005228.
1053 52. Ose T, Oikawa A, Nakamura Y, Maenaka K, Higuchi Y, Satoh Y, Fujiwara S, Demura M, Sone T,
1054 Kamiya M: **Solution structure of an avirulence protein, AVR-Pia, from *Magnaporthe oryzae***.
1055 *Journal of Biomolecular NMR* 2015, **63**(2):229-235.
1056 53. Zhang X, He D, Zhao Y, Cheng X, Zhao W, Taylor IA, Yang J, Liu J, Peng YL: **A positive-charged**
1057 **patch and stabilized hydrophobic core are essential for avirulence function of AvrPib in the**
1058 **rice blast fungus**. *The Plant Journal* 2018, **96**(1):133-146.
1059 54. De la Concepcion JC, Franceschetti M, Maqbool A, Saitoh H, Terauchi R, Kamoun S, Banfield
1060 MJ: **Polymorphic residues in rice NLRs expand binding and response to effectors of the blast**
1061 **pathogen**. *Nature Plants* 2018, **4**(8):576-585.
1062 55. Nyarko A, Singarapu KK, Figueroa M, Manning VA, Pandelova I, Wolpert TJ, Ciuffetti LM,
1063 Barbar E: **Solution NMR structures of *Pyrenophora tritici-repentis* ToxB and its inactive**
1064 **homolog reveal potential determinants of toxin activity**. *Journal of Biological Chemistry*
1065 2014, **289**(37):25946-25956.
1066 56. Sarma GN, Manning VA, Ciuffetti LM, Karplus PA: **Structure of Ptr ToxA: An RGD-containing**
1067 **host-selective toxin from *Pyrenophora tritici-repentis*** *The Plant Cell* 2005, **17**(11):3190-3202.

1068 57. Ma L, Cornelissen BJC, Takken FLW: **A nuclear localization for Avr2 from *Fusarium oxysporum***
1069 **is required to activate the tomato resistance protein I-2**. *Frontiers in Plant Science* 2013, **4**:94.
1070 58. Di X, Cao L, Hughes RK, Tintor N, Banfield MJ, Takken FLW: **Structure-function analysis of the**
1071 ***Fusarium oxysporum* Avr2 effector allows uncoupling of its immune-suppressing activity**
1072 **from recognition**. *New Phytologist* 2017, **216**(3):897-914.
1073 59. Guncar G, Wang C-IA, Forwood JK, Teh T, Catanzariti A-M, Ellis JG, Dodds PN, Kobe B: **The use**
1074 **of Co^{2+} for crystallization and structure determination, using a conventional monochromatic**
1075 **X-ray source, of flax rust avirulence protein**. *Crystallization Communications* 2007, **63**(Pt
1076 3):209-213.
1077 60. Wang C-IA, Gunčar G, Forwood JK, Teh T, Catanzariti A-M, Lawrence GJ, Loughlin FE, Mackay
1078 JP, Schirra HJ, Anderson PA *et al*: **Crystal structures of flax rust avirulence proteins AvrL567-**
1079 **A and -D reveal details of the structural basis for flax disease resistance specificity**. *The Plant*
1080 *Cell* 2007, **19**(9):2898-2912.
1081 61. Blondeau K, Blaise F, Graille M, Kale SD, Linglin J, Ollivier B, Labarde A, Lazar N, Daverdin G,
1082 Balesdent MH *et al*: **Crystal structure of the effector AvrLm4-7 of *Leptosphaeria maculans***
1083 **reveals insights into its translocation into plant cells and recognition by resistance proteins**.
1084 *The Plant Journal* 2015, **83**(4):610-624.
1085 62. Lazar N, Mesarich CH, Petit-Houdenot Y, Talbi N, Li de la Sierra-Gallay I, Zélie E, Blondeau K,
1086 Gracy J, Ollivier B, Blaise F *et al*: **A new family of structurally conserved fungal effectors**
1087 **displays epistatic interactions with plant resistance proteins**. *PLOS Pathogens* 2022,
1088 **18**(7):e1010664.
1089 63. Yu DS, Outram MA, Smith A, McCombe CL, Khambalkar PB, Rima SA, Sun X, Ma L, Ericsson DJ,
1090 Jones DA *et al*: **The structural repertoire of *Fusarium oxysporum* f. sp. *lycopersici* effectors**
1091 **revealed by experimental and computational studies**. *bioRxiv* 2021: 2021.12.14.472499.

1092 64. Li N, Erman M, Pangborn W, Duax WL, Park C-M, Bruenn J, Ghosh D: **Structure of *Ustilago*
1093 *maydis* killer toxin KP6 α-subunit: a multimeric assembly with a central pore**. *Journal of*
1094 *Biological Chemistry* 1999, **274**(29):20425-20431.
1095 65. Allen A, Chatt E, Smith TJ: **The atomic structure of the virally encoded antifungal protein,**
1096 **KP6**. *Journal of Molecular Biology* 2013, **425**(3):609-621.
1097 66. Padilla A, Hoh, F, De Guillen, K: **Zt-KP6-1: an effector from *Zymoseptoria tritici***. 2019: PDB
1098 DOI: 10.2210/pdb6QPK/pdb.
1099 67. Fox NK, Brenner SE, Chandonia J-M: **SCOPe: Structural classification of proteins-extended,**
1100 **integrating SCOP and ASTRAL data and classification of new structures**. *Nucleic Acids*
1101 *Research* 2014, **42**(D1):D304-D309.
1102 68. Huber A, Hajdu D, Bratschun-Khan D, Gáspári Z, Varbanov M, Philippot S, Fizil Á, Czajlik A, Kele
1103 Z, Sonderegger C *et al*: **New antimicrobial potential and structural properties of PAFB: a**
1104 **cationic, cysteine-rich protein from *Penicillium chrysogenum* Q176**. *Scientific Reports* 2018,
1105 **8**(1):1751.
1106 69. Marx F, Binder U, Leiter E, Pócsi I: **The *Penicillium chrysogenum* antifungal protein PAF, a**
1107 **promising tool for the development of new antifungal therapies and fungal cell biology**
1108 **studies**. *Cellular and Molecular Life Sciences* 2008, **65**(3):445-454.
1109 70. Sonderegger C, Fizil Á, Burtscher L, Hajdu D, Muñoz A, Gáspári Z, Read ND, Batta G, Marx F:
1110 **D19S mutation of the cationic, cysteine-rich protein PAF: novel insights into its structural**
1111 **dynamics, thermal unfolding and antifungal function**. *PLOS ONE* 2017, **12**(1):e0169920.
1112 71. Takahashi T, Maeda H, Yoneda S, Ohtaki S, Yamagata Y, Hasegawa F, Gomi K, Nakajima T, Abe
1113 K: **The fungal hydrophobin RolA recruits polyesterase and laterally moves on hydrophobic**
1114 **surfaces**. *Molecular Microbiology* 2005, **57**(6):1780-1796.
1115 72. Koller W, Parker DM, Becker CM: **Role of cutinase in the penetration of apple leaves by**
1116 ***Venturia inaequalis.*** *Phytopathology* 1991, **81**(11):1375-1379.

1117 73. Martínez-Cruz J, Romero D, Hierrezuelo J, Thon M, de Vicente A, Pérez-García A: **Effectors**
1118 **with chitinase activity (EWCAs), a family of conserved, secreted fungal chitinases that**
1119 **suppress chitin-triggered immunity**. *The Plant Cell* 2021, **33**(4):1319-1340.
1120 74. Gamas P, Niebel Fde C, Lescure N, Cullimore J: **Use of a subtractive hybridization approach**
1121 **to identify new *Medicago truncatula* genes induced during root nodule development**.
1122 *Molecular Plant-Microbe Interactions* 1996, **9**(4):233-242.
1123 75. Kimura M, Yamamoto YY, Seki M, Sakurai T, Sato M, Abe T, Yoshida S, Manabe K, Shinozaki K,
1124 Matsui M: **Identification of *Arabidopsis* genes regulated by high light-stress using cDNA**
1125 **microarray**. *Photochemistry and Photobiology* 2003, **77**(2):226-233.
1126 76. Doss RP: **Treatment of pea pods with Bruchin B results in up-regulation of a gene similar to**
1127 ***MtN19***. *Plant Physiology Biochemistry* 2005, **43**(3):225-231.
1128 77. Naya L, Paul S, Valdés-López O, Mendoza-Soto AB, Nova-Franco B, Sosa-Valencia G, Reyes JL,
1129 Hernández G: **Regulation of copper homeostasis and biotic interactions by microRNA 398b**
1130 **in common bean**. *PLOS ONE* 2014, **9**(1):e84416.
1131 78. Zhang Z-N, Wu Q-Y, Zhang G-Z, Zhu Y-Y, Murphy RW, Liu Z, Zou C-G: **Systematic analyses**
1132 **reveal uniqueness and origin of the CFEM domain in fungi**. *Scientific Reports* 2015,
1133 **5**(1):13032.
1134 79. Zhao S, Shang X, Bi W, Yu X, Liu D, Kang Z, Wang X, Wang X: **Genome-wide identification of**
1135 **effector candidates with conserved motifs from the wheat leaf rust fungus *Puccinia triticina***.
1136 *Frontiers in Microbiology* 2020, **11**(1188).
1137 80. Zhu W, Wei W, Wu Y, Zhou Y, Peng F, Zhang S, Chen P, Xu X: **BcCFEM1, a CFEM domain-**
1138 **containing protein with putative GPI-anchored site, is involved in pathogenicity, conidial**
1139 **production, and stress tolerance in *Botrytis cinerea***. *Frontiers in Microbiology* 2017, **8**(1807).
1140 81. Wang J-X, Long F, Zhu H, Zhang Y, Wu J-Y, Shen S, Dong J-G, Hao Z-M: **Bioinformatic analysis**
1141 **and functional characterization of CFEM proteins in *Setosphaeria turcica***. *Journal of*
1142 *Integrative Agriculture* 2021, **20**(9):2438-2449.

1143 82. Choi W, Dean RA: **The adenylate cyclase gene *MAC1* of *Magnaporthe grisea* controls**
1144 **appressorium formation and other aspects of growth and development**. *The Plant Cell* 1997,
1145 **9**(11):1973-1983.
1146 83. Skrzydeł J, Borowska-Wykręt D, Kwiatkowska D: **Structure, assembly and function of cuticle**
1147 **from mechanical perspective with special focus on perianth**. *International Journal of*
1148 *Molecular Sciences* 2021, **22**(8):4160.
1149 84. Sacristán S, Vigouroux M, Pedersen C, Skamnioti P, Thordal-Christensen H, Micali C, Brown
1150 JKM, Ridout CJ: **Coevolution between a family of parasite virulence effectors and a class of**
1151 **LINE-1 retrotransposons**. *PLOS ONE* 2009, **4**(10):e7463.
1152 85. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, Ver Loren van Themaat E,
1153 Brown JK, Butcher SA, Gurr SJ *et al*: **Genome expansion and gene loss in powdery mildew**
1154 **fungi reveal tradeoffs in extreme parasitism**. *Science* 2010, **330**(6010):1543-1546.
1155 86. Dutheil JY, Mannhaupt G, Schweizer G, C MKS, Münsterkötter M, Güldener U, Schirawski J,
1156 Kahmann R: **A tale of genome compartmentalization: the evolution of virulence clusters in**
1157 **smut fungi**. *Genome Biology and Evolution* 2016, **8**(3):681-704.
1158 87. Fouché S, Plissonneau C, Croll D: **The birth and death of effectors in rapidly evolving**
1159 **filamentous pathogen genomes**. *Current Opinion in Microbiology* 2018, **46**:34-42.
1160 88. Amoozadeh S, Johnston J, Meisrimler C-N: **Exploiting structural modelling tools to explore**
1161 **host-translocated effector proteins**. *International Journal of Molecular Sciences* 2021,
1162 **22**(23):12962.
1163 89. Seong K, Krasileva KV: **Computational structural genomics unravels common folds and novel**
1164 **families in the secretome of fungal phytopathogen *Magnaporthe oryzae***. *Molecular Plant-*
1165 *Microbe Interactions* 2021:MPMI-03-21-0071-R.
1166 90. Seong K, Krasileva KV: **Comparative computational structural genomics highlights divergent**
1167 **evolution of fungal effectors**. *bioRxiv* 2022: 2022.05.02.490317.

1168 91. Cesari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, Rivas S, Alaux L, Kanzaki H,
1169 Okuyama Y *et al*: **The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe***
1170 ***oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding**. *The Plant Cell* 2013, **25**(4):1463-
1171 1481.
1172 92. Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, Tosa Y, Chuma I, Takano
1173 Y, Win J *et al*: **Association genetics reveals three novel avirulence genes from the rice blast**
1174 **fungal pathogen *Magnaporthe oryzae***. *The Plant Cell* 2009, **21**(5):1573-1591.
1175 93. Ashikawa I, Hayashi N, Yamane H, Kanamori H, Wu J, Matsumoto T, Ono K, Yano M: **Two**
1176 **adjacent nucleotide-binding site–leucine-rich repeat class genes are required to confer**
1177 ***Pikm*-specific rice blast resistance**. *Genetics* 2008, **180**(4):2267.
1178 94. Zhang S, Wang L, Wu W, He L, Yang X, Pan Q: **Function and evolution of *Magnaporthe oryzae***
1179 **avirulence gene *AvrPib* responding to the rice blast resistance gene *Pib***. *Scientific Reports*
1180 2015, **5**:11642.
1181 95. Oikawa K, Fujisaki K, Shimizu M, Takeda T, Saitoh H, Hirabuchi A, Hiraka Y, Białas A, Langner T,
1182 Kellner R *et al*: **The blast pathogen effector AVR-Pik binds and stabilizes rice heavy metal-**
1183 **associated (HMA) proteins to co-opt their function in immunity**. *bioRxiv* 2020.
1184 96. Guo L, Cesari S, de Guillen K, Chalvon V, Mammri L, Ma M, Meusnier I, Bonnot F, Padilla A,
1185 Peng Y-L *et al*: **Specific recognition of two MAX effectors by integrated HMA domains in plant**
1186 **immune receptors involves distinct binding surfaces**. *Proceedings of the National Academy*
1187 *of Sciences* 2018, **115**(45):11637.
1188 97. Franceschetti M, Maqbool A, Jiménez-Dalmaroni Maximiliano J, Pennington Helen G, Kamoun
1189 S, Banfield Mark J: **Effectors of filamentous plant pathogens: commonalities amid diversity**.
1190 *Microbiology and Molecular Biology Reviews* 2017, **81**(2):e00066-00016.
1191 98. Schouten HJ, Brinkhuis J, Burgh AMvd, Schaart JG, Groenwold R, Broggini GAL, Gessler C:
1192 **Cloning and functional characterization of the *Rvi15* (*Vr2*) gene for apple scab resistance**.
1193 *Tree Genetics & Genomes* 2013, **10**:251-260.

1194 99. Rafiqi M, Gan PH, Ravensdale M, Lawrence GJ, Ellis JG, Jones DA, Hardham AR, Dodds PN:
1195 **Internalization of flax rust avirulence proteins into flax and tobacco cells can occur in the**
1196 **absence of the pathogen**. *The Plant Cell* 2010, **22**(6):2017-2032.
1197 100. Cao L, Blekemolen MC, Tintor N, Cornelissen BJC, Takken FLW: **The *Fusarium oxysporum* Avr2-**
1198 **Six5 effector pair alters plasmodesmatal exclusion selectivity to facilitate cell-to-cell**
1199 **movement of Avr2**. *Molecular Plant* 2018, **11**(5):691-705.
1200 101. Dagvadorj B, Outram MA, Williams SJ, Solomon PS: **The necrotrophic effector ToxA from**
1201 ***Parastagonospora nodorum* interacts with wheat NHL proteins to facilitate Tsn1-mediated**
1202 **necrosis**. *The Plant Journal* 2022,**110**(2):407-418.
1203 102. Belfanti E, Silfverberg-Dilworth E, Tartarini S, Patocchi A, Barbieri M, Zhu J, Vinatzer BA,
1204 Gianfranceschi L, Gessler C, Sansavini S: **The *HcrVf2* gene from a wild apple confers scab**
1205 **resistance to a transgenic cultivated variety**. *Proceedings of the National Academy of Sciences*
1206 2004, **101**(3):886-890.
1207 103. Ghanbarnia K, Ma L, Larkan NJ, Haddadi P, Fernando WGD, Borhan MH: ***Leptosphaeria***
1208 ***maculans* AvrLm9: a new player in the game of hide and seek with AvrLm4-7**. *Molecular*
1209 *Plant Pathology* 2018, **19**(7):1754-1764.
1210 104. Plissonneau C, Daverdin G, Ollivier B, Blaise F, Degrave A, Fudal I, Rouxel T, Balesdent M-H: **A**
1211 **game of hide and seek between avirulence genes *AvrLm4-7* and *AvrLm3* in *Leptosphaeria***
1212 ***maculans***. *New Phytologist* 2016, **209**(4):1613-1624.
1213 105. Houterman PM, Cornelissen BJC, Rep M: **Suppression of plant resistance gene-based**
1214 **immunity by a fungal effector**. *PLOS Pathogens* 2008, **4**(5):e1000061.
1215 106. Outram MA, Figueroa M, Sperschneider J, Williams SJ, Dodds PN: **Seeing is believing:**
1216 **Exploiting advances in structural biology to understand and engineer plant immunity**.
1217 *Current Opinion in Plant Biology* 2022, **67**: 102210.

1218 107. Petit-Houdenot Y, Degrave A, Meyer M, Blaise F, Ollivier B, Marais CL, Jauneau A, Audran C,
1219 Rivas S, Veneault-Fourrey C *et al*: **A two genes-for-one gene interaction between**
1220 ***Leptosphaeria maculans* and *Brassica napus***. *New Phytologist* 2019, **223**(1):397-411.
1221 108. Mosquera G, Giraldo MC, Khang CH, Coughlan S, Valent B: **Interaction transcriptome analysis**
1222 **identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice**
1223 **blast disease**. *The Plant Cell* 2009, **21**(4):1273-1290.
1224 109. Ebert MK, Rangel LI, Spanner RE, Taliadoros D, Wang X, Friesen TL, de Jonge R, Neubauer JD,
1225 Secor GA, Thomma BPHJ *et al*: **Identification and characterization of *Cercospora beticola***
1226 **necrosis-inducing effector CbNip1**. *Molecular Plant Pathology* 2021, **22**(3):301-316.
1227 110. Postic G, Gracy J, Périn C, Chiche L, Gelly J-C: **KNOTTIN: the database of inhibitor cystine knot**
1228 **scaffold after 10 years, toward a systematic structure modeling**. *Nucleic Acids Research* 2018,
1229 **46**(D1):D454-D458.
1230 111. Vervoort J, van den Hooven HW, Berg A, Vossen P, Vogelsang R, Joosten MHAJ, de Wit PJGM:
1231 **The race-specific elicitor AVR9 of the tomato pathogen *Cladosporium fulvum*: a cystine knot**
1232 **protein: sequence-specific ^{1}H NMR assignments, secondary structure and global fold of the**
1233 **protein**. *FEBS Letters* 1997, **404**(2):153-158.
1234 112. van den Hooven HW, van den Burg HA, Vossen P, Boeren S, de Wit PJGM, Vervoort J: **Disulfide**
1235 **bond structure of the AVR9 elicitor of the fungal tomato pathogen *Cladosporium fulvum*:**
1236 **Evidence for a cystine knot**. *Biochemistry* 2001, **40**(12):3458-3466.
1237 113. de Guillen K, Lorrain C, Tsan P, Barthe P, Petre B, Saveleva N, Rouhier N, Duplessis S, Padilla A,
1238 Hecker A: **Structural genomics applied to the rust fungus *Melampsora larici-populina* reveals**
1239 **two candidate effector proteins adopting cystine knot and NTF2-like protein folds**. *Scientific*
1240 *Reports* 2019, **9**(1):18084.
1241 114. Snelders NC, Petti GC, van den Berg GCM, Seidl MF, Thomma BPHJ: **An ancient antimicrobial**
1242 **protein co-opted by a fungal plant pathogen for in planta mycobiome manipulation**.
1243 *Proceedings of the National Academy of Sciences* 2021, **118**(49):e2110968118.

1244 115. Lee H-T, Lee C-C, Yang J-R, Lai JZC, Chang KY: **A large-scale structural classification of**
1245 **antimicrobial peptides**. *BioMed Research International* 2015, **2015**:475062.
1246 116. Vogt E, Künzler M: **Discovery of novel fungal RiPP biosynthetic pathways and their**
1247 **application for the development of peptide therapeutics**. *Applied Microbiology and*
1248 *Biotechnology* 2019, **103**(14):5567-5581.
1249 117. Slazak B, Kapusta M, Strömstedt AA, Słomka A, Krychowiak M, Shariatgorji M, Andrén PE,
1250 Bohdanowicz J, Kuta E, Göransson U: **How does the sweet violet (*Viola odorata* L.) fight**
1251 **pathogens and pests – cyclotides as a comprehensive plant host defense system**. *Frontiers*
1252 *in Plant Science* 2018, **9**(1296).
1253 118. Rocafort M, Arshed S, Hudson D, Sidhu JS, Bowen JK, Plummer KM, Bradshaw RE, Johnson RD,
1254 Johnson LJ, Mesarich CH: **CRISPR-Cas9 gene editing and rapid detection of gene-edited**
1255 **mutants using high-resolution melting in the apple scab fungus, *Venturia inaequalis***. *Fungal*
1256 *Biology* 2021, **126**(1):35-46.
1257 119. Chen C, Bock CH, Wood BW: **Draft genome sequence of *Venturia carpophila*, the causal agent**
1258 **of peach scab**. *Standards in Genomic Sciences* 2017, **12**:68-68.
1259 120. Jaber MY, Bao J, Gao X, Zhang L, He D, Wang X, Wang A, Wang Z, Wang B: **Genome sequence**
1260 **of *Venturia oleaginea*, the causal agent of olive leaf scab**. *Molecular Plant-Microbe*
1261 *Interactions* 2020, **33**(9):1095-1097.
1262 121. Johnson S, Jones D, Thrimawithana AH, Deng CH, Bowen JK, Mesarich CH, Ishii H, Won K, Bus
1263 VGM, Plummer KM: **Whole genome sequence resource of the Asian pear scab pathogen**
1264 ***Venturia nashicola***. *Molecular Plant-Microbe Interactions* 2019, **32**(11):1463-1467.
1265 122. Prokchorchik M, Won K, Lee Y, Choi ED, Segonzac C, Sohn KH: **High contiguity whole genome**
1266 **sequence and gene annotation resource for two *Venturia nashicola* isolates**. *Molecular*
1267 *Plant-Microbe Interactions* 2019, **32**(9):1091-1094.

1268 123. Winter DJ, Charlton ND, Krom N, Shiller J, Bock CH, Cox MP, Young CA: **Chromosome-level**
1269 **reference genome of *Venturia effusa*, causative agent of pecan scab**. *Molecular Plant*
1270 *Microbe Interactions* 2020, **33**(2):149-152.
1271 124. Zhou Y, Zhang L, Fan F, Wang ZQ, Huang Y, Yin LF, Yin WX, Luo CX: **Genome sequence of**
1272 ***Venturia carpophila*, the causal agent of peach scab**. *Molecular Plant Microbe Interactions*
1273 2021, **34**(7):852-856.
1274 125. **The *Venturia populina* CBS256.38 v1.0 genome**
1275 [https://mycocosm.jgi.doe.gov/Venpo1/Venpo1.home.html]
1276 126. Stehmann C, Pennycook S, Plummer KM: **Molecular identification of a sexual interloper: the**
1277 **pear pathogen, *Venturia pirina*, has sex on apple**. *Phytopathology* 2001, **91**(7):633-641.
1278 127. Caffier V, Patocchi A, Expert P, Bellanger M-N, Durel C-E, Hilber-Bodmer M, Broggini GAL,
1279 Groenwold R, Bus VGM: **Virulence characterization of *Venturia inaequalis* reference isolates**
1280 **on the differential set of *Malus* hosts**. *Plant Disease* 2014, **99**(3):370-375.
1281 128. Xu X, Harvey A, Barbara DJ: **Population variation of apple scab (*Venturia inaequalis*) within**
1282 **mixed orchards in the UK**. *European Journal of Plant Pathology* 2013, **135**:97-104.
1283 129. Parker D, Hilber U, Bodmer M, Smith F, Yao C, Köller W: **Production and transformation of**
1284 **conidia of *Venturia inaequalis***. *Phytopathology* 1995, **85**(1):87-91.
1285 130. Win J, Greenwood DR, Plummer KM: **Characterisation of a protein from *Venturia inaequalis***
1286 **that induces necrosis in *Malus* carrying the *Vm* resistance gene**. *Physiological and Molecular*
1287 *Plant Pathology* 2003, **62**(4):193-202.
1288 131. Bruzzese E, Hasan S: **A whole leaf clearing and staining technique for host specificity studies**
1289 **of rust fungi**. *Plant Pathology* 1983, **32**(3):335-338.
1290 132. Atkinson RG, Johnston SL, Yauk Y-K, Sharma NN, Schröder R: **Analysis of xyloglucan**
1291 **endotransglucosylase/hydrolase (XTH) gene families in kiwifruit and apple**. *Postharvest*
1292 *Biology and Technology* 2009, **51**(2):149-157.

1293 133. Wu TD, Watanabe CK: **GMAP: a genomic mapping and alignment program for mRNA and EST**
1294 **sequences**. *Bioinformatics* 2005, **21**(9):1859-1875.
1295 134. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements**.
1296 *Nature Methods* 2015, **12**(4):357-360.
1297 135. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL: **Graph-based genome alignment and**
1298 **genotyping with HISAT2 and HISAT-genotype**. *Nature Biotechnology* 2019, **37**(8):907-915.
1299 136. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
1300 McCarthy SA, Davies RM *et al*: **Twelve years of SAMtools and BCFtools**. *Gigascience* 2021,
1301 **10**(2).
1302 137. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
1303 Raychowdhury R, Zeng Q *et al*: **Full-length transcriptome assembly from RNA-Seq data**
1304 **without a reference genome**. *Nature Biotechnology* 2011, **29**(7):644-652.
1305 138. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
1306 Markowitz S, Duran C *et al*: **Geneious Basic: an integrated and extendable desktop software**
1307 **platform for the organization and analysis of sequence data**. *Bioinformatics* 2012,
1308 **28**(12):1647-1649.
1309 139. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A,
1310 Nuka G *et al*: **InterProScan 5: genome-scale protein function classification**. *Bioinformatics*
1311 2014, **30**(9):1236-1240.
1312 140. Törönen P, Medlar A, Holm L: **PANNZER2: a rapid functional annotation web server**. *Nucleic*
1313 *Acids Research* 2018, **46**(W1):W84-W88.
1314 141. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von
1315 Heijne G, Nielsen H: **SignalP 5.0 improves signal peptide predictions using deep neural**
1316 **networks**. *Nature Biotechnology* 2019, **37**(4):420-423.

1317 142. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein**
1318 **topology with a hidden Markov model: application to complete genomes**. *Journal of*
1319 *Molecular Biology* 2001, **305**(3):567-580.
1320 143. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y: **dbCAN2: a meta**
1321 **server for automated carbohydrate-active enzyme annotation**. *Nucleic Acids Research* 2018,
1322 **46**(W1):W95-W101.
1323 144. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N: **The carbohydrate-active**
1324 **enzyme database: functions and literature**. *Nucleic Acids Research* 2022, **50**(D1):D571-d577.
1325 145. Sperschneider J, Dodds PN: **EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors**
1326 **in fungi and oomycetes**. *Molecular Plant-Microbe Interactions* 2022, **35**(2):146-156.
1327 146. Nepusz T, Sasidharan R, Paccanaro A: **SCPS: a fast implementation of a spectral method for**
1328 **detecting protein families on a genome-wide scale**. *BMC Bioinformatics* 2010, **11**(1):120.
1329 147. Dainat J, Hereñú D, Davis E, Crouch K, Sol L, Agostinho N, Git P, Tayyrov A: **AGAT: Another gff**
1330 **analysis toolkit to handle annotations in any gtf/gff format**. *Zenodo* 2022,
1331 https://www.doi.org/10.5281/zenodo.3552717.
1332 148. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features**.
1333 *Bioinformatics* 2010, **26**(6):841-842.
1334 149. Li H, Durbin R: **Fast and accurate long-read alignment with Burrows–Wheeler transform**.
1335 *Bioinformatics* 2010, **26**(5):589-595.
1336 150. Wingett SW, Andrews S: **FastQ Screen: a tool for multi-genome mapping and quality control**.
1337 *F1000Research* 2018, **7**:1338.
1338 151. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning**
1339 **sequence reads to genomic features**. *Bioinformatics* 2014, **30**(7):923-930.
1340 152. Ewels P, Magnusson M, Lundin S, Käller M: **MultiQC: summarize analysis results for multiple**
1341 **tools and samples in a single report**. *Bioinformatics* 2016, **32**(19):3047-3048.

1342 153. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-**
1343 **seq data with DESeq2**. *Genome Biology* 2014, **15**(12):550.
1344 154. Wickham H: **ggplot2: elegant graphics for data analysis**. *Springer-Verlag New York* 2016.
1345 155. Gu Z, Eils R, Schlesner M: **Complex heatmaps reveal patterns and correlations in**
1346 **multidimensional genomic data**. *Bioinformatics* 2016, **32**(18):2847-2849.
1347 156. Alexa A, Rahnenführer J, Lengauer T: **Improved scoring of functional groups from gene**
1348 **expression data by decorrelating GO graph structure**. *Bioinformatics* 2006, **22**(13):1600-
1349 1607.
1350 157. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M: **ColabFold: making**
1351 **protein folding accessible to all**. *Nature Methods* 2022, **19**(6):679-682.
1352 158. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M,
1353 Söding J *et al*: **Fast, scalable generation of high-quality protein multiple sequence alignments**
1354 **using Clustal Omega**. *Molecular Systems Biology* 2011, **7**(1):539.
1355 159. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R: **A new bioinformatics**
1356 **analysis tools framework at EMBL–EBI**. *Nucleic Acids Research* 2010, **38**:W695-W699.
1357 160. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN,
1358 Alva V: **A completely reimplemented MPI bioinformatics toolkit with a new HHpred server**
1359 **at its core**. *Journal of Molecular Biology* 2018, **430**(15):2237-2243.
1360 161. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V: **Protein sequence**
1361 **analysis using the MPI bioinformatics toolkit**. *Current Protocols in Bioinformatics* 2020,
1362 **72**(1):e108.
1363 162. Necci M, Piovesan D, Dosztányi Z, Tosatto SCE: **MobiDB-lite: fast and highly specific**
1364 **consensus prediction of intrinsic disorder in proteins**. *Bioinformatics* 2017, **33**(9):1402-1404.
1365 163. Ishida T, Kinoshita K: **PrDOS: prediction of disordered protein regions from amino acid**
1366 **sequence**. *Nucleic Acids Research* 2007, **35**:W460-W464.

1367 164. Shindyalov IN, Bourne PE: **Protein structure alignment by incremental combinatorial**
1368 **extension (CE) of the optimal path**. *Protein Engineering* 1998, **11**(9):739-747.
1369 165. Zhang Y, Skolnick J: **TM-align: a protein structure alignment algorithm based on the TM-**
1370 **score**. *Nucleic Acids Research* 2005, **33**(7):2302-2309.
1371 166. Ayoub R, Lee Y: **RUPEE: A fast and accurate purely geometric protein structure search**. *PLOS*
1372 *ONE* 2019, **14**(3):e0213712.
1373 167. Ayoub R, Lee Y: **Protein structure search to support the development of protein structure**
1374 **prediction methods**. *Proteins: Structure, Function, and Bioinformatics* 2021, **89**(6):648-658.
1375 168. Chandonia J-M, Fox NK, Brenner SE: **SCOPe: classification of large macromolecular structures**
1376 **in the structural classification of proteins—extended database**. *Nucleic Acids Research* 2019,
1377 **47**(D1):D475-D481.
1378