A new family of structurally conserved fungal effectors displays epistatic interactions with plant resistance proteins

Short title: A structural family of effectors displaying epistatic interactions

Author list

N. Lazar[1], C. H. Mesarich[2], Y. Petit-Houdenot[3], N. Talbi[3], I. Li de la Sierra-Gallay[1], E. Zélie[1], K. Blondeau[1], J. Gracy[4], B. Ollivier[3], F. Blaise[3], T. Rouxel[3], M.H. Balesdent[3], A. Idnurm[5], H. van Tilbeurgh[1, ¥] and I. Fudal[3, ¥]

[1] Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CEA, CNRS, Gif-sur-Yvette, France

[2] Laboratory of Molecular Plant Pathology, School of Agriculture and Environment, Massey University, Palmerston North, New Zealand

[3] Université Paris-Saclay, INRAE, AgroParisTech, UMR BIOGER, 78850, Thiverval-Grignon, France

[4] CNRS UMR 5048, INSERM U1054, Centre de Biochimie Structurale, Université Montpellier, 34090 Montpellier, France

[5] School of BioSciences, The University of Melbourne, VIC 3010, Australia

[¥] Corresponding authors:
Herman van Tilbeurgh
herman.van-tilbeurgh@i2bc.paris-saclay.fr
and
Isabelle Fudal
Isabelle.Fudal@inrae.fr

Abstract

Recognition of a pathogen avirulence (AVR) effector protein by a cognate plant resistance (R) protein triggers a set of immune responses that render the plant resistant. Pathogens can escape this so-called Effector-Triggered Immunity (ETI) by different mechanisms including the deletion or loss-of-function mutation of the *AVR* gene, the incorporation of point mutations that allow recognition to be evaded while maintaining virulence function, and the acquisition of new effectors that suppress AVR recognition. The Dothideomycete *Leptosphaeria maculans*, causal agent of oilseed rape stem canker, is one of the few fungal pathogens where suppression of ETI by an AVR effector has been demonstrated. Indeed, AvrLm4-7 suppresses the recognition of AvrLm3 and AvrLm5-9 by the R proteins Rlm3 and Rlm9, respectively. The presence of *AvrLm4-7* does not impede *AvrLm3* and *AvrLm5-9* expression, and the three AVR proteins do not appear to physically interact. To decipher the epistatic interaction between these *L. maculans* AVR effectors, we determined the crystal structure of AvrLm5-9 and obtained a 3D model of AvrLm3, based on the crystal structure of Ecp11-1, a homologous avirulence effector candidate from *Fulvia fulva.* Despite a lack of sequence similarity, AvrLm5-9 and AvrLm3 are structural analogues of AvrLm4-7 (structure previously characterized). Structure-informed sequence database searches identified a larger number of putative structural analogues among *L. maculans* effector candidates, including the avirulence effector AvrLmS-Lep2, all produced during the early stages of oilseed rape infection, as well as among effector candidates from other phytopathogenic fungi. These structural analogues were called LARS (for Leptosphaeria AviRulence-Suppressing) effectors. Remarkably, transformants of *L. maculans* expressing one of these structural analogues, Ecp11-1, triggered Rlm3-mediated immunity. Furthermore, this recognition could be suppressed by AvrLm4-7. These results suggest that Ecp11-1 has the same function as AvrLm3, or that the Ecp11-1 structure is sufficiently close to that of AvrLm3 to be recognized by Rlm3.

Introduction

Fungi are the most devastating pathogens of plants, including crops of major economic importance. They represent a recurrent threat to agriculture and possess extreme adaptive abilities, resulting in the constant outbreaks of diseases (Fones et al., 2017; Fisher et al., 2018). Host invasion relies on effectors, key elements of pathogenesis, that modulate plant immunity and facilitate infection (Oliva et al., 2010; Rocafort et al., 2020). Fungal effector genes are diverse and typically encode small proteins, predicted to be secreted, with no or few homologues present in sequence databases, and an absence of known sequence motifs. In most phytopathogenic fungi, no large effector gene families have been identified (Lo Presti et al., 2015). Notably, effectors can have a dual role in plant-pathogen interactions, both targeting plant components and being targeted by plant resistance (R) proteins. Such dual-role effectors are known as avirulence (AVR) proteins because, in the presence of a corresponding R protein, they render the pathogen that produces them avirulent. Recognition of a pathogen AVR protein triggers a set of immune responses grouped under the term Effector-Triggered Immunity (ETI), frequently leading to a rapid localized cell death termed the hypersensitive response (HR) (Jones and Dangl, 2006). Breeding cultivars carrying major *R* genes against pathogens is a common and powerful tool to control disease. However, the massive deployment of single *R* genes in the field exerts a strong selection pressure against avirulent pathogens that can become virulent through evolution of their *AVR* gene repertoire. Mechanisms leading to virulence include deletion, inactivation or down-regulation of the *AVR* gene, point mutations allowing recognition to be evaded while maintaining the virulence function of the AVR protein, or the acquisition of new effectors that suppress ETI (Jones and Dangl, 2006; Guttman et al., 2014; Sanchez-Vallet et al., 2018). Suppression of ETI by a fungal effector represents an efficient way to evade the selection pressure exerted by *R* genes in the field while maintaining the function of non-dispensable effectors. In some cases, the effector that suppresses ETI can itself be recognized by an R protein. A few examples of such strategies have been described in fungi (reviewed by Petit-Houdenot and Fudal, 2017), but the underlying mechanisms for the suppression of ETI by fungal effectors remain unexplained.

The Dothideomycete *Leptosphaeria maculans*, causal agent of oilseed rape stem canker or blackleg disease, is one of the fungal pathogens in which suppression of ETI by the presence of an *AVR* gene has been demonstrated. *L. maculans* can be controlled by combining qualitative and quantitative resistance of the host plant (Brun et al., 2010). To date, nine *AVR* genes (called *AvrLm*) recognized by the products of *R* genes (called *Rlm*) from *Brassica napus* or other

Brassica species have been cloned (Rouxel and Balesdent, 2017; Ghanbarnia et al., 2018; Petit-Houdenot et al., 2019; Neik et al., 2020) and share common characteristics: they encode small secreted proteins with no or low homologies in sequence databases, are located in repeat-rich regions of the genome, and are specifically expressed during the early stages of leaf infection. Among them, AvrLm4-7 hides AvrLm3 from Rlm3-mediated recognition and AvrLm5-9 from Rlm9-mediated recognition (Plissonneau et al., 2016; Ghanbarnia et al., 2018). How AvrLm4-7 suppresses recognition of AvrLm3 and AvrLm5-9 by Rlm3 and Rlm9, respectively, is not known: the presence of *AvrLm4-7* does not impede *AvrLm3* and *AvrLm5-9* expression, and yeast two-hybrid (Y2H) assays suggest the absence of a physical interaction between AvrLm4-7, AvrLm5-9 and AvrLm3. While AvrLm5-9 and AvrLm3 share 29 % of amino acid sequence identity, a very low level of identity was found with AvrLm4-7 (15 %). AvrLm4-7 confers a dual recognition specificity by two distinct R proteins of oilseed rape, Rlm4 and Rlm7 (Parlange et al., 2009), and loss of *AvrLm4-7* is associated with a strong fitness cost (Huang et al., 2006; Novakova et al., 2016). *AvrLm5*-9 and *AvrLm3*, on the other hand, are always present in *L. maculans* isolates, and only point mutation polymorphisms were reported, suggesting a high importance of these two effectors in pathogenicity towards *B. napus* (Van de Wouw et al., 2014; Plissonneau et al. 2017; Ghanbarnia et al., 2018). Moreover, the silencing of *AvrLm3* led to a reduced aggressiveness (Plissonneau et al. 2016).

Elucidation of the 3D structures of effectors may provide an effective strategy to resolve functional traits. Indeed, structure determination of effectors and the proteins with which they interact has provided key advances in our understanding of plant-pathogen interactions, including: the identification of protein functions that were not apparent from sequence analysis alone, the visualization of molecular interfaces of relevance to pathogen virulence and to plant immunity, and the identification of structural homologies in effectors that were not visible by sequence comparisons (reviewed by Franceschetti et al., 2017). The crystal structure of AvrLm4-7 did not reveal similarities with documented effectors, but suggested a positively charged surface patch could be involved in AvrLm4-7 translocation into the cytoplasm of plant cells (Blondeau et al. (2015). AvrLm4-7 escapes Rlm4-mediated recognition through a single point mutation (Parlange et al., 2009) and Rlm7-mediated recognition through more drastic DNA changes or point mutations (Daverdin et al., 2012). Blondeau et al. (2015) identified a protein region involved in Rlm4-mediated recognition, and two regions involved in Rlm7-mediated recognition.

Here we describe the 3D structures of AvrLm5-9 and AvrLm3, whose recognition by Rlm9 and Rlm3, respectively, is masked by the presence of AvrLm4-7. Surprisingly, despite low sequence similarity, AvrLm5-9 and AvrLm3 are structural analogues of AvrLm4-7, sharing an anti-parallel β-sheet covered by α-helices. Structure-informed and pattern-based searches identified a larger number of putative structural analogues among avirulence effectors and effector candidates of *L. maculans*, but also of other phytopathogenic fungi, including Ecp11-1 from the biotrophic tomato leaf mold fungus *Fulvia fulva* (formerly *Cladosporium fulvum*). Remarkably, transformants of *L. maculans* producing *F. fulva* Ecp11-1 triggered Rlm3-mediated immunity and this recognition could be suppressed by AvrLm4-7. These findings give insight into the way effectors suppress ETI and can guide recommendations on how to use plant resistance genes targeting AVRs belonging to structural families of effectors.

Results

- Determination of the 3D structures of AvrLm5-9 and AvrLm3

To explore the putative molecular relationships between AvrLm4-7, AvrLm5-9 and AvrLm3, we set out to determine the 3D structures of the AvrLm3 and AvrLm5-9 effectors. AvrLm5-9 and AvrLm3 are rich in cysteines and therefore are difficult to express in soluble form in *Escherichia coli*. For the recombinant production of AvrLm5-9 and AvrLm3, we therefore chose the well-established *Pichia pastoris* eukaryotic expression system. The genes coding for the AvrLm5-9 and AvrLm3 proteins without their secretion signal peptides were cloned into expression vectors as fusion proteins with a purification His-tag and with thioredoxin. A TEV proteolytic cleavage site was inserted between thioredoxin and the effectors. The AvrLm5-9 fusion protein was well expressed and purified to homogeneity (Fig S1), but the yields of the AvrLm3 fusion protein were insufficient (*i.e.* about 50 mg of pure AvrLm5-9 per liter of cell culture against less than 1 mg of purified AvrLm3).

Recently, a small secreted protein with 37 % amino acid sequence similarity with AvrLm3 was identified from *F. fulva* (Mesarich et al., 2018). This protein, named Ecp11-1 (Extracellular protein 11-1), was found in apoplastic washing fluid samples harvested from compatible *F. fulva–Solanum lycopersicum* (tomato) interactions. Curiously, Ecp11-1 triggers an HR in multiple wild accessions of tomato. It is therefore likely that Ecp11-1 is an AVR effector recognized by a corresponding R protein (tentatively named Cf-Ecp11-1) in wild accessions of tomato (Mesarich et al., 2018). We decided to determine the 3D structure of Ecp11-1 using the same *P. pastoris*-based strategy as described for AvrLm5-9 and AvrLm3.

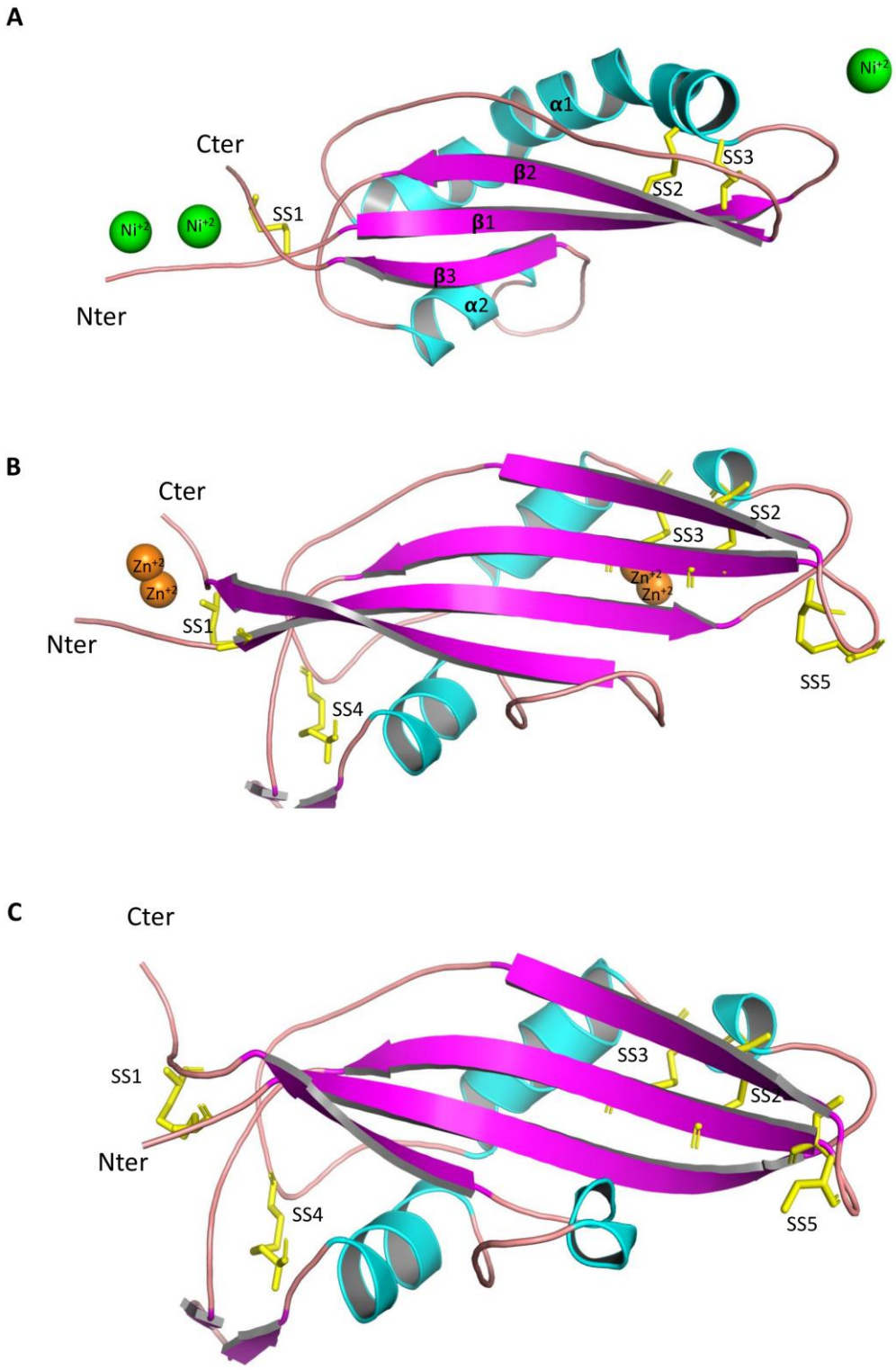

Fig 1. Crystal structures of AvrLm5-9 and Ecp11-1 and a 3D-model of AvrLm3.

The proteins are represented as cartoons and coloured by secondary structure (α-helix in cyan, β-strand in magenta). N- and C-termini are labelled. The S-S bonds are represented by yellow sticks, and are numbered (numbering refers to sequence alignment in figure 2A). (A) Structure of AvrLm5-9. Ni^{+2} ions involved in crystal packing are shown as green spheres. (B) Structure of ECP11-1. Zn^{+2} ions involved in crystal packing are shown as orange spheres. (C) Swiss-Model structure of AvrLm3. The AvrLm3 amino acid sequence and the Ecp11-1 X-ray structure were used to compute the AvrLm3 three-dimensional model.

The yields of Ecp11-1 production were sufficient to start structural studies (Fig S1, *i.e.* about 5 mg of purified Ecp11-1 per liter of cell culture).

After cleavage by the TEV protease and removal of thioredoxin, both AvrLm5-9 and Ecp11-1 provided good quality crystals. The structure of AvrLm5-9 was solved using iodine single wavelength anomalous diffusion (SAD) signal from a derivative crystal and then refined at 2.14 Å resolution using native data (Table S1). The crystallization liquor of AvrLm5-9 contained 80 mM of Ni^{2+} ions that proved mandatory for obtaining crystals. The structure revealed the presence of three Ni^{2+} ions that are involved in crystal contacts: two are found at the N- and C-termini of AvrLm5-9 as a ligand with two histidines from the native protein, by one histidine from a linker peptide and one from a crystal neighbor. The third Ni^{2+} ion is bound at the opposite end, and also interacts with histidines from two neighboring copies of AvrLm5-9 in the crystal. This suggests the bound Ni^{2+} ions are a crystallographic artefact. The complete AvrLm5-9 sequence could be placed into the electron density, which also accounted for two residues from the linker peptide at the N-terminus. The structure of AvrLm5-9 consists of a central β-sheet made of three anti-parallel β-strands (Fig 1A). An elongated peptide (residues 54 to 64) runs anti-parallel to β-strand 2, but only establishes a few main-chain H-bonds, and therefore is not categorized as a β-strand. One face of the β-sheet is covered by the long connections between the stands. The connection between the β1 and β2 strands is a curved α-helix and the connection between the β3 and β4 strands contains a shorter helix surrounded by two irregular peptide loops. AvrLm5-9 has three disulfide bridges, which are C^{3}-C^{119}, C^{22}-C^{69} and C^{26}-C^{71}, named SS1, SS2 and SS3, respectively, according to protein numeration without the signal peptide (Fig 2) or C^{22}-C^{138}, C^{41}-C^{88} and C^{45}-C^{90} according to protein numeration with the signal peptide. The SS1 bridge knits the N- and C-termini together, while the other two disulfide bridges fix the long helix onto the β-sheet.

The presence of a metal, in this case Zn^{2+}, was also mandatory for the crystallization of Ecp11-1. Speculating that the Ecp11-1 crystal would also bind a bivalent ion, we successfully solved its structure at 1.6 Å resolution by using the SAD-signal of Zn^{2+}. The crystals contained one copy of Ecp11-1 in the asymmetric unit. The complete sequence could be positioned into the electron density. The analysis of the anomalous signal revealed the presence of two Zn-clusters in the structure that resemble those of the Ni-ions in AvrLm5-9. One Zn-cluster is bound near the N- and C-termini and involves histidines from the native protein and from the linker peptide and side chains from a crystal neighbor. Another Zn-cluster is found at the opposite end and is

composed by residues emanating from neighboring Ecp11-1 molecules. Ecp11-1 forms an anti-parallel four-stranded β-sheet with (+2,-1,-2) topology (Fig 1B and 2C). Strands β1 and β2 are connected by a peptide composed of a helical turn and a short helix surrounded by irregular peptides. The long connection between strands β3 and β4 contains a β-hairpin, a short helix and some irregular peptide stretches. Ecp11-1 contains five disulfide bridges: C^{3}-C^{137} (named SS1) that connects the N- and C-termini, C^{22}-C^{71} (named SS2) and C^{26}-C^{73} (named SS3) that link the β1-β2 connection to strand β3. The two remaining disulfide bridges (C^{44}-C^{96} and C^{62}-C^{65}, named SS4 and SS5) (according to protein numbering in Fig 2) are found in the irregular loop regions.

AvrLm3 and Ecp11-1 share 37 % identity and 59% similarity, and there are only minor insertions/deletions between the two sequences. We consequently constructed a 3D-model of AvrLm3 using the Swiss-Model web server (https://swissmodel.expasy.org/). The sequence alignment and structure superposition show conservation of all 10 cysteines, suggesting both proteins form an identical set of disulfide bridges (Fig 1B and C).

- AvrLm4-7, AvrLm5-9, AvrLm3 and Ecp11-1 form a structural family

Although the sequence similarity between AvrLm5-9 and Ecp11-1/AvrLm3 is weak, their 3D structures are strongly related. Interestingly, their folding is also shared with AvrLm4-7 (Fig 2B), which displays suppressive interactions with AvrLm5-9 and AvrLm3. The sequence similarity between AvrLm4-7 on the one hand and AvrLm5-9 and Ecp11-1 on the other is even weaker than between the latter two (Fig 2A-B). Both AvrLm4-7 (β4) and AvrLm5-9 (β2) have an irregular β-strand, but these are oriented and positioned in the same way as the corresponding strands in Ecp11-1. To fully appreciate the similarities between the three effectors, we superposed their structures using the DALI webserver (http://ekhidna2.biocenter.helsinki.fi/dali/). Superposition of Ecp11-1 and AvrLm5-9 gives a Z-score of 8.9 and an RMSD value of 3.3 Å (106 residues aligned) and superposition of Ecp11-1 and AvrLm4-7 gives a Z-score of 5.6 and an RMSD value of 3.8 Å (106 residues aligned). The main differences between the three proteins are found in the regions connecting the strands (Fig 2B and 2C). The structures of the three effectors are stabilized by disulfide bridges, two of which are overlapping in the three proteins. All three effectors have a disulfide bridge that connects the N- and C-termini, and also at least one disulfide bridge that links the helical connection between β1-β2 and strand β3.

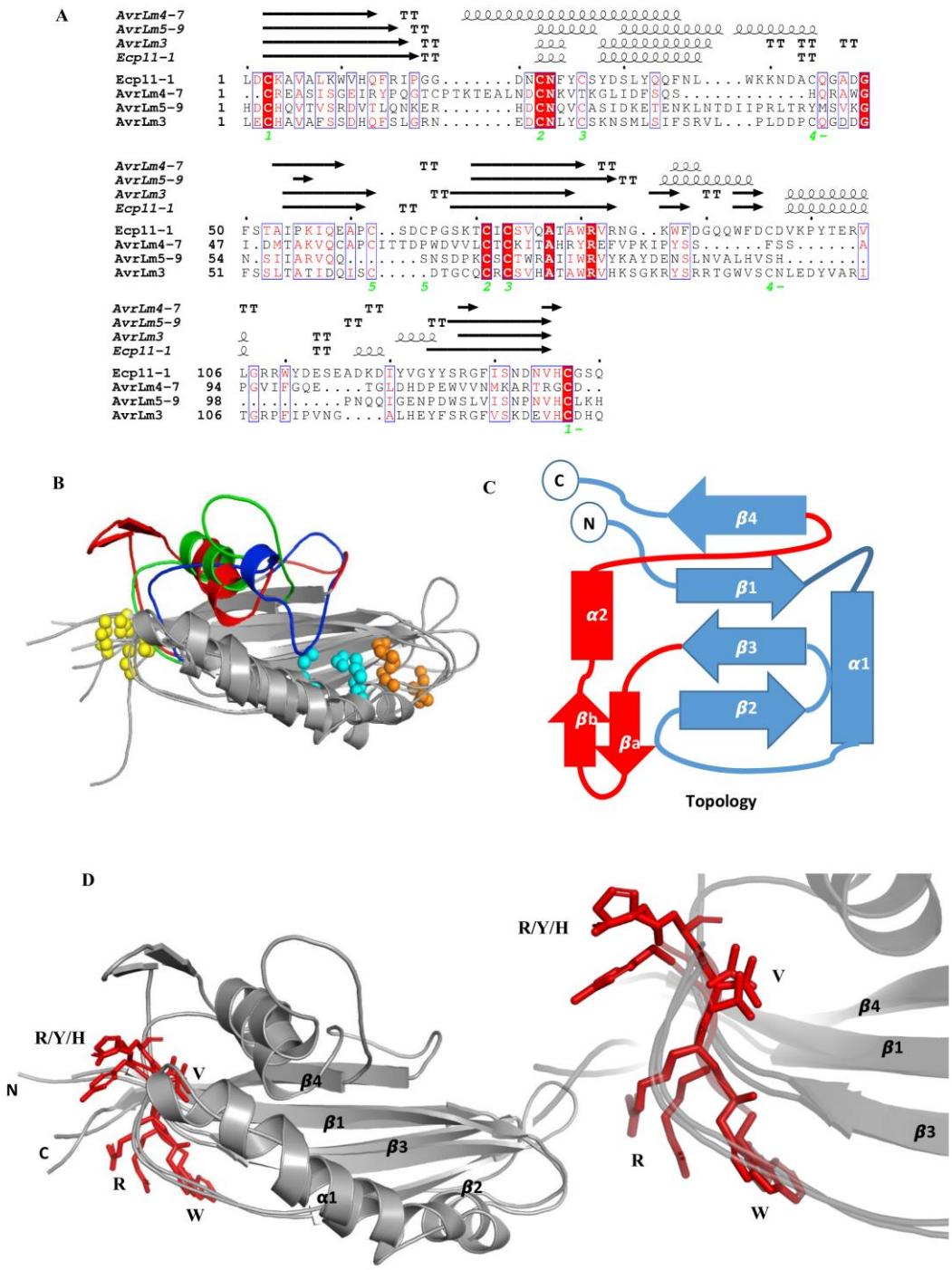

Fig 2. AvrLm4-7, AvrLm5-9, AvrLm3 and Ecp11-1 are structural analogues.
(A) Structure-based protein sequence alignment of AvrLm4-7, AvrLm5-9, AvrLm3 and Ecp11-1. The S-S bridges of Ecp11-1 are labelled green and numbered to show their connectivity. Secondary structure for each protein (β-sheets, α-helices, and β-turns are rendered as arrows, squiggles and TT letters respectively) is shown above the alignment. Identical residues are in red boxes and similar residues are in red. The Figure was made with the ESPript server (Robert and Gouet, 2014). (B) Superposition of AvrLm4-7, AvrLm5-9 and AvrLm3. The variable connections between β3 and β4 are coloured in red (AvrLm3), blue (AvrLm4-7) and green (AvrLm5-9). The conserved or neighbouring S-S bridges are indicated as yellow, orange and cyan spheres (indicated as 1, 2 and 3, respectively, in panel A). (C) Representation of conserved topology, with the variable connexion between β3 and β4 coloured in red. (D) Superposition of AvrLm5-9, Ecp11-1 and AvrLm3 (all in grey), with their conserved WR(F/L/V)(R/K) motif at the exit of β3 represented by red sticks. The last Arg of the motif happens to be either His in AvrLm3 or Tyr in AvrLm5-9. A zoom on the motif is represented in the right corner.

A search in the Protein Data Bank (PDB) for structural analogues of Ecp11-1 and AvrLm5-9 returned AvrLm4-7 and the yeast elongation factor 1B. The latter protein (100 residues) indeed has the same β-sheet topology and similar connections, but shares no significant sequence identity with the effectors and has no disulfide bridges. The similarity is therefore probably not biologically relevant.

- **Polymorphic residues identified in *L. maculans* and *F. fulva* populations are mainly located on external loops of AvrLm5-9, AvrLm3 and Ecp11-1**

We then set out to identify regions putatively required for recognition by cognate R proteins by mapping polymorphic residues onto their respective structures. We therefore exploited previous population studies that had reported polymorphisms in AvrLm5-9, AvrLm3 and Ecp11-1. Only three polymorphic alleles of AvrLm5-9 were reported in *L. maculans* populations (Van de Wouw et al., 2014; Ghanbarnia et al., 2018). One event caused a switch to virulence towards both *Rlm5* and *Rlm9* (R^{29} to stop) (R^{10} according to protein numbering in Fig 2), a point mutation at residue R^{55} to T or K leads to virulence towards Rlm9. A third polymorphism had no effect on the interaction with Rlm5 or Rlm9 (R^{38} to L) (R^{19} according to protein numbering in Fig 2). The three polymorphic residues were mapped onto the AvrLm5-9 3D structure (Fig 3A). The stop mutation evidently leads to a truncated and non-functional protein. The mutation leading to virulence towards *Rlm9* is situated in the middle of the long helix, and the mutation having no effect on the interaction with Rlm9 and Rlm5 is in a loop connecting β1 and β2. Neither mutation is expected to perturb the 3D structure.

Sequencing of AvrLm3 in a collection of worldwide *L. maculans* isolates defined 22 different alleles, corresponding to 14 non-synonymous mutations leading to 11 isoforms of the protein (Plissonneau et al., 2017). We projected eight of the 11 polymorphic amino acids onto the AvrLm3 3D structure, (the remaining three are located in the signal peptide and thus could not be plotted) (Fig 3B). Most of the AvrLm3 variants concern amino acid positions with surface-exposed side chains, and none of the variants affects the cysteines. Only one polymorphic substitution (L^{78} to F) (L^{54} according to protein numbering in Fig 2) is found in a residue involved in hydrophobic packing, but the mutation is conservative. These observations suggest that all alleles should lead to well-folded proteins. Six of these AvrLm3 isoforms were only found in isolates avirulent towards *Rlm3*. These amino acids are scattered over the surface of the protein. Only two amino acid residues consistently differed between virulent and avirulent isoforms: I/L^{58}H and G^{131}R (I^{34} and G^{107} according to protein numbering in Figure 2).

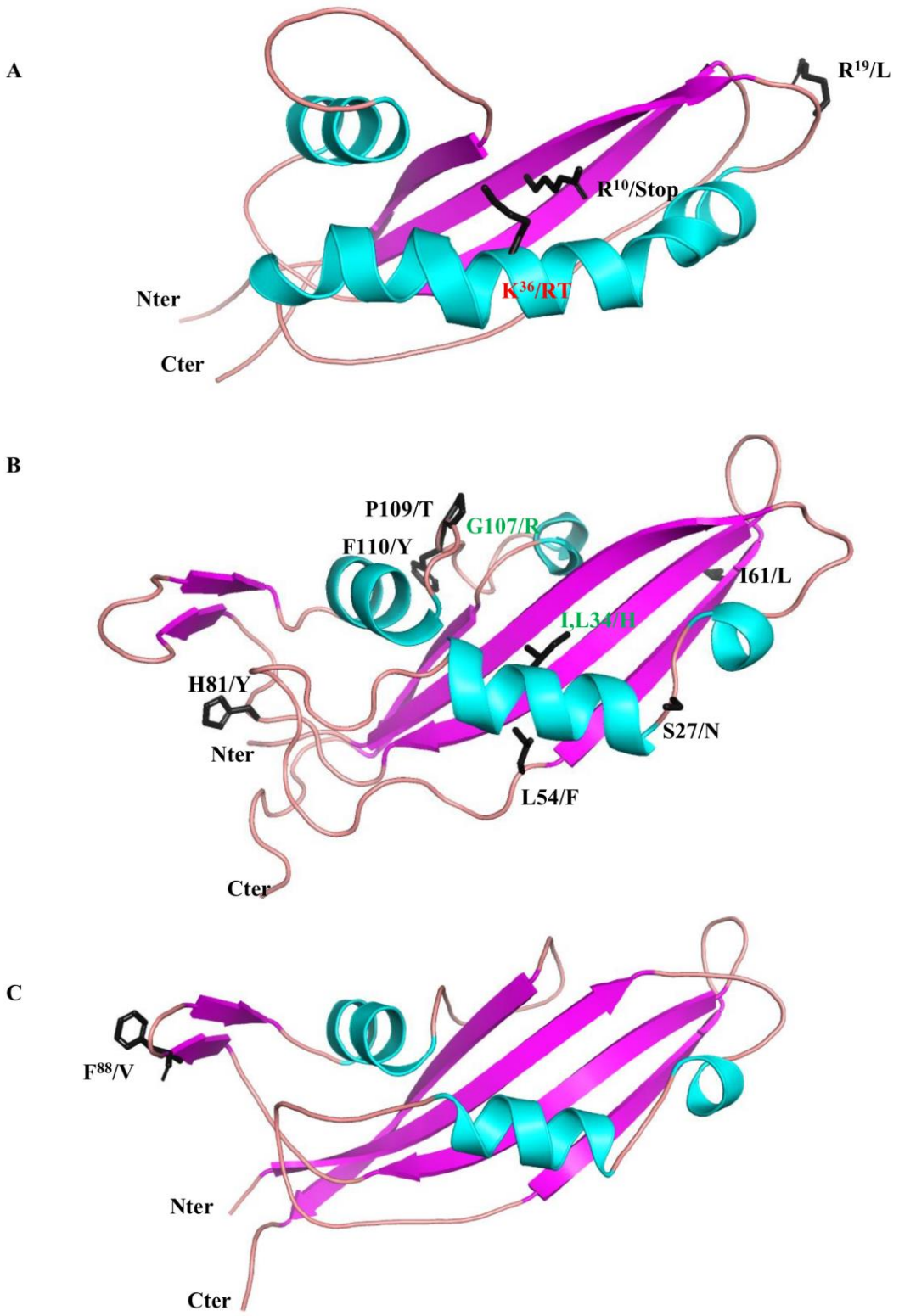

Fig 3. Localisation on the 3D structures of AvrLm5-9, AvrLm3 and Ecp11-1 of polymorphic residues.

Polymorphic residues were identified in populations of *L. maculans* 'brassicae' and *F. fulva*. (A) Polymorphic positions in AvrLm5-9 (Van de Wouw et al., 2014 and Ghanbarnia et al., 2018). Amino acid change involved in virulence toward *Rlm9* is indicated in red. (B) Polymorphic positions in AvrLm3 (Plissonneau et al., 2017). Amino acid changes shared among isolates virulent toward *Rlm3* are indicated in green. (C) Polymorphic position in Ecp11-1 (Mesarich et al., 2018). It is currently unclear if this variant impacts protein function.

These two substitutions interestingly are always coupled and likely responsible for the virulent phenotype towards *Rlm3*. They are positioned in the regions connecting the strands, and the folding brings them relatively close together in space, suggesting this site could be important for its function. Remarkably, the positions in the 3D structures of the polymorphic residues $G^{131}R$ (AvrLm3) and $G^{120}R$ (AvrLm4-7; Blondeau et al., 2015) are very similar, suggesting the same protein regions could be involved in the virulence phenotype.

From a population study on a worldwide collection of *F. fulva* strains (Mesarich et al., 2018), a single non-synonymous $F^{119}V$ substitution was identified in Ecp11-1. W^{118}, F^{119} and W^{124} (W^{87}, F^{88} and W^{93} according to protein numbering in Figure 2) form an exposed hydrophobic surface patch contiguous to a conserved WR(F/L/V)(R/K) motif (see below for more information on this motif). It is not yet known whether this non-synonymous substitution affects the ability of Ecp11-1 to trigger an HR in tomato plants carrying the putative Cf-Ecp11-1 R protein (Fig 3C).

- **Structural analogues of AvrLm4-7, AvrLm5-9 and Ecp11-1 are present in *Leptosphaeria maculans***

The strong structural similarities between AvrLm4-7, AvrLm5-9, AvrLm3 and Ecp11-1 suggest that these four proteins belong to a fungal effector family, characterized by a four-stranded β-sheet, an α-helix and three conserved disulfide bridges (Fig 2A and 2C). To identify other structurally related family members, a hidden Markov model (HMM)-based profile search was performed on the predicted protein repertoire from *L. maculans* (v23.1.3 isolate). An iterative search was carried out with each effector structure, using a cut-off E-value of 1 and an overlap cut-off of 50 %. A total of three iterations were performed. At each iteration, proteins longer than 160 amino acids were removed. Seven potential structural analogues were found for AvrLm4-7, six for Ecp11-1 and four for AvrLm5-9. The search with Ecp11-1 retrieved both AvrLm3 and AvrLm4-7, the remaining four were also found with the search using AvrLm4-7. The search with AvrLm5-9 retrieved AvrLm3 together with other candidates that were not found with AvrLm4-7 or Ecp11-1, including the avirulence protein AvrLmS-Lep2 (Table 1; Neik et al., 2020). A total of thirteen structural analogues were thus identified in *L. maculans* (Table 1).

Table 1. Characteristics of the structural analogues identified in *L. maculans*

Accession number[a]	Size (aa)	Cysteine number[b]	Localization[c]	Genomic context[c]	Signal peptide[d]	Structure allowing identification[e]			Best structural analogue[f]
						AvrLm5-9	Ecp11-1	AvrLm4-7	
Lmb_jn3_00001 (*AvrLm3*)	160	10	BAC01	TE-rich	Yes	X	X		
Lmb_jn3_01426	139	6	SC1	Gene-rich	Yes		X	X	AvrLm4-7
Lmb_jn3_01427	137	8	SC1	Gene-rich	Yes		X	X	AvrLm4-7
Lmb_jn3_01428	130	8	SC1	Gene-rich	Yes		X	X	AvrLm4-7
Lmb_jn3_03262 (*AvrLm4-7*)	143	8	SC3	TE-rich	Yes		X	X	
Lmb_jn3_03263	142	8	SC3	Gene-rich	Yes			X	AvrLm4-7
Lmb_jn3_03815	140	8	SC3	TE-rich	Yes	X			2hreD
Lmb_jn3_08343 (*AvrLmS-Lep2*)	141	8	SC9	TE-rich	Yes	X			2w53B
Lmb_jn3_08418	138	6	SC10	TE-rich	Yes			X	AvrLm4-7
Lmb_jn3_08419	174	8	SC10	TE-rich	No		X	X	AvrLm4-7
Lmb_jn3_08421	127	8	SC10	TE-rich	Yes			X	AvrLm4-7
Lmb_jn3_10106 (*AvrLm5-9*)	141	7	SC12	TE-rich	Yes	X			
Lmb_jn3_12986	143	8	SC18	TE-rich	Yes	X			2xjaA

[a]When the gene corresponds to a known avirulence gene, it is shown.

[b]The number of cysteines is calculated on the mature protein without the signal peptide.

[c]The localization of the different genes is based on the recent *L. maculans* genome assembly by Dutreux et al. (2018).

SC: Super Contig. BAC: Bacterial Artificial Chromosome. TE: Transposable Element.

[d]Signal peptide was predicted using SignalP 4.1 (Petersen *et al.*, 2011).

[e]Structure which allowed retrieval of the structural analogue through HMMER search.

[f]The best structural analogue was predicted using the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/).

The Protein Data Bank accession number is indicated.

We then wanted to confirm that the retrieved protein analogues have similar structures. We therefore predicted the 3D structure of the different analogues found by the HMM search using the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/; Yang and Zhang, 2015). For all identified hits using either AvrLm4-7 or Ecp11-1 as the template, the server proposed AvrLm4-7 as the best structural analogue. For those hits that were identified using AvrLm5-9 as template, and even using AvrLm5-9 itself, I-TASSER was not able to find AvrLm4-7 as the best structural analogue (Table 1). Structure prediction suggests that these sequences possess the same fold and belong to a family that we will name from this point forward as the LARS (for Leptosphaeria AviRulence-Suppressing) effector family.

- LARS effectors of *L. maculans* share common characteristics and are expressed *in planta*

The characteristics of LARS structural analogues identified in *L. maculans* using a HMM search are summarized in Table 1. The genomic location of the genes encoding structural analogues was investigated using the recent new assembly of the *L. maculans* genome (Dutreux et al., 2018). The majority of these genes are located in genomic regions rich in remnants of transposable elements, with the exception of a group of three neighboring genes (*Lmb_jn3_01426*, *Lmb_jn3_01427* and *Lmb_jn3_01428*), located in a gene-rich region, and a gene that had been previously described as a paralogue of *AvrLm4-7* (*Lmb_jn3_03263*; Parlange et al., 2009), which is located at the border of a gene-rich region. These genes are mostly located on different Super Contigs. However, it was found that several genes are neighbors. This is the case for (i) *Lmb_jn3_01426*, *Lmb_jn3_01427* and *Lmb_jn3_01428*; (ii) *Lmb_jn3_08418*, *Lmb_jn3_08419* and *Lmb_jn3_08421*; (iii) *Lmb_jn3_03262* (*AvrLm4-7*) and *Lmb_jn3_03263*. The average size of these proteins is 140 amino acids, and they all have between six and 10 cysteines. For the whole family, apart from Lmb_jn3_08419, we were able to predict a signal peptide, suggesting that these proteins are secreted by the fungus.

RNAseq data corresponding to different stages of infection of oilseed rape cotyledons by *L. maculans* as well as during axenic growth of *L. maculans* on V8 solid medium were recently generated (Gay et al., 2020), and expression kinetics of the LARS effector genes analyzed (Fig 4). All the genes showed over-expression during the biotrophic / asymptomatic phase of cotyledon infection. A closer look at the expression profiles distinguished three different patterns.

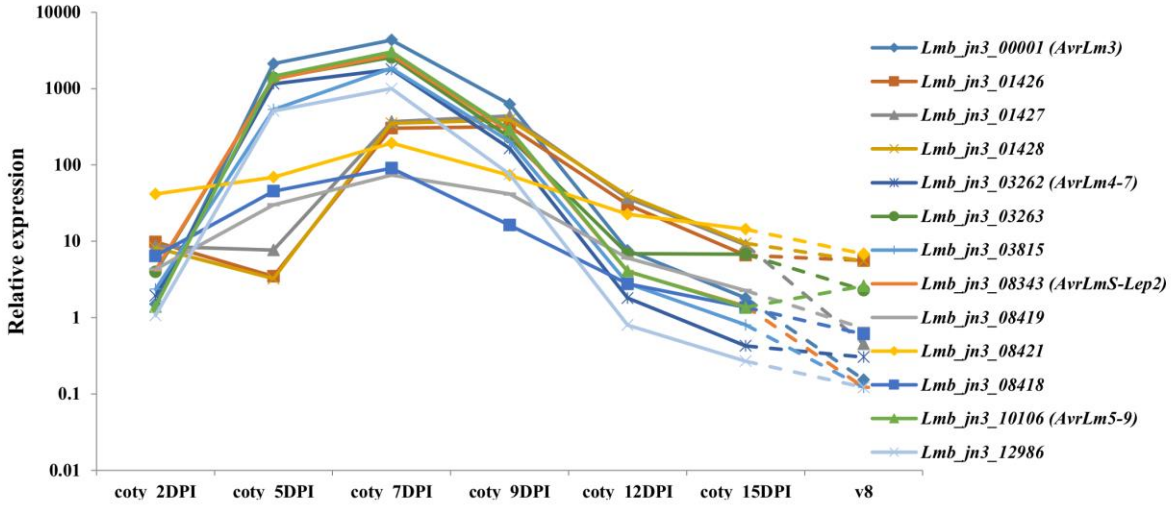

Fig 4. Expression kinetics of the LARS structural analogues identified in *L. maculans* 'brassicae'. Expression is based on RNAseq data described in Gay et al. (2020) normalized by the total number of sequences per condition, counts per million (CPM), and represented on a logarithmic scale, following inoculation of cotyledons from a susceptible oilseed rape cultivar, Darmor-Bzh. Coty (cotyledons of the susceptible oilseed rape cultivar, Darmor-Bzh infected by *L. maculans* 'brassicae' 2, 5, 7, 9, 12 and 15 days after infection (DPI)); v8 (growth on solid medium V8 for one week).

The first pattern represented genes that were highly expressed as early as 5 days post infection (DPI) with a peak at 7 DPI, and included the *AVR* genes of the family: *Lmb_jn3_00001* (*AvrLm3*), *Lmb_jn3_03262* (*AvrLm4-7*), *Lmb_jn3_03263*, *Lmb_jn3_03815*, *Lmb_jn3_08343* (*AvrLmS-Lep2*), *Lmb_jn3_10106* (*AvrLm5-9*) and *Lmb_jn3_12986*. The second pattern of genes was characterized by a low expression at 5 DPI and a plateau between 7 and 9 DPI: *Lmb_jn3_01426*, *Lmb_jn3_01427* and *Lmb_jn3_01428*. The last pattern grouped genes whose expression peaked lower at 7 days post-infection: *Lmb_jn3_08418*, *Lmb_jn3_08419* and *Lmb_jn3_08421*. Of interest, genes with the same expression profile are neighbors in the *L. maculans* genome. Finally, on V8 medium, all the genes were very lowly expressed, which indicates that they are overexpressed during the primary infection of oilseed rape by *L. maculans*.

379 **- LARS effectors are present in other phytopathogenic fungi**

380 To find out whether the LARS effector family has members in other fungi, a new HMM-based
381 profile search was performed on an in-house database, composed of annotated proteomes of
382 163 fungal strains, corresponding to 116 species, mostly Dothideomycetes and
383 Sordariomycetes with contrasting lifestyles (phytopathogens, entomopathogens, endophytes,
384 saprophytes, mycoparasites, Table S3). The in-house database was iteratively searched with
385 each effector structure file, using a cut-off E-value of 1 and a cut-off overlap of 50 %. At each
386 iteration, proteins longer than 160 amino acids were removed. This HMM-search identified,
387 after three iterations, 34 potential structural analogues using AvrLm4-7, 32 with Ecp11-1 and
388 two using AvrLm5-9. Interestingly, Ecp11-1 was found using AvrLm5-9 as a template, but not
389 using AvrLm4-7. Combined with the analogues found in *L. maculans* 'brassicae', 49 non-
390 redundant proteins were identified (Fig 5). These potential structural analogues originate from
391 13 fungal species, with the majority from species closely related to *L. maculans* 'brassicae' (*L.*
392 *maculans* 'lepidii', *L. biglobosa* 'brassicae' and *L. biglobosa* 'thlaspii'), Dothideomycetes
393 (*Macrophomina phaseolina*, *Pyrenophora tritici-repentis*, *P. teres*, *Corynespora cassiicola*,
394 *Stemphylium lycopersici* and *F. fulva*), and a few Sordariomycetes (*Colletotrichum orbiculare,*
395 *C. gloeosporioides* and *C. higginsianum*; Fig 5A, Fig S2 and Table S3).

396 A sequence logo derived from the multiple alignment of the structural analogues highlighted
397 conserved features of the LARS effectors (Fig 5B). Although the number of cysteines is
398 variable between the different members, six cysteines can be aligned between the majority of
399 the LARS members. The cysteines near the N- and C-termini form a disulfide bridge in all
400 available structures, and this is likely the case for all of the proteins identified. The remaining
401 aligned cysteines are not always involved in superimposable disulfide bridges in the three
402 available structures. All structures have a putative disulfide bridge that connects α-helix 1 to β-
403 strand 3. The third conserved cysteine pair establishes a different disulfide bridge, however, in
404 AvrLm4-7 on the one hand and AvrLm5-9, Ecp11-1 and AvrLm3 on the other (Fig 2A). Apart
405 from the cysteines, a WR(F/L/V)(R/K) sequence motif is very well conserved in all sequences
406 (Fig 2D), positioned at the exit of the third β strand. The motif on one side crosses the disulfide
407 bridge that connects the N- and C-termini, and on the other lies against the α–loop that precedes
408 the second strand. Residues of this strand are less well conserved.

409

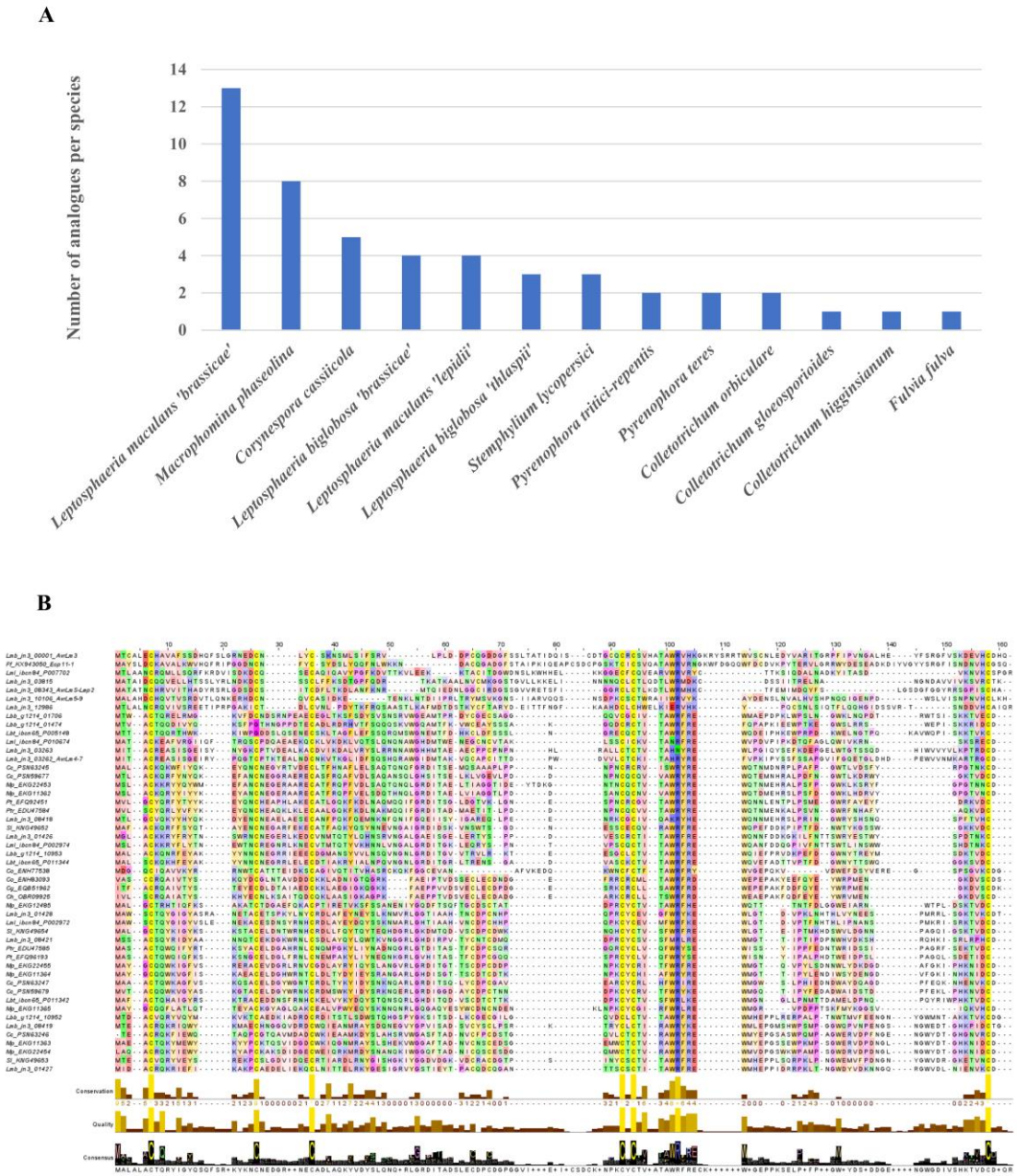

Fig 5. Identification of LARS effector structural analogues in *L. maculans* 'brassicae' and other phytopathogenic fungi.

Species distribution of number of structural analogues of AvrLm4-7, AvrLm5-9 and Ecp11-1 found with a low-stringency HMM search on a database encompassing 163 predicted proteomes of 116 fungal species. (B) Multiple sequence alignment of the 50 potential structural analogues and the sequences of the known structures (AvrLm4-7=4fprA / AvrLm5-9=0a59A / Ecp11-1=0cp1A). The secondary structures of the latter were calculated using the software STRIDE and were added at the bottom of the alignment (H=helix, G=3-10 helix, E= β strand, B=β bridge, C=coil, T=turn) above the residue conservation measure, the local alignment quality and the consensus logo. The alignment was displayed using the software Jalview. In the displayed alignment, amino acids which do not align with any of the three sequences AvrLm4-7, AvrLm5-9 and Ecp11-1 have been removed.

Cg, *Colletotrichum gloeosporioides*; Ch, *Colletotrichum higginsianum*; Co, *Colletotrichum orbiculare;* Cc, *Corynespora cassiicola*; Ff, *Fulvia fulva*; Lbb*, Leptosphaeria biglobosa* 'brassicae'; Lbt, *Leptosphaeria biglobosa* 'thlaspii'; Lmb, *Leptosphaeria maculans* 'brassicae'; Lml, *Leptosphaeria maculans* 'lepidii'; Mp, *Macrophomina phaseolina*; Pt, *Pyrenophora teres;* Ptr, *Pyrenophora tritici-repentis*; Sl, *Stemphylium lycopersici*

Remarkably, the tryptophan residue of the motif is fully exposed to the solvent. The F/L/V residue fits into the hydrophobic core of the protein. The conservation of the motif and its exposure on the surface suggest that it might be involved in a functional interaction.

- A structural analogue from *F. fulva* triggers recognition by oilseed rape R protein Rlm3, and this recognition is masked by the presence of AvrLm4-7

In an attempt to determine whether members of the LARS family from other phytopathogenic fungi could be recognized by R proteins from oilseed rape and whether that recognition could be masked by AvrLm4-7, we set out to determine whether Rlm3 can recognize Ecp11-1 of *F. fulva*, and if so, whether AvrLm4-7 can 'hide' Ecp11-1 from Rlm3-mediated recognition.

As a first stage, a construct containing the *ECP11-1* coding sequence and terminator under the control of the *AvrLm4-7* promoter was introduced via *Agrobacterium tumefaciens*-mediated transformation into two isolates virulent towards *Rlm3*, *Rlm4* and *Rlm7*: Nz-T4 and v45.15. The corresponding transformants (10 Nz-T4-*ECP11-1* and 9 v45.15-*ECP11-1*) were inoculated onto *B. napus* cvs Pixel (*Rlm4*) and 15.22.4.1 (*Rlm3*; Fig 6). All transformants, as the wild type isolates, were virulent on Pixel. In contrast, and differing from their parental isolates, all transformants were avirulent on 15.22.4.1. We conclude that Ecp11-1, as AvrLm3, is able to be recognized by Rlm3.

As a second stage, we introduced *ECP11-1* into isolate IBCN18, which is virulent towards *Rlm3* and avirulent towards *Rlm4* and *Rlm7* (carrying a functional allele of *AvrLm4*-7) in order to test whether AvrLm4-7 was able to mask Ecp11-1 recognition by Rlm3. Nine IBCN18-*ECP11-1* transformants were obtained. All transformants, as IBCN18, were avirulent on Pixel and virulent on 15.22.4.1 (Fig 6), suggesting that AvrLm4-7 is masking Rlm3-mediated recognition of Ecp11-1.

Finally, we complemented one Nz-T4-*ECP11-1* transformant (number 3) with *AvrLm4-7*. One Nz-T4-*ECP11-1-AvrLm4*-7 transformant was obtained. It was avirulent on Pixel, confirming that *AvrLm4-7* was expressed in the transformant. Contrasting with Nz-T4-*ECP11-1* transformants, the Nz-T4-*ECP11-1-AvrLm4-7* transformant was able to cause disease on cultivar 15.22.4.1, confirming that the presence of *AvrLm4-7* is masking Rlm3-mediated recognition of Ecp11-1.

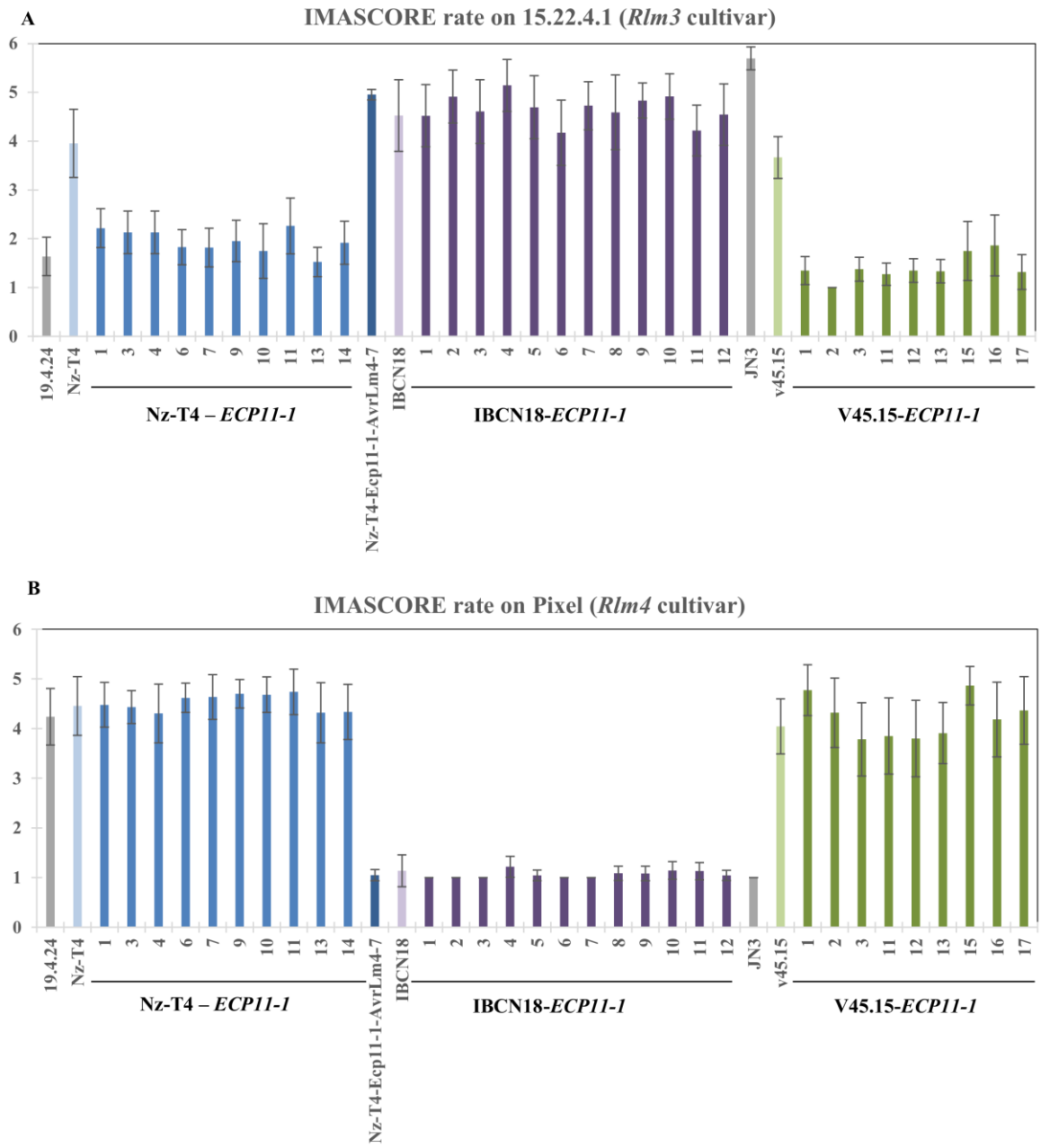

Fig 6. Ecp11-1 of *F. fulva* triggers Rlm3-mediated recognition in *L. maculans*, masked by AvrLm4-7.

Wild type isolates Nz-T4 (a3a4a7), IBCN18 (a3A4A7) and v45.15 (a3a4a7), as well as Nz-T4, IBCN18 and v45.15 transformants carrying *ECP11-1*, and Nz-T4 transformants carrying both *ECP11-1* and *AvrLm4-7* were inoculated onto cotyledons of a cultivar carrying *Rlm3* (15.22.4.1, A) or *Rlm4* (Pixel, B). Pathogenicity was measured 13 days post infection. Results are expressed as a mean scoring using the IMASCORE rating comprising six infection classes (IC), where IC1 to IC3 correspond to resistance, and IC4 to IC6 to susceptibility (Balesdent et al., 2005). Error bars indicate the standard deviation of technical replicates. 19.4.24 (A3a4a7) and JN3 (a3A4A7) were used as controls of the AvrLm3/Rlm3, and AvrLm4-7/Rlm4 and Rlm7 interaction phenotypes, respectively.

Discussion

In this study, we determined the crystal structure of *L. maculans* AvrLm5-9 and *F. fulva* Ecp11-1, and obtained a good quality model for AvrLm3 built via the crystal structure of Ecp11-1. Despite no evident sequence similarity, these three effectors are structural analogues of AvrLm4-7. All have a four-stranded β-sheet and helical connections with the same topology. The main differences reside in the conformations of the connections between the strands. Six cysteines involved in disulfide bridges are shared by the three effectors. One disulfide bridge ties together the N- and C-terminal regions, and two others connect the main helical region to the β-sheet. Structure-based pattern searches identified a large number of LARS effector candidates displaying sequence diversity, but likely sharing basic structural elements. Sequence comparison of these candidates shows the strong conservation of six cysteines, which are involved in the aforementioned structure stabilizing disulfide bridges. All of the retrieved sequence analogues are likely compatible with the structures, as confirmed by the structure prediction server I-TASSER. The alignment of the putative analogues highlights a conserved sequence patch WRXR, with X being a hydrophobic/aromatic residue. These residues are situated at the end of the third β-strand, close to the N and C termini. The tryptophan and arginine are exposed and could provide an interaction surface with plant targets. The hydrophobic X residue is involved in hydrophobic packing of this loop against the β-sheet. Interestingly, a previous functional characterization of AvrLm4-7 and its recognition by Rlm4 and Rlm7 supported a possible role of that conserved sequence patch in the translocation of AvrLm4-7 into plant cells and a role of the second arginine in the recognition by Rlm7 (Blondeau et al., 2015). Indeed, mutation of the R100 or F102 residue located close to the conserved patch resulted in the loss of Rlm7-mediated recognition. The polymorphic residues identified in AvrLm5-9 and AvrLm3 from *L. maculans* populations and on Ecp11-1 from *F. fulva* isolates are mainly located on external loops of the proteins, with the only exceptions being a few amino acid changes (putatively) involved in the switch to virulence towards cultivars with *Rlm3* or *Rlm9* resistance genes. Remarkably, the positions in the 3D structures of the polymorphic residues $G^{131}R$ in AvrLm3 and $G^{120}R$ in AvrLm4-7 are very similar, suggesting the same protein regions could be involved in the virulence phenotype.

Structure-informed pattern searches specifically identified LARS-effectors in phytopathogenic ascomycetes from the Dothideomycetes and Sordariomycetes classes. One or two LARS-effector(s) per species were detected in the Sordariomycetes *Colletotrichum* sp. Two structural

analogues were identified in the Dothideomycetes closely related to *L. maculans* 'brassicae', specifically *P. tritici repentis* and *P. teres*, while between three and four structural analogues were identified in the species from the complex comprising *L. maculans* 'brassicae', suggesting a recent expansion of the LARS family in the species complex. In *L. maculans* 'brassicae', thirteen LARS effectors could be detected and their expression during the primary biotrophic stages of oilseed rape cotyledon infection suggests they are *bona fide* effectors. They represent 14 % of the candidate effectors specifically overexpressed during the biotrophic stages of oilseed rape infection (nine LARS effectors among the 63 effector genes in Cluster 2 'biotrophy' defined by Gay et al. (2020)). The LARS family also comprises four out of the nine cloned *AVR* genes from *L. maculans*. Most of the *L. maculans* 'brassicae' LARS effectors are located in TE-rich regions (9/13) and eight are grouped in three genomic regions as neighbor genes, suggesting their expansion could be partly due to local duplications and that their location in TE-rich compartments could have led to their rapid diversification, with these regions being subjected to Repeat-Induced Point mutations (RIP; Rouxel *et al.*, 2011). An expansion of the LARS family, with 5-8 structural analogues, was also detected in two other Dothideomycetes, *M. phaseolina* and *C. cassiicola*. We conclude that LARS effectors probably have the same origin and that their expansion in some Dothideomycetes results from duplications, and, at least in the *L. maculans* 'brassicae' genome, diversification in TE-rich compartments due to RIP. Two other structural families of effectors were reported in fungi: the RALPH effectors identified in *Blumeria graminis* and the MAX effectors identified in *Magnaporthe oryzae* (RALPH for RNAse-Like Proteins Associated with Haustoria, Pedersen et al., 2012; MAX for *Magnaporthe* Avrs and ToxB like, de Guillen et al., 2015). RALPH effectors represent about 25 % of the *B. graminis* predicted effectors and three out of the four AVR effectors identified to date, and most of them are highly expressed during plant infection (Pedersen et al., 2012; Bourras et al., 2015; Lu et al., 2016; Praz et al., 2017). Pedersen et al. (2012) hypothesized that RALPH effectors originated from an ancestral gene, encoding a secreted ribonuclease, duplicated by TE-driven processes and recently diversified within the grass and cereal powdery mildew lineage. The same way, MAX effectors represent 5-10 % of the *M. oryzae* effectors and 25 % of the cloned AVR effectors, and most of them are expressed during early biotrophic stages of rice infection. De Guillen et al. (2015) hypothesized that the expansion of MAX effectors occurred in a common ancestor of *M. oryzae* and *M. grisea*. The scenario observed for the LARS, RALPH and MAX examples suggests that a wide variety of effectors, without any apparent sequence relationship, could in fact constitute a limited set of

structurally conserved effector families and that they have expanded in some fungal lineages or even in several fungal classes.

AvrLm4-7 suppresses recognition of AvrLm3 by Rlm3 and of AvrLm5-9 by Rlm9. Other cases of effectors with a suppressive function have been described in fungi. In *Fusarium oxysporum* f. sp. *lycopersici*, the Avr1 effector suppresses recognition of Avr2 and Avr3 by the I-2 and I-3 R proteins of tomato (Houterman et al., 2008). The structure of Avr2 has recently been determined (Di et al., 2017), but is unrelated to the structures of AvrLm4-7, AvrLm5-9 or AvrLm3, and the 3D structures of Avr1 and Avr3 are unknown. In the necrotrophic fungus *P. tritici repentis*, the Host Selective Toxin (HST) ToxA suppresses the activity of other HSTs (Manning and Ciuffetti, 2015). Finally, *B. graminis* secretes a suppressor of avirulence (*SvrPm3*) acting on interaction between the *AVR* gene *AvrPm3* and the barley *R* gene *Pm3* (Bourras et al., 2015). However, neither in *F. oxysporum*, in *P. tritici-repentis* nor in *B. graminis* have the mechanisms underlying suppressive function yet been determined. In non-fungal models, several mechanisms explaining suppressive interactions have been highlighted. (i) The AVR effector can act downstream of another AVR recognition by an R protein to suppress HR induction: in *Xanthomonas campestris* pv. *vesicatoria*, AvrBsT interacts in the plant cell with SnRK1, an SNF1-related kinase, to inhibit the HR induced by AvrBs1 recognition (Szczesny et al., 2010). (ii) Effectors displaying suppressive interaction can share a common plant target, but differ in their actions on that target: in *Pseudomonas syringae*, the effector proteins AvrRpm1, AvrRpt2 and AvrB target the *Arabidopsis thaliana* protein RIN4, a key regulator of plant immunity (Mackey et al., 2002; Axtell and Staskawicz, 2003). While AvrB and AvrRpm1 trigger Rpm1-mediated recognition through phosphorylation of RIN4, AvrRpt2 triggers plant immunity through cleavage of RIN4, thus preventing recognition of AvrB and AvrRpm1 by Rpm1. (iii) Suppressive effectors can directly act on the R proteins: in *Phytophthora infestans*, the IPI-O4 effector suppresses the HR triggered by recognition of IPI-O1 by the potato RB R protein. IPI-O4 interacts with the coiled-coil domain of RB which is also the domain targeted by IPI-O1 (Chen et al., 2012). Since AvrLm4-7, AvrLm3 and AvrLm5-9 share the same structural fold, we hypothesize that they could (i) target the same plant components or cellular processes and / or (ii) be recognized by the same R proteins. While we do not have any information on the plant components targeted by AvrLm3, AvrLm5-9 and AvrLm4-7, such a target could be guarded by the R proteins or, in the case of a direct interaction with the R proteins, could be Rlm9 or Rlm3 themselves. *Rlm9, Rlm4, Rlm7* and *Rlm3* are located in the same genetic cluster and could possibly be allelic (Delourme et al., 2004). *Rlm9* was recently

cloned and found to encode a wall-associated kinase-like (WAKL) protein, a newly described class of Receptor-Like Kinase (RLK) R protein (Larkan et al., 2020). Using a yeast two-hybrid (Y2H) assay, no direct interaction between the extracellular region of Rlm9 and AvrLm5-9 could be detected. However, a direct interaction between AvrLm5-9 and Rlm9 cannot be excluded since Y2H is not an optimal technique to test interaction with a membrane protein. Larkan et al. (2020) identified at the *Rlm9* locus, in other resistant accessions of oilseed rape, three WALK genes that could potentially correspond to *Rlm3*, *Rlm4* and *Rlm7*, suggesting they could be variants or duplications of the same WAKL locus. Based on the zig-zag model (Jones and Dangl, 2006), we hypothesize that *Rlm4* and *Rlm7* evolved from *Rlm3* or *Rlm9* in response to the suppressive effect of AvrLm4-7 on AvrLm3 and AvrLm5-9 recognition. We propose a model in which Rlm3 directly recognizes AvrLm3 (and Ecp11-1), and Rlm9 directly recognizes AvrLm5-9, but that this recognition only occurs after AvrLm3 and AvrLm5-9 have bound to their plant target (i.e., the region of AvrLm3 (Ecp11-1) and AvrLm5-9 recognized by Rlm3 and Rlm9, respectively, is only exposed upon interaction with their plant target). AvrLm4-7 would have a higher affinity for the same host target, thereby preventing interaction with AvrLm3 or AvrLm5-9, and thus masking their presence to Rlm3 and Rlm9. Although AvrLm4-7 binds the same host virulence target, we hypothesize that it does not possess the protein region recognized by Rlm3 and Rlm9. Instead, upon binding the plant target, AvrLm4-7 presents a protein region that is recognized by Rlm4 and Rlm7. A slightly alternative model would be that the R proteins recognize the complexes between the AVR proteins and their plant targets but not the AVR proteins or their targets on their own.

We have identified a large structural family of effectors that, in *L. maculans*, are expressed during the early stages of infection and are potentially targeted by R proteins. This structural information on effectors could be used to improve the management and durability of *R* genes in the field. Indeed, among the nine *AVR* genes identified to date in *L. maculans*, four belong to the LARS family. The corresponding *R* genes are, at least in part, present in commercial varieties currently used in the fields (*Rlm7*, *Rlm3*, *Rlm4*, *Rlm9*, *RlmS*). We hypothesize that the presence of *R* genes targeting members of the LARS family potentially exerts a selection pressure on the other members of the family, and that an efficient strategy to improve durability of *R* genes would consist in alternating or pyramiding *R* genes corresponding to different structural classes of effectors. We have also determined that Ecp11-1, a homologue of AvrLm3 and AVR effector candidate from *F. fulva*, is able to trigger *Rlm3*-mediated resistance in oilseed

rape. This finding significantly alters our understanding about the degree of host-microbe specificity as developed by Flor in the 1940s, in that this is an example of cross-species effector recognition. A next step will be to determine whether other homologues identified in Dothideomycetes and Sordariomycetes can also trigger recognition by R proteins of oilseed rape and, in the longer term, to evaluate the possible use of broad-spectrum resistances for multi-pathogen management of diseases.

Materials and Methods

Leptosphaeria maculans, bacteria and plant growth conditions

The v23.1.3 isolate (also known as JN3) belongs to the race Av1-4-5-6-7-8-10-11-S, i.e. is avirulent towards *Rlm1*, *Rlm4-Rlm8*, *Rlm10-11* and *RlmS*, and is the reference isolate whose genome was sequenced (Rouxel et al., 2011) and recently re-assembled and re-annotated (Dutreux et al., 2018). Nz-T4 is a field isolate from New Zealand, IBCN18 is a field isolate from Australia (Balesdent et al., 2005), and v45.15 is progeny from a cross between isolates v29.3.1 and 19.2.1 (Leflon et al., 2007). Nz-T4, IBCN18 and v45.15 were used as recipient isolates for genetic transformation. All fungal cultures were maintained on V8 juice agar medium, and highly sporulating cultures were obtained on V8 juice, as previously described by Ansan-Melayah *et al*. (1995).

E. coli strain DH5α (Invitrogen) was grown in LB medium at 37 °C. *Agrobacterium tumefaciens* strain C58::pGV2260 was grown in LB medium at 28 °C. Antibiotics were used at the following concentrations: rifampicin 25 µg/ml, ampicillin 50 µg/ml, kanamycin, 100 µg/ml.

Brassica napus plants were grown in growth chambers with 16 h of day (22 °C, 80 % humidity) and 8 h of night (18 °C, 100 % humidity).

Fungal transformation

A. tumefaciens–mediated transformation (ATMT) was performed on *L. maculans* as described by Gout *et al.* (2006). Transformants were selected on minimal medium supplemented with 50 µg/ml nourseothricin or 50 µg/ml hygromycin, purified by single conidium isolation and maintained on selective medium.

Inoculation tests on oilseed rape

All *L. maculans* isolates and transformants were phenotyped for their virulence towards *Rlm3* and *Rlm4* oilseed rape genotypes using a cotyledon-inoculation test (Balesdent *et al.*, 2002). Spore suspensions of each isolate or transformant were inoculated on 10-12 plants of each of the *B. napus* cvs or lines Pixel (*Rlm4*) and 15.22.4.1 (*Rlm3*). Symptoms were scored 12-21 days post-inoculation (dpi) using a semi-quantitative 1 (avirulent) to 6 (virulent) rating scale in which scores 1-3 represent different levels of resistance (from typical HR to delayed resistance) and 4-6 different levels of susceptibility (mainly evaluated by the intensity of sporulation on lesions; Balesdent *et al.*, 2005). Inoculation tests were repeated twice.

Bacterial and yeast strains and DNA manipulation

For protein production, all synthetic gene constructs were obtained from Genscript (Piscataway, USA). The DNA sequences coding for AvrLm5-9, AvrLm3 and Ecp11-1 were cloned into a pPICZαA backbone using the *Eco*RI and *Not*I restriction sites in frame with the *Saccharomyces cerevisiae* α–mating factor signal sequence, and under the control of the alcohol oxidase *AOX1* promoter that allows methanol-inducible expression in *P. pastoris*. *E. coli* strain Top10F' was used for the amplification of recombinant plasmids. *E. coli* transformants were selected on LBLS plates with 10 mg/ml tetracycline and 25 mg/L Zeocin^{TM}. The AvrLm5-9, AvrLm3 and Ecp11-1 constructs were expressed in the eukaryotic expression system *P. pastoris* strain X33 using the EasySelect *Pichia* expression kit (Invitrogen Life Technologies, cat K1740-01) according to the manufacturer's instructions. These plasmids were used to express AvrLm5-9, AvrLm3 and ECP11-1 with a N-terminal 6His-Trx-tag followed by a TEV proteolytic recognition site to remove the 6His-Trx-tag.

In order to be expressed in *L. maculans*, *ECP11-1* from *F. fulva* was placed under the control of *AvrLm4-7* promoter. The promoter of the *AvrLm4-7* gene was amplified with the primers promAvrLm4-7_Up (GTTTTGGTTAGGTTTAGGGTCT) and promAvrLm4-7_Down BamHI (GAGAGAGGATCCGTTGTTAACTGTCAAAGGGTT). It was then digested with *Nhe*I and *Bam*HI and ligated into a *Spe*I-*Bam*HI-digested pPZPNat1 vector. *ECP11-1* was amplified from its starting codon to its terminator region using primers Ecp11_ATG_EcoU (GAGAGAGAATTCATGTTGTCGTCAGCGAAGACC) and Ecp11_3UTR_XhoL (GAGAGACTCGAGCGACTCCGTAAACTAAGAGATCC) and then digested with *Eco*RI and *Xho*I. This fragment was ligated into the pPZPNat1 vector containing the *AvrLm4-7*

promoter and digested with the same enzymes in order to obtain the vector pPZPNat1-pAvrLm4-7-Ecp11-1.

A plasmid pPZPNat1 containing *AvrLm4-7* (including its promoter and terminator regions) had been constructed previously by Parlange et al. (2009). *AvrLm4-7*, from its promoter to its terminator, was transferred into the binary vector pBht2, which carries a hygromycin cassette (Mullins et al., 2001). pBht2 and pPZPNat1-AvrLm4-7 were digested by *Sac*I and *Sal*I. *AvrLm4-7* was ligated into the pBht2 vector to obtain the vector pBht2-AvrLm4-7.

Protein production using *Pichia pastoris* and purification

Proteins were produced in the *P. pastoris* expression system by fed-batch cultivation. Upon transformation of *P. pastoris* with pPICZαA-6His-Trx-AvrLm5–9, pPICZαA-6His-Trx-AvrLm3 and pPICZαA-6His-Trx-Ecp11-1, transformants were screened for zeocin resistance in shake-flask cultures. A clone secreting a protein of ~30 kDa (as judged by SDS-PAGE) was selected for large-scale production using a fed-batch mode of cultivation. Recombinant protein was collected from the supernatant, and concentrated. For purification, supernatant was applied onto Ni-NTA resin (Qiagen, Inc.) column affinity chromatography previously equilibrated with Tris–HCl buffer (20 mM Tris–HCl pH 8.0, 300 mM NaCl, 5 % glycerol). Resin was washed with 40 ml of the same buffer and the proteins were eluted using three fractions of 6 ml of the previous buffer supplemented with 100, 200 and 400 mM imidazole. Fractions containing the proteins of interest were pooled and concentrated by ultra-filtration and loaded on Superdex 75 (GE-Healthcare) equilibrated in buffer (20 mM Tris pH 8.0, 300 mM NaCl, 5 % glycerol, 0.5 mM EDTA). The purified fusion proteins were cleaved overnight at 4 °C using the TEV protease. The sample was further purified using an ion exchange column (HiTrap Q HP 5 ml) and the fraction containing the proteins of interest were loaded on a Superdex 75 (GE-Healthcare) gel filtration column equilibrated in buffer (20 mM Tris pH 8.0, 300 mM NaCl, 5 % glycerol). Purified proteins were analyzed by SDS-PAGE (14 %) under denaturing and reducing conditions and protein concentrations were determined with a NanoDrop 2000 spectrophotometer (Thermo Scientific). Finally, fractions containing the proteins of interest were concentrated to 7.2 $mg \cdot ml^{-1}$ or 6 $mg \cdot ml^{-1}$ for AvrLm5-9 and Ecp11-1, respectively. The identity of the recombinant proteins was confirmed by mass spectrometry.

Crystallization and resolution of the protein structures

Crystals for the AvrLm5-9 protein were obtained at 18 °C from a 0.1 µl:0.2 µl mixture of a 2.7 mg/ml protein solution (stored in 20 mM Tris pH 8.0, 300 mM NaCl and 5 % glycerol) with crystallization liquor composed of 0.7 M sodium acetate, 80 mM nickel sulfate, 0.1 M HEPES pH 7.5. Native crystals were soaked for 10 seconds in 1.2 M sodium acetate, 50 mM nickel sulfate, 0.1 M HEPES pH 7.5 supplemented with 0.2 M sodium iodine and then transferred to a solution composed of 1.25 M sodium acetate, 50 mM nickel sulfate, 0.1 M HEPES pH 7.5 and 30 % of glycerol before flash-cooling in liquid nitrogen. Diffraction data at 2.7 Å resolution were recorded on beam line FIP-BM30A (synchrotron ESRF, France). The space group was $P6_122$ and the asymmetric unit contained one copy of AvrLm5-9 (solvent content of the crystals was 66 %). The structure was determined by the SAD method using the anomalous signal of the iodine atoms and the automated structure solution pipeline implemented in ccp4 with the program CRANK2 (Skubak and Pannu, 2013). Data were processed using the XDS package (Kabsch, 2010). Two iodine sites were located using the program SHELXD (Schneider and Sheldrick, 2002). The iterative substructure improvement and phasing of these Iodine sites as well as phasing and hand determination was performed using the REFMAC5, PEAKMAX, MAPRO, Solomon and Multicomb programs. The model was built semi-automatically after density modification using Parrot, REFMAC5 and BUCCANEER (Cowtan, 2006). At this step, 123 residues were placed in the electron density with a FOM value of 0.85 and R/R_{free}=31/35. The model was further improved by iterative cycles of manual rebuilding using COOT (Emsley and Cowtan, 2004) and refinement using REFMAC5 program (Murshudov *et al.*, 2011). Refinement was pursued using 2.14 Å resolution native data recorded on beam line PROXIMA 2 (synchrotron SOLEIL, France). The space group native data (different from those of the I-soaked crystals) was $P4_12_12$ with one copy of AvrLm5-9 per asymmetric unit. The structure was solved by molecular replacement using the PHASER program (McCoy *et al.*, 2007) and this model was refined using REFMAC5 (R/R_{free} was 22/28 %.) The final model contains residues -2 to 121, three nickel atoms, two acetates and 72 water molecules. Statistics for data collection and refinement are summarized in Table S1. The atomic coordinates and structure factors have been deposited into the Brookhaven Protein Data Bank under the accession numbers 7B76 (AvrLm5-9 I-derivative) and 7AD5 (AvrLm5-9 native).

Crystals for the Ecp11-1 protein were obtained at 4 °C from a 0.1 µl:0.2 µl mixture of a 6 mg/ml protein solution (stored in 20 mM Tris pH 8.0, 300 mM NaCl and 5 % glycerol) and

crystallization liquor composed of 21 % of PEG550MME, 10 mM zinc sulfate, 0.09 M MES pH 6.5 and 15 % glycerol. Native crystals were flash-cooled in liquid nitrogen and 1.94 Å resolution data were collected on beam line PROXIMA 2 (synchrotron SOLEIL, France). Data were processed using the XDS package (Kabsch, 2010). The space group was $P2_12_12_1$ with one Ecp11-1 molecule per asymmetric unit (solvent content of the crystal was 57 %). The structure was determined by the SAD method using the anomalous signal of the zinc atoms and the automated structure solution pipeline implemented in ccp4 with the program CRANK2 (Skubak and Pannu, 2013). Two Zinc sites were located using the program SHELXD (Schneider and Sheldrick, 2002). The iterative substructure improvement and phasing of these zinc sites as well as phasing and hand determination was performed using the REFMAC5, PEAKMAX, MAPRO, Solomon and Multicomb programs. The model was built semi-automatically after density modification using Parrot, REFMAC5 and BUCCANEER (Cowtan 2006). The initial model contained 142 residues with a FOM value of 0.90 and R/R_{free}=26/27. The model was further improved by iterative cycles of manual rebuilding using COOT (Emsley and Cowtan, 2004) and refinement using REFMAC5 (Murshudov *et al.*, 2011). A native set data at 1.62 Å resolution was recorded on beam line PROXIMA 2 (synchrotron SOLEIL, France) and the model was refined using REFMAC5. This yielded a R/R_{free} factor of 19/20 %. Statistics for data collection and refinement are summarized in Table S1. The atomic coordinates and structure factors have been deposited into the Brookhaven Protein Data Bank under the accession numbers 6ZUQ (Ecp11-1 Zn-derivative) and 6ZUS (Ecp11-1 native).

Expression analyses using RNAseq data

RNAseq data were previously generated by Gay et al. (2020). The expression levels of structural family members were analyzed in *L. maculans* after extraction of RNAseq data corresponding to the following stages of infection: 2, 5, 7, 12 and 15 days after infection of oilseed rape cotyledons (susceptible cultivar Darmor-Bzh). Growth in V8 axenic medium was used as a control. Two biological replicates per condition were tested. We measured the number of RNAseq reads aligned on the different genes after normalization according to the library size (Counts Per Million, CPM). All the genes were detected as being differentially expressed in at least one infection condition compared to the *in vitro* growth condition on V8 solid medium, according to the criterion: Log2 (Fold change)> 2.0 and p-value <0.05.

Bio-informatics analyses

For the HMM searches, a homemade database of 163 fungal proteomes was used (the list of the proteomes used can be found in Table S2). The analyses were performed separately for AvrLm4-7, AvrLm5-9 and Ecp11-1. The HMMER search program from the HMMER package v 3.0 was used to perform four iterations with a cut-off E-value of 1 and a minimum overlap of 50 % (Potter et al., 2018). The full sequences of the proteins detected by each of these three homology searches were retrieved. Proteins longer than 160 residues were removed. The remaining sequences were aligned with the MAFFT program using parameters --localpair --maxiterate 1000. This final list of sequences was used to build diversity trees using the Neighbor-Joining method using the MEGA6 package. In order to confirm that the retrieved protein analogues have similar structures we used the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/; Yang and Zhang, 2015) to predict their 3D structure.

Signal peptide predictions were performed using SignalP version 4.1 software (Petersen et al., 2011).

Acknowledgments

Authors wish to thank all members of the "Effectors and Pathogenesis of *L. maculans*" group. We thank the beamline staff at the ESRF synchrotron (Grenoble France, beamline FIP-BM30A) and Synchrotron SOLEIL (Saclay France, beamline PROXIMA1 and PROXIMA2) for assistance and advice during data collection. David Cobessi (FIP-BM30A beamline) is gratefully acknowledged for his valuable help in the AvrLm5-9 data collection and analysis.

Funding

Y. Petit-Houdenot was funded by a "Contrat Jeune Scientifique" grant from INRA. The "Effectors and Pathogenesis of *L. maculans*" group benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). This work was supported by the French National Research Agency projects StructuraLEP (ANR-14-CE19-0019) and Ln23 (ANR-13-BS07-0007-01 to EG), the Australian Grains Research and Development Corporation (UM00050), French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INBS-05, and by funds from the Centre National de la Recherche Scientifique and the University of Paris-Saclay.

Author contribution

CM, MHB, AI, HVT and IF planned and designed the research. NL, YPH, NT, ILSG, EZ, KB, JG, BO, FB, TR and IF performed the experiments. NL, YPH, ILSG, JG, HVT and IF analysed the data. CM, YPH, KB, JG, HVT and IF drafted the manuscript, which was reviewed and revised by all authors.

References

Ansan-Melayah D, Balesdent MH, Buée M, Rouxel T. Genetic characterization of *AvrLm1*, the first avirulence gene of *Leptosphaeria maculans*. Phytopathology. 1995; 85: 1525-1529.

Axtell MJ, Staskawicz BJ. Initiation of *RPS2*-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. Cell. 2003; 112: 369-377.

Balesdent MH, Attard A, Kühn ML, Rouxel T. New avirulence genes in the phytopathogenic fungus *Leptosphaeria maculans*. Phytopathology. 2002; 92: 1122–1133.

Balesdent MH, Barbetti MJ, Li H, Sivasithamparam K, Gout L, Rouxel T. Analysis of *Leptosphaeria maculans* race structure in a worldwide collection of isolates. Phytopathology. 2005; 95: 1061–1071.

Blondeau K, Blaise F, Graille M, Kale SD, Linglin J, Ollivier B, et al. Crystal structure of the effector AvrLm4–7 of *Leptosphaeria maculans* reveals insights into its translocation into plant cells and recognition by resistance proteins. Plant J. 2015; 83: 610–624. doi:10.1111/tpj.12913.

Bourras S, McNally KE, Ben-David R, Parlange F, Roffler S, Praz CR, et al. Multiple avirulence loci and allele-specific effector recognition control the *Pm3* race-specific resistance of wheat to powdery mildew. Plant Cell. 2015; 27: 2991– 3012. DOI: 10.1105/tpc.15.00171

Brun H, Chèvre AM, Fitt BD, Powers S, Besnard AL, Ermel M, et al. Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. New Phytol. 2010; 185, 285–299.

Chen Y, Liu Z and Halterman DA. Molecular determinants of resistance activation and suppression by *Phytophthora infestans* effector IPI-O. PLOS Pathog. 2012; 8: e1002595.

Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr. 2006; 62:1002-1011. doi:10.1107/S0907444906022116

Daverdin G, Rouxel T, Gout L, Aubertot JN, Fudal I, Meyer M, et al. Genome structure and reproductive behaviour influence the evolutionary potential of a fungal phytopathogen. PLOS Pathog. 2012; 8: e1003020. doi:10.1371/journal.ppat.1003020.

Delourme R, Pilet-Nayel ML, Archipiano M, Horvais R, Tanguy X, Rouxel T, et al. A cluster of major specific resistance genes to *Leptosphaeria maculans* in *Brassica napus*. Phytopathology. 2004; 94: 578–583. doi:10.1094/PHYTO.2004.94.6.578.

Di X, Cao L, Hughes RK, Tintor N, Banfield MJ, Takken FLW. Structure-function analysis of the *Fusarium oxysporum* Avr2 effector allows uncoupling of its immune-suppressing activity from recognition. New Phytol. 2017; 216: 897-914. doi: 10.1111/nph.14733.

Dutreux F, Da Silva C, d'Agata L, Couloux A, Gay EJ, Istace B, et al. De novo assembly and annotation of three *Leptosphaeria* genomes using Oxford Nanopore MinION sequencing. Sci Data. 2018; 5: 180235.

Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60: 2126-2132. doi:10.1107/S0907444904019158

Fisher MC, Hawkins NJ, Sanglard D, Gurr SJ. Worldwide emergence of resistance to antifungal drugs challenges human health and food security. Science. 2018; 360: 739-742.

Flor HH. Inheritance of pathogenicity in *Melampsora lini*. Phytopath. 1942; 32: 653–669.

Fones HN, Fisher MC, Gurr SJ. Emerging fungal threats to plants and animals challenge agriculture and ecosystem resilience. Microbiol Spectr. 2017; 5: doi: 10.1128/microbiolspec.FUNK-0027-2016.

Franceschetti M, Maqbool A, Jiménez-Dalmaroni MJ, Pennington HG, Kamoun S, Banfield MJ. Effectors of filamentous plant pathogens: commonalities amid diversity. Microbiol Mol Biol Rev. 2017; 81:e00066-16. doi:10.1128/MMBR.00066-16

Gay EJ, Soyer JL, Lapalu N, Linglin J, Fudal I, Da Silva C, et al. Large-scale transcriptomics to dissect two years of the life of a fungal phytopathogen interacting with its host plant. bioRxiv. 2020; doi: https://doi.org/10.1101/2020.10.13.331520

Ghanbarnia K, Ma L, Larkan NJ, Haddadi P, Fernando WGD, Borhan MH. *Leptosphaeria maculans AvrLm9*: a new player in the game of hide and seek with *AvrLm4-7*. Mol Plant Pathol. 2018; 19: 1754–1764.

Gout L, Fudal I, Kuhn ML, Blaise F, Eckert M, Cattolico L, et al. Lost in the middle of nowhere: the *AvrLm1* avirulence gene of the Dothideomycete *Leptosphaeria maculans*. Mol Microbiol. 2006; 60: 67–80.

de Guillen K, Ortiz-Vallejo D, Gracy J, Fournier E, Kroj T, Padilla A. Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. PLOS Pathog. 2015; 11: e1005228. doi:10.1371/journal.ppat.1005228.

Guttman DS, McHardy AC, Schulze-Lefert P. Microbial genome-enabled insights into plant-microorganism interactions. Nature Rev Genet. 2014; 15: 797–813. doi:10.1038/nrg3748.

Huang YJ, Li ZQ, Evans N, Rouxel T, Fitt BDL, Balesdent M-H. Fitness cost associated with loss of the *AvrLm4* avirulence function in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). Eur J Plant Pathol. 2006; 114: 77-89. doi:10.1007/1-4020-4525-5-7.

Houterman PM, Cornelissen BJC, Rep M. Suppression of plant resistance gene-based immunity by a fungal effector. PLOS Pathog. 2008; 4: e1000061.

Jones JDG, Dangl JL. The plant immune system. Nature. 2006; 444: 323–329. doi:10.1038/nature05286.

Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010; 66:125-132. doi:10.1107/S0907444909047337

Larkan NJ, Ma L, Haddadi P, Buchwaldt M, Parkin IAP, Djavaheri M, et al. The *Brassica napus* wall-associated kinase-like (WAKL) gene *Rlm9* provides race-specific blackleg resistance. Plant J. 2020; Doi:10.1111/tpj.14966.

Leflon M, Brun H, Eber F, Delourme R, Lucas M, Vallee P, et al.. Detection, introgression and localization of genes conferring specific resistance to *Leptosphaeria maculans* from *Brassica rapa* into *B. napus*. Theor Appl Genet. 2007; 115: 897-906.

Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal Effectors and Plant Susceptibility. Annu Rev Plant Biol. 2015; 66: 513–545.

Lu X, Kracher B, Saur IML, Bauer S, Ellwood SR, Wise R, et al. Allelic barley MLA immune receptors recognize sequence-unrelated avirulence effectors of the powdery mildew pathogen. Proc Natl Acad Sci USA. 2016; 113: E6486–E6495. DOI: 10.1073/pnas.1612947113

Mackey D, Holt BF, Wiig A, Dangl JL. RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for *RPM1*-mediated resistance in *Arabidopsis*. Cell. 2002; 108: 743-754.

Manning VA, Ciuffetti LM. Necrotrophic effector epistasis in the *Pyrenophora tritici-repentis*-wheat interaction. PLOS ONE. 2015; 10: e0123548.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40: 658-674.

Mesarich CH, Ökmen B, Rovenich H, Griffiths SA, Wang C, Karimi Jashni M, et al. Specific hypersensitive response–associated recognition of new apoplastic effectors from *Cladosporium fulvum* in wild tomato. Mol Plant Microbe Interact. 2018; 31: 145-162.

Mullins ED, Chen X, Romaine P, Raina R, Geiser DM, Kang S. Agrobacterium-mediated transformation of *Fusarium oxysporum*: an efficient tool for insertional mutagenesis and gene transfer. Phytopathology. 2001; 91: 173–180.

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011; 67: 355-367. doi:10.1107/S0907444911001314

Neik TX, Ghanbarnia K, Ollivier B, Scheben A, Severn-Ellis A, Larkan NJ, et al. Two independent approaches converge to the cloning of a new *Leptosphaeria maculans* avirulence effector gene, *AvrLmS-Lep2*. bioRxiv. 2020; 10.02.322479; doi: https://doi.org/10.1101/2020.10.02.322479

Nováková M, Šašek V, Trdá L, Krutinová H, Mongin T, Valentová O, et al. *Leptosphaeria maculans* effector AvrLm4-7 affects salicylic acid (SA) and ethylene (ET) signalling and hydrogen peroxide (H_2O_2) accumulation in *Brassica napus.* Mol Plant Pathol. 2016; 17: 818–831.

Oliva R, Win J, Raffaele S, Boutemy L, Bozkurt TO, Chaparro-Garcia A, et al. Recent developments in effector biology of filamentous plant pathogens. Cell Microbiol .2010; 12: 705-715.

Parlange F, Daverdin G, Fudal I, Kuhn ML, Balesdent MH, Blaise F, et al. *Leptosphaeria maculans* avirulence gene *AvrLm4-7* confers a dual recognition specificity by the *Rlm4* and *Rlm7* resistance genes of oilseed rape, and circumvents Rlm4-mediated recognition through a single amino acid change. Mol Microbiol. 2009; 71: 851–863. doi:10.1111/j.1365-2958.2008.06547.x.

Pedersen C, Ver Loren van Themaat E, McGuffin LI, Abbott JC, Burgis TA, Barton G, et al. Structure and evolution of barley powdery mildew effector candidates. BMC Genomics. 2012; 13: 694. DOI: 10.1186/1471-2164-13-694

Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011; 8: 785-786.

Petit-Houdenot Y, Fudal I. Complex interactions between fungal avirulence genes and their corresponding plant resistance genes and consequences for disease resistance management. Front Plant Sci. 2017; 8: 1072

Petit-Houdenot Y, Degrave A, Meyer M, Blaise F, Ollivier B, Marais CL, et al. A two genes for one gene interaction between *Leptosphaeria maculans* and *Brassica napus*. New Phytol. 2019; 223: 397-411.

Plissonneau C, Daverdin G, Ollivier B, Blaise F, Degrave A, Fudal I, et al. A game of hide and seek between avirulence genes *AvrLm4-7* and *AvrLm3* in *Leptosphaeria maculans*. New Phytol. 2016; 209: 1613–1624.

Plissonneau C, Blaise F, Ollivier B, Leflon M, Carpezat J, Rouxel T, et al. Unusual evolutionary mechanisms to escape effector-triggered immunity in the fungal phytopathogen *Leptosphaeria maculans.* Mol Ecol. 2017; 26: 2183–2198.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R and Finn RD. HMMER web server: 2018 update. Nucleic Ac Res. 2018; 46: W200-W204.

Praz CR, Bourras S, Zeng F, Sánches-Martín J, Menardo F, Xue M, et al. *AvrPm2* encodes an RNase-like avirulence effector which is conserved in the two different specialized forms of wheat and rye powdery mildew fungus. New Phytol. 2017; 213: 1301-1314. DOI: 10.1111/nph.14372

Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucl Acids Res. 2014; 42: 320-324 - doi: 10.1093/nar/gku316.

Rocafort M, Fudal I, Mesarich CH. Apoplastic effector proteins of plant-associated fungi and oomycetes. Curr Opin Plant Biol. 2020; 56: 9-19.

Rouxel T, Balesdent MH. Life, death and rebirth of avirulence effectors in a fungal pathogen of *Brassica* crops, *Leptosphaeria maculans.* New Phytol. 2017; 214: 526–532

Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, et al. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat Commun. 2011; 2: 202.

Sánchez-Vallet A., Fouché S., Fudal I., Hartmann F.E., Soyer J.L., Tellier A., et al. The genome biology of effector gene evolution in filamentous plant pathogens. Annu Rev Phytopathol. 2018; 56: 21–40.

Schneider TR, Sheldrick GM. Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr. 2002; 58: 1772-1779. doi:10.1107/s0907444902011678

Skubák P, Pannu NS. Automatic protein structure solution from weak X-ray data. Nat Commun. 2013; 4: 2777. doi:10.1038/ncomms3777

Szczesny R, Büttner D, Escolar L, Schulze S, Seiferth A, Bonas U. Suppression of the *AvrBs1*-specific hypersensitive response by the YopJ effector homolog AvrBsT from *Xanthomonas* depends on a SNF1-related kinase. New Phytol. 2010; 187: 1058-1074.

Van de Wouw AP, Lowe RGT, Elliott CE, Dubois DJ, Howlett BJ. An avirulence gene, *AvrLmJ1*, from the blackleg fungus, *Leptosphaeria maculans*, confers avirulence to *Brassica juncea* cultivars. Mol Plant Pathol. 2014; 15: 523–530.

Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic Ac Res. 2015; 43: W174-W181.

Supporting information captions

Fig S1. Purification of the recombinant AvrLm5-9 and Ecp11-1 proteins.

(A) Size-exclusion chromatogram (Superdex 75 16/60 column; GE Healthcare); elution with 20 mM Tris pH 8.0, 300 mM NaCl, 5 % glycerol. (B) SDS-PAGE gel analysis of the peak fractions of the gel-filtration step. Protein sizes are shown in kDa.

Fig S2. Diversity of LARS effector structural analogues identified by HMM analyses in *L. maculans* 'brassicae' and other phytopathogenic fungi.

The multiple sequence alignment generated in Figure 5 was used to generate a diversity tree using the Neighbor-joining method. Branch supports are based on 1000 bootstraps and horizontal branch length reflects sequence divergence.

Cg, *Colletotrichum gloeosporioides*; Ch, *Colletotrichum higginsianum*; Co, *Colletotrichum orbiculare;* Cc, *Corynespora cassiicola*; Ff, *Fulvia fulva*; Lbb, *Leptosphaeria biglobosa* 'brassicae'; Lbt, *Leptosphaeria biglobosa* 'thlaspii'; Lmb, *Leptosphaeria maculans* 'brassicae'; Lml, *Leptosphaeria maculans* 'lepidii'; Mp, *Macrophomina phaseolina*; Pt, *Pyrenophora teres;* Ptr, *Pyrenophora tritici-repentis*; Sl, *Stemphylium lycopersici*

Table S1. Data collection and refinement statistics for crystal structures of AvrLm5-9 and Ecp11-1

Table S2. Fungal proteome databases used for the HMM analysis

Table S3. Characteristics of LARS structural analogues identified in fungi