



HAL
open science

Analyse Structurale d'effecteurs de *Magnaporthe oryzae* impliqués dans la reconnaissance de l'agent pathogène par le récepteur immunitaire du riz RGA-5

Karine de Guillen

► **To cite this version:**

Karine de Guillen. Analyse Structurale d'effecteurs de *Magnaporthe oryzae* impliqués dans la reconnaissance de l'agent pathogène par le récepteur immunitaire du riz RGA-5. Sciences du Vivant [q-bio]. Université de Montpellier, 2016. Français. <NNT : 2016MONT3502>. <tel-04712707>

HAL Id: tel-04712707

<https://hal.science/tel-04712707v1>

Submitted on 27 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

THÈSE

Pour obtenir le grade de
Docteur

Délivré par **Université de Montpellier**

Préparée au sein de l'école doctorale CBS2
Et de l'unité de recherche Centre de Biochimie
Structurale

UMR CNRS 5048 - INSERM U1054 - UM

Spécialité : **Biologie Santé**

Présentée par **Karine de Guillen**

Analyse Structurale d'effecteurs de
Magnaporthe oryzae impliqués dans la
reconnaissance de l'agent pathogène par le
récepteur immunitaire du riz RGA-5

Soutenue le 04 Juillet 2016 devant le jury composé de

M. Sébastien DUPLESSIS, DR, INRA

M^{me} Virginie GERVAIS, CR, CNRS

M. Thomas KROJ, DR, INRA

M. André PADILLA, CR, CNRS

M. Christian ROUMESTAND, PR, UM

M. Herman VAN TILBEURGH, PR, Université Paris Sud

Rapporteur

Examineur

Examineur

Directeur de thèse

Président

Rapporteur



UNIVERSITÉ DE MONTPELLIER

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

Spécialité : Biologie Santé

Ecole Doctorale CBS2

Présentée et soutenue publiquement par

Karine de Guillen

Le 04 Juillet 2016

**Analyse Structurale d'effecteurs
de *Magnaporthe oryzae*
impliqués dans la reconnaissance de l'agent
pathogène par le récepteur immunitaire du riz RGA-5**

JURY

M. Sébastien DUPLESSIS, DR, INRA	Rapporteur
M ^{me} Virginie GERVAIS, CR, CNRS	Examineur
M. Thomas KROJ, DR, INRA	Examineur
M. André PADILLA, CR, CNRS	Directeur de thèse
M. Christian ROUMESTAND, PR, UM	Président
M. Herman VAN TILBEURGH, PR, Université Paris Sud	Rapporteur

Doctorat réalisé au sein de l'équipe « Structure, Dynamique et Fonction des Biomolécules par RMN » du Centre de Biochimie Structurale, UMR CNRS 1058, INSERM 1054, UM, 29 rue de Navacelles, 34090 Montpellier, FRANCE

RÉSUMÉ

Pour promouvoir l'infection, le champignon *Magnaporthe oryzae*, phytopathogène du riz, sécrète des protéines nommées effecteurs qui vont interférer avec les processus cellulaires de son hôte. La reconnaissance de certains de ces effecteurs, appelés Avr, par des récepteurs immunitaires de la plante, nommés protéines de résistance (R) est à l'origine de la résistance.

Cette thèse a permis la caractérisation structurale par Résonance Magnétique Nucléaire de deux effecteurs d'avirulence du champignon *Magnaporthe oryzae*, AVR1-CO39 et AVR-Pia. Bien que très différents au niveau de leurs séquences primaires, la découverte inattendue d'un repliement similaire pour ces deux effecteurs et AVR-Pizt, effecteur de *M. oryzae*, ainsi que ToxB, toxine du pathogène du blé *Pyrenophora tritici-repentis* nous a permis de mettre en évidence une famille d'effecteurs appelée MAX-effecteurs (*M. oryzae* Avrs and et ToxB) principalement retrouvée chez *M.oryzae* et *M.Grisea*.

Parallèlement à cette étude, nous nous sommes intéressé à caractériser l'interaction entre l'AVR-Pia et le domaine « leurre intégré » RATX1 de la protéine de résistance RGA5. La structure d'un allèle inactif AVR-PIA-H3, qui a perdu la capacité de liaison au RATX1, a été déterminée par RMN. Ces interactions ont été caractérisées par calorimétrie et la surface de liaison du RATX1 sur l'AVR-Pia a été identifiée par RMN. Les études *in planta* de mutants de l'effecteur AVR-Pia et de domaines du RGA5 dans le riz montrent que le RATX1, bien que nécessaire n'est pas le seul domaine de RGA5 impliqué dans la reconnaissance de l'effecteur.

ABSTRACT

To promote infection, the rice phytopathogen fungus *Magnaporthe oryzae* secretes proteins, called effectors, which will interfere with cellular process in the host. The recognition of some of them (avirulence effector, or Avr) by the plant immune receptors, named resistance proteins (R), induces the resistance.

This thesis has allowed the structural characterisation, by Nuclear Magnetic Resonance, of two Avr-effectors proteins of *Magnaporthe oryzae*, AVR1-CO39 and AVR-Pia. Although very different in term of their primary sequences, the unexpected discovery that they share a similar fold with AVR-Pizt, another *M.oryzae* effector, and ToxB, a toxin of the wheat pathogen *Pyrenophora tritici-repentis*, allow us to highlight an effector family named MAX-effector (*M. oryzae* Avrs and ToxB) mostly found in *M.oryzae* et *M.Grisea*.

At the same time, we investigated the interaction between AVR-Pia and the « integrated decoy » domain RATX1 from RGA5. The structure of AVR-Pia-H3, an inactive allele of AVR-Pia which has lost binding to RATX1, was determined by NMR. These interactions were characterized by calorimetry and the binding surface of RATX1 on AVR-Pia was determined by NMR. *In planta* studies of AVR-Pia mutants and various RGA5 domains in rice show that even if the RATX1 domain is necessary, it is not the only RGA5 domain involved in the effector recognition.

Remerciements

Au moment d'écrire ces mots de remerciements, je me rends compte de tout ce chemin parcouru et de toutes les personnes qui m'ont encouragée dans cette démarche et qui m'ont soutenue tout au long de ces années de thèse.

Tout d'abord, je tiens à remercier tous les membres du jury, rapporteurs et examinateurs, d'avoir accepté de juger ce travail.

Je remercie le Docteur Catherine Royer, alors directrice du CBS quand j'ai entrepris ce projet de thèse, d'avoir accepté et de m'avoir encouragée dans cette voie.

Je remercie également Pierre Emmanuel Milhiet, actuel directeur du laboratoire, pour m'avoir soutenue dans l'aboutissement de cette thèse.

Je remercie le Professeur Christian Roumestand, mon chef d'équipe, pour sa confiance et ses encouragements au cours de ces années. Merci pour tous tes conseils avisés, tu m'as « presque tout appris ». Merci pour ton soutien tout au long de cette aventure, notamment pendant l'épreuve difficile de la rédaction de ce manuscrit.

J'aimerais remercier particulièrement mon directeur de thèse André Padilla sans qui ce travail n'aurait pas pu aboutir. André, merci pour ta gentillesse durant toutes ces années, ta patience, ta grande pédagogie et ton extrême rigueur scientifique. Toutes ces qualités font de toi un belle personne auprès de qui j'ai énormément appris. Merci pour tes encouragements, pour tous tes conseils scientifiques ou autres, et pour le temps que tu as passé pour les corrections de ce manuscrit qui a été si compliqué pour moi à écrire. J'espère continuer longtemps à travailler à tes côtés.

Ce travail est le fruit d'une superbe collaboration avec Thomas Kroj du BGP1 et de son étudiante Diana Ortiz. Thomas, merci de t'intéresser à la biologie structurale car sans cet intérêt ce projet n'aurait jamais existé, merci pour ta gentillesse et pour tous les conseils que tu as pu me donner. Diana merci pour toutes ces conversations et ces échanges que nous avons eus sur ce projet et sur l'avenir. Bientôt ce sera ton tour d'écrire ces lignes, alors courage. Je te souhaite une belle réussite scientifique et personnelle.

Je remercie Jérôme Gracy, qui s'est penché activement sur la partie bioinformatique du projet, merci pour ton efficacité et ton aide précieuse.

Je tenais à remercier tous les membres de mon comité de thèse et particulièrement Nathalie Declerck pour tous ses précieux conseils notamment en biochimie. Nathalie c'est un grand bonheur de partager la paillasse avec toi, merci pour ton sourire et ta disponibilité. Merci pour toutes les idées que nous avons pu partager, j'en ai perdu le compte depuis longtemps.

Je remercie Diego Cattoni pour son aide pour la réalisation des expériences d'anisotropie, pour sa disponibilité et ses conseils.

Je remercie les membres de mon équipe que je n'ai pas encore cités, Yingsang Yang, Hélène Déméné et Philippe Barthe. Yang merci pour le temps passé devant le spectro, maintenant je sais ce qu'est un bon shim. Hélène, merci pour tes encouragements et pour les discussions que l'on a eues autour du café.

Merci à toi, Philou, de m'avoir supportée dans le même bureau et d'avoir supporté mes questions et sollicitations fréquentes lors de l'attribution ou lorsque j'avais de petits problèmes de code, maintenant je connais l'intérêt d'un oreadme...ainsi que pour tes conseils avisés pour le tri des nOes et les calculs de structures. Merci pour ta gentillesse, ton soutien et ton aide durant cette thèse et merci pour les pauses café.

Je tenais également à remercier Martin Cohen-Gonsaud et Angélique Devich pour leur aide et la bonne humeur qu'ils font régner dans la bioch 1. Martin, merci pour tous tes conseils et ta disponibilité, merci d'avoir toujours cru à ce projet. Merci à toi Angélique, pour ta gentillesse, ton aide et ta sympathie. Je te souhaite un bel avenir et une belle carrière.

Merci à Florence et à Didier, pour leur sympathie et leur soutien tout au long de ces années, c'est toujours un grand plaisir de déjeuner avec vous.

Un grand merci à mes voisins du bureau B17 pour leur bonne humeur et leurs encouragements dans l'aboutissement de ce projet, Jérôme et Manouk merci.

Merci à tous les autres membres du laboratoire pour toutes les discussions autour du café, merci pour cette bonne tranche de vie au laboratoire. Je ne regrette pas d'être là et compte bien y rester encore un long moment.

Enfin, un grand merci à mes amis qui se sont toujours intéressés à mon projet ; même si pour certains ce que je fais paraît toujours très abstrait, merci pour votre soutien.

Bien sûr, je souhaite également remercier ma famille, ma grand-mère, mes parents et mes sœurs, qui m'ont toujours encouragée et soutenue dans cette démarche. Merci à ma maman qui a pris en charge toute l'intendance de ma maison pour que je puisse finir de rédiger ce manuscrit dans les meilleures conditions.

Je dédie cette thèse à mon mari Denis ainsi qu'à mes deux merveilleux enfants Lucie et Paul pour tout l'amour et l'équilibre qu'ils m'apportent au quotidien.

ABBREVIATIONS

1D,2D,3D	Spectres à 1,2,3 dimensions
BIC	Biotrophy Interfacial Complex
BLAST	Basic Local Alignment Search Tool
<i>C. fulvum</i>	<i>Cladosporium fulvum</i>
CC	Coiled-Coil
CC-NB-LRR	Coiled-Coil Nucleotide Binding site Leucin Rich Repeat
CNL	Coiled-Coil Nucleotide Binding site Leucin Rich Repeat
CNS	Crsitallography and NMR System
DAMP	Damage Associated Molecular Pattern
Da	Dalton : poids moléculaire (1Da ≈ 1g/mol)
DO	Densité Optique
EDTA	Ethylène diamine tétra-acétate
ETI	Effector Triggered Immunity
ETS	Effector Triggered Suceptibility
<i>E.coli</i>	<i>Escherichia coli</i>
FPLC	Fast Pressure Liquid Chromatography
<i>H. arabidopsidis</i>	<i>Hyalpoperonospora arabidopsidis</i>
HMA	Heavy Metal Associated
HMM	Hidden Markov Model
HR	Hypersensitive Response
HSQC	Heteronuclear Single Quantum Coherance
Hz	Hertz
IPTG	IsoPropylThioGalactoside
ITC	IsoThermal Calorimetry
IUPAC	International Union of Pure Applied Chemistry
K_d	Constante de dissociation
LRR	Leucin Rich Repeat
mg	milligramme
min	minutes
mL	mililitre
μ M	micromolaire
<i>M. lini</i>	<i>Melampsoro lini</i>
<i>M. oryzae</i>	<i>Magnaporthe oryzae</i>
MAMP	Microbe Associated Molecular Pattern
NiNTA	Nickel NitriliTriAcitic Acid
<i>N. benthamiana</i>	<i>Nicotiana benthamiana</i>
<i>N. tabacum</i>	<i>Nicotiana tabacum</i>

nOe	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Enhancement Spectroscopy
NB	Nucleotide Binding
NB-ARC	Nucleotide Binding adaptor shared by Apaf-1, resistance proteins and CED-4
NLR	protéine NB-LRR
PAMP	Pathogen Associated Molecular Pattern
PRR	Pattern Recognition Receptor
PTI	PAMP Triggered Immunity
PAGE	PolyAcrylamide Gel Electrophoresis
<i>P. infestans</i>	<i>Phytophthora infestans</i>
<i>P. sojae</i>	<i>Phytophthora sojae</i>
<i>P. syringae</i>	<i>Phytophthora syringae</i>
<i>PMV</i>	<i>Potato Mosaic Virus</i>
PDB	Protein Data Bank
ppm	partie par million
QTL	Quantitative Trait Loci
RLK	Receptor Like Kinase
RLP	Receptor Like Protein
R	Résistance ou Résistant
RMN	Résonance Magnétique Nucléaire
RATX1	Related to ATX1
RGA	Resistance Gene Analog
RLK	Receptor Like Kinase
RLP	Receptor Like Protein
rpm	Rotation par minute
STAND	Signal Transduction ATPases with Numerous Domains
SDS	Sodium DodécylSulfate
TALOS	Torsion Angle Likelihood Obtained from Shift and sequence similarity
TOCSY	Total Correlation Spectroscopy
TIR	Toll/Interleukine 1-like Receptor
TIR-NB-LRR	Toll/Interleukine 1-like Receptor Nucleotide Binding site Leucin Rich Repeat
TNL	Toll/Interleukine 1-like Receptor Nucleotide Binding site Leucin Rich Repeat

TABLE DES FIGURES et DES TABLEAUX

Fig i.1 Les différents niveaux de reconnaissance du pathogène par la plante.....	5
Fig i.2 Modèle en zig-zag : une réponse immunitaire induite à deux niveaux.....	9
Fig i.3 Interaction compatible et incompatible, illustration de la relation gène pour gène	10
Fig i.4 Architecture des protéines NLR.....	12
Fig i.5 Modèles de structures des différents domaines trouvés chez les protéines NB-LRR.....	13
Fig i.6 Structures des domaines CC.....	15
Fig i.7 Architecture du domaine NB-ARC.....	17
Fig. i.8 Modèle d'activation des protéines R.....	19
Fig.i.9 Structure du LRR.....	20
Fig i.10 Structures permettant la sécrétions des effecteurs chez les Champignons et les Oomycètes.....	22
Fig i.11 Structures d'effecteurs de type RxLR.....	23
Fig i.12 Modèles de détection de l'effecteur par les protéines NB-LRR.....	25
Fig i.13 Symptômes de la pyriculariose provoqué par <i>M.oryzae</i>	29
Fig i.14 Cycle de vie de <i>M.oryzae</i>	31
Fig i.15 Description d'une cellule végétale.....	32
Fig i.16 Schéma de l'infection d'une cellule végétale par <i>M.oryzae</i>	32
Fig i.17 Modèle de sécrétion d'effecteurs par <i>M.oryzae</i>	35
Table i.1 Gène de résistance cloné chez le riz et effecteur avirulent de <i>M.oryzae</i> correspondant.....	37
Tableau 1.1 Listes d'effecteurs dont la structure a été déterminée.....	41
Tableau 1.2 Tableau des effecteurs d'avirulence connus chez <i>M.oryzae</i>	43
Fig 1.1 Structure d'AvrPiz-t (2LW6)	44
Fig 1.2 Déplacement chimique de référence pour les 20 des acides aminés.....	45
Fig 1.3 Stratégie d'attribution en RMN ¹³ C/ ¹⁵ N.....	47
Fig 1.4 Système de SPIN.....	48
Fig 1.5 Représentation des angles de torsions φ et ψ sur une chaîne polypeptidique.....	48

TABLE DES FIGURES et DES TABLEAUX

(suite)

Fig 2.1 Modèle de fonctionnel de RGA4/RGA5	117
Fig 2.2 Schéma représentant les protéines recombinantes des différents domaines du C-ter de RGA5.....	119
Fig 2.3 Mesure d'anisotropie de fluorescence.....	121
Fig M1. Plasmide et séquence des effecteurs recombinants.....	192
FigM 2. Protocole pour mini tests d'induction et stocks glycérols.....	193
Table M1. Milieux de culture.....	194
Table M2. Protocole de préparation des solutions et des gels de polyacrylamide pour l'électrophorèse.....	196
Fig M3. Plasmide et séquences des différents construits de RGA5 recombinants.....	200
Fig M4. Plasmide et séquence du RATX1 recombinant.....	205
Fig M5. Courbe de calibration BCA.....	207
Fig M6 Schéma du calorimètre.	208

TABLE DES MATIERES

<i>INTRODUCTION</i>	1
<i>A. SPECIFICITE DE L'IMMUNITE VEGETALE</i>	2
1. GENERALITES SUR L'IMMUNITE	2
1.1 L'IMMUNITE CHEZ LES PLANTES ET LES ANIMAUX	2
2. L'IMMUNITE DANS LE REGNE VEGETAL	4
2.1 LES BARRIERES PHYSIQUES	6
2.2 LA REPONSE IMMUNITAIRE INDUITE	6
2.2.1 REPONSE IMMUNITAIRE A LARGE SPECTRE (PTI/MTI)	7
2.2.2 CONTOURNEMENT DE LA REPONSE PRIMAIRE PTI/MTI, PAR LE PATHOGENE	9
2.2.3 REPONSE IMMUNITAIRE SPECIFIQUE DECLENCHEE PAR LES EFFECTEURS ETI.....	9
2.2.4 CONTOURNEMENT DE L'ETI PAR LE PATHOGENE	11
2.3 LES PRINCIPAUX ACTEURS DE L'ETI	12
2.3.1 LES PROTEINES DE RESISTANCES R.....	12
2.3.1.1 Caractéristiques des NLR.....	12
• Domaines N-terminaux TIR et CC	14
• Domaine NB-ARC	16
• Domaine LRR.....	17
2.3.2 LES EFFECTEURS	20
2.3.2.1 Caractéristiques des effecteurs	21
2.3.3 LES MODELES DE RECONNAISSANCE ENTRE EFFECTEUR ET PROTEINE R	24
2.3.4 SIGNALISATION EN AVAL DE LA RECONNAISSANCE DE L'EFFECTEUR.....	26
<i>B. LE PATHOSYSTEME RIZ-MAGNAPORTHE ORYZAE</i>	<i>27</i>
1. LE RIZ	27
2. LA PYRICULARIOSE	30
2.1 CYCLE INFECTIEUX DE <i>MAGNAPORTHE ORYZAE</i>	30
2.2 LES EFFECTEURS DU PATHOGENE <i>M. ORYZAE</i>	33
• Effecteur de type Avr	33
• Effecteur de type non-Avr	34
• Localisation, sécrétion et translocation.....	34

• Modifications post-traductionnelles.....	36
2.3 LES PROTEINES R CHEZ LE RIZ	36
3. PROJET DE THESE	39
<i>CHAPITRE 1.....</i>	<i>40</i>
1. INTRODUCTION	41
1.1 LES EFFECTEURS DE <i>M. ORYZAE</i>	42
1.2 CONNAISSANCES STRUCTURALES RECENTES (AVRPIZ-T)	43
2. METHODES	45
3. ARTICLE 1 : STRUCTURE ANALYSIS UNCOVERS A HIGHLY DIVERSE BUT STRUCTURALLY CONSERVED EFFECTOR FAMILI IN PHYTOPATHOGENIC FUNGI.....	50
<i>CHAPITRE 2.....</i>	<i>115</i>
1. INTRODUCTION	116
1.1 LE COUPLE DE RECEPTEUR IMMUNITAIRE RGA4 - RGA5	116
1.2 LE DOMAINE RATX1.....	118
2. METHODES.....	118
2.1 LES INTERACTIONS BIOLOGIQUES.....	119
2.2 TITRAGE PAR ANISOTROPIE DE FLUORESCENCE.....	120
2.3 TITRAGE PAR RMN	121
2.4 TITRAGE PAR CALORIMETRIE.....	122
3. ARTICLE 2 : RECOGNITION OF <i>MAGNAPORTHE ORYZAE</i> EFFECTOR AVR-PIA BY THE DECOY DOMAIN OF THE RICE NLR IMMUNE RECEPTOR RGA5	122
<i>DISCUSSION ET PERSPECTIVES.....</i>	<i>180</i>
1. DISCUSSION.....	181

1.1	LA FAMILLE D'EFFECTEURS MAX	181
1.2	CARTOGRAPHIE AVR PIA-RATX1	184
2.	PERSPECTIVES	186

MATERIELS ET METHODES.....190

1. MATERIEL BIOLOGIQUE.....191

1.1 SOUCHES ET PLASMIDES191

2. TECHNIQUES DE BIOCHIMIE193

2.1 PRODUCTION.....193

2.2 EXTRACTION DE LA FRACTION PERIPLASMIQUE195

2.3 PURIFICATION DES PROTEINES195

2.3.1 CHROMATOGRAPHIE PAR AFFINITE.....195

2.3.2 CHROMATOGRAPHIE PAR FILTRATION SUR GEL.....195

2.3.3 COUPURE A LA PROTEASE TEV195

3. ANALYSES BIOCHIMIQUES196

3.1 ELECTROPHORESE PAR GELS SDS-PAGE196

3.2 DOSAGE STANDARD197

3.3 DETERMINATION DES PONTS DIS-SULFURES.....197

1. MATERIEL BIOLOGIQUE.....199

1.1 SOUCHES ET PLASMIDES199

2. TECHNIQUES DE BIOCHIMIE199

2.1 PRODUCTION.....199

2.2 EXTRACTION DE LA FRACTION SOLUBLE.....201

2.3 PURIFICATION DES PROTEINES201

2.3.1 CHROMATOGRAPHIE D’AFFINITE201

2.3.2 CHROMATOGRAPHIE PAR FILTRATION SUR GEL.....201

2.4 COUPURE A LA PROTEASE 3C.....202

3. ETUDE D’INTERACTION PAR ANISOTROPIE DE FLUORESCENCE.....202

3.1 ANISOTROPIE DE FLUORESCENCE.....202

3.1.1	MARQUAGE A L'ALEXA 488 SUCCI-ESTER.....	203
3.1.2	TESTS D'INTERACTION PAR ANISOTROPIE DE FLUORESCENCE.....	203
1.	MATERIEL BIOLOGIQUE.....	204
1.1	SOUCHES ET PLASMIDES	204
2.	TECHNIQUES DE BIOCHIMIE	204
2.1	PRODUCTION.....	204
2.2	EXTRACTION DE LA FRACTION SOLUBLE.....	205
2.3	PURIFICATION DES PROTEINES	205
2.3.1	CHROMATOGRAPHIE PAR AFFINITE.....	205
2.3.2	CHROMATOGRAPHIE PAR FILTRATION SUR GEL	206
3.	ANALYSES BIOCHIMIQUES	206
3.1	DOSAGE BCA	206
4.	ETUDE D'INTERACTION PAR TITRAGE CALORIMETRIQUE ISOTHERME	207
4.1	TITRAGE CALORIMETRIQUE ISOTHERME (ITC)	207
4.1.1	PRINCIPE DE LA MESURE	208
4.1.2	PREPARATION DES ECHANTILLONS	209
4.1.3	MESURE ET INTERPRETATION.....	210
	<i>BIBLIOGRAPHIE.....</i>	<i>211</i>

INTRODUCTION

A. Spécificité de l'immunité végétale

1. Généralités sur l'immunité

En biologie, l'immunité est définie comme la capacité des organismes vivants à se défendre contre les agents étrangers, en particulier les agents infectieux (virus, champignons, bactéries...). Un corollaire est la propriété d'un système immunitaire pour discriminer le « soi » qui doit être préservé, du « non-soi » qui doit être détruit.

Tout organisme est en premier lieu protégé par des barrières physiques. Pour les Végétaux, cette barrière se caractérise par l'écorce ou la paroi des cellules végétales. Pour les Animaux, les cellules épithéliales, les muqueuses ou encore l'acidité gastrique agissent comme une barrière protectrice.

Si ces barrières sont franchies, deux types d'immunités se distinguent, l'immunité innée et l'immunité acquise.

L'immunité dite innée ou héréditaire est la première ligne de défense contre l'invasion des micro-organismes. La pénétration d'agents pathogènes implique généralement une reconnaissance du « non-soi » permettant la réponse immunitaire, évitant ainsi la propagation du pathogène.

L'immunité acquise ou spécifique permet l'établissement d'une « mémoire immunitaire ». Présente uniquement chez les vertébrés en plus de l'immunité innée, elle permet une réponse immunitaire plus efficace et plus rapide en cas de réinfection. Cette réponse, hautement spécifique, est dirigée contre des antigènes. Ce sont les lymphocytes (globules blancs) qui sont impliqués dans l'immunité acquise. Les lymphocytes B produisent des anticorps dirigés contre un antigène, c'est la réponse immunitaire adaptative humorale. Les lymphocytes T peuvent reconnaître de façon spécifique les cellules infectées par des agents pathogènes et leur destruction représente la réponse immunitaire adaptative cellulaire. Après l'infection, des cellules mémoires provenant des lymphocytes permettront à l'organisme de se souvenir de l'antigène (Spoel and Dong, 2012). C'est sur ce principe qu'est fondée la vaccination.

1.1 L'immunité chez les Plantes et les Animaux.

Il y a des différences fondamentales entre le système immunitaire des plantes et celui des animaux. Les plantes ne possèdent pas de système immunitaire adaptatif ni de cellules spécialisées dans la réponse immunitaire. Chez les vertébrés, l'immunité adaptative est responsable de la

spécificité de la réponse de défense en générant une vaste gamme de récepteurs spécifiques à l'antigène, ainsi qu'en régissant l'expansion clonale des cellules effectrices (lymphocytes) spécifiques d'antigène ; mais aussi de la formation de cellules de mémoires « à longue vie » qui produisent des récepteurs avec la même spécificité face à l'antigène, permettant une réponse immunitaire correspondant à l'antigène plus rapide et plus efficace (Spoel and Dong, 2012).

Cependant, n'ayant qu'une immunité innée, les plantes ont la capacité, au niveau de chacune de leur cellules, de reconnaître et d'induire des réponses de résistance face à l'invasion d'agents pathogènes divers (Maekawa et al., 2011a). Un élément clé du système immunitaire des plantes est la présence de deux classes de récepteurs pour la perception du non-soi qui peuvent déclencher des réponses de résistance.

La première classe implique la reconnaissance des « motifs moléculaires associés aux pathogènes » PAMP (pathogen associated molecular patterns) plus récemment définis comme MAMP (microbes associated molecular patterns) constitués, par exemple, par des liposaccharides, le peptidoglycane ou la flagelline bactérienne (Iriti and Faoro, 2007; Nummerger et al., 2004). La capacité à détecter un motif particulier est habituellement conservée à l'intérieur d'une espèce végétale et entre différentes espèces de plantes. Cette reconnaissance extracellulaire est réalisée par des récepteurs de reconnaissance de formes PRR (pattern-recognition receptor). Ces récepteurs sont constitués d'un domaine transmembranaire et d'un domaine de récepteur extracellulaire variable. Comme par exemple le domaine LysM qui reconnaît des fragments de chitine ou des peptidoglycanes ou des domaines contenant des répétitions riches en leucines (LRR) impliqués dans la reconnaissance de motifs peptidiques (Böhm et al., 2014). Souvent, ces récepteurs comportent également un domaine kinase impliqué dans la signalisation en aval et sont alors appelés RLK pour receptor like kinases. Sinon, ils sont nommés RLP pour receptor like proteins. Les réponses déclenchées par ces récepteurs, comprennent des flux ioniques, la production d'espèces d'oxygène actives, l'activation de cascade de kinases de la classe de MAPK, des changements transcriptionnels et la production de composés anti-microbiens, comme des protéines de défense tel que les défensines ou des protéines PR (pathogenesis-related proteins) et des métabolites secondaires nommés phytoalexines (Spoel and Dong, 2012). Chez les animaux, les PRR sont présents à la surface des cellules spécialisées (macrophages, cellules dendritiques) qui sont retrouvées au niveau du système de circulation sanguine. Leur reconnaissance implique l'activation des réponses immunitaires comme la phagocytose ou la réaction inflammatoire (Nüumberger et al., 2004).

La seconde classe permet la détection de facteurs de virulence aussi nommés protéines effectrices ou effecteurs. Ces effecteurs sont souvent spécifiques d'une espèce d'agent pathogène et polymorphes entre différents isolats. La reconnaissance des effecteurs est surtout assurée par des protéines, de la famille des récepteurs NLR (Nucleotide-Binding and Leucine-Rich Repeat domain Proteins), qui détectent l'action ou la structure des effecteurs de l'agent pathogène à l'intérieur des cellules hôtes. Les NLR végétales sont souvent polymorphes au sein de l'espèce de plante et l'ensemble des gènes codant les NLR dans une population hôte définit le répertoire pour la détection

des effecteurs. Au contraire, la plupart des NLR animales sont activées par la détection de MAMP ou de substances endogènes libérées après l'invasion par un agent pathogène, appelé « dommages associés au motifs moléculaires » DAMP (Damage associated molecular pattern). Dans ce sens, les protéines NLR animales sont fonctionnellement analogues aux protéines membranaires PRRs animales, comme les Toll-like récepteurs (TLR) qui présentent un domaine extracellulaire LRR (Maekawa et al., 2011a).

Ainsi, bien qu'au niveau moléculaire il existe des récepteurs immunitaires présentant des similitudes chez les animaux et les plantes, comme par exemple des domaines avec des répétitions riches en leucines (LRR), les mécanismes impliqués dans cette immunité déclenchée par des motifs moléculaires sont différents (Ausubel, 2005; Spoel and Dong, 2012). En conséquence, bien que des protéines équivalentes soient utilisées, elles ne fonctionnent pas nécessairement de la même façon. Ainsi, les plantes et les mammifères ont des récepteurs analogues au niveau de la membrane cellulaire qui détectent la flagelline (protéine bactérienne) - FLS2 (dans la plante *Arabidopsis thaliana*) et TLR5 (humain). Cependant les deux récepteurs détectent différentes régions de la flagelline. (Maekawa et al., 2011a). Les agents pathogènes sont une menace omniprésente, influençant l'évolution des organismes, l'hypothèse d'un ancêtre commun est encore débattue. Le dernier ancêtre commun des plantes et des animaux se situe il y a environ un milliard d'années. Les plantes et les animaux ont pu échanger leurs gènes mais la plupart du temps ils ont contourné la pression de sélection indépendamment (Ronald and Beutler, 2010).

En conclusion, si d'une part les plantes et les animaux diffèrent en terme d'immunité adaptative, ils conservent d'autre part la même capacité à discerner un agent pathogène de façon générale (voie des PAMP, MAMP) ou spécifique (voie des effecteurs), bien que cela se fasse par des mécanismes différents (Iriti and Faoro, 2007). Par conséquent, la principale caractéristique de la réponse immunitaire responsable de la survie des plantes et des animaux via l'évolution, repose sur la meilleure spécificité entre l'hôte et son pathogène.

2. L'immunité dans le règne végétal

A première vue l'immunité chez les plantes semble assez simple. Les végétaux apparaissent souvent comme des organismes figés, inertes, voir fragiles. Cependant ils sont assaillis par une multitude d'organismes pathogènes comme les insectes, les bactéries, les virus ou encore les champignons. Dans certains cas la maladie se développe pouvant entraîner de grosses pertes économiques. Malgré tout, la maladie reste l'exception dans le règne végétal, la plante ayant élaboré un système de défense efficace basé sur différents modes de reconnaissance des agents pathogènes.

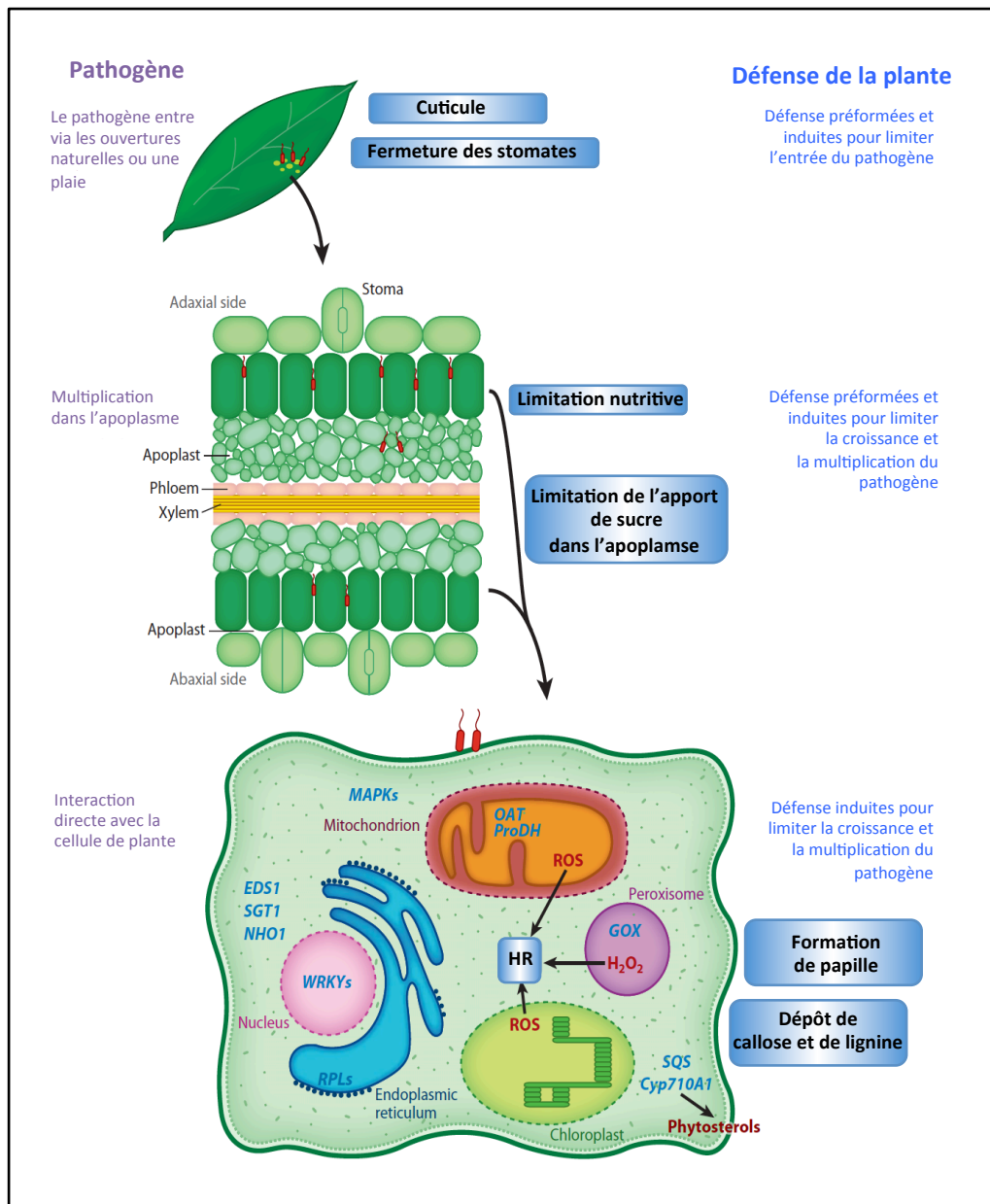


Fig i.1 Les différents niveaux de reconnaissance du pathogène par la plante :

Les premières barrières de défenses rencontrées par les agents pathogènes sont des barrières physiques comme la cuticule de la feuille composée de cire. Le pathogène va essayer de pénétrer dans la plante par des voies d'accès facilitées comme les stomates ou une plaie.

Le pathogène se retrouve ensuite dans l'apoplasme ou espace extracellulaire. A ce niveau la plante dispose de récepteurs transmembranaires reconnaissant des motifs associés à différents pathogène (PAMP). Cela permet à la plante d'induire une réponse de défense dite basale pour éviter la multiplication et la croissance du pathogène.

Le pathogène va produire des protéines effecteurs dont l'un des rôles est d'inhiber cette première réponse et de promouvoir l'infection. L'interaction entre ces effecteurs et la cellule de plante conduit à une réponse hypersensible (HR) ou à d'autres voies de défenses (libération d'H₂O₂, réponse oxydative ROS) conduisant à l'arrêt du développement du pathogène.

(Adapté de Senthil-Kumar and Mysore, 2013).

Tout d'abord, des barrières permanentes préformées offrent à la plante une protection minimale contre toute intrusion (Fig. i.1). En cas de pénétration du pathogène s'ensuit une réponse immunitaire induite à deux niveaux. Le premier niveau de détection, basé sur la reconnaissance de motifs conservés chez les microbes, permet de discerner un large spectre de pathogènes et d'induire la réponse de défense. Finalement le second niveau de détection, basé sur la reconnaissance spécifique d'un pathogène grâce à certaines de ses protéines qui servent souvent de facteurs de virulence, permet de détecter des pathogènes propres à la plante hôte et d'initier des réponses de défense plus efficaces. Mais bien sûr, les pathogènes tentent d'esquiver la réponse de défense initiée par le système immunitaire afin de pouvoir coloniser leur plante hôte. Une course à l'armement constante décrit bien la relation qui existe entre une plante hôte et son pathogène (Fig i.2) (Jones and Dangl, 2006).

2.1 Les barrières physiques

Les végétaux sont séparés du monde extérieur par des barrières physiques comme la cuticule ou la paroi cellulaire. Elles constituent un mécanisme de défense passif, et représentent le premier obstacle à franchir pour les agents pathogènes. La cuticule est composée d'acides gras (cutine) et de cires. Elle protège la paroi cellulaire exposée vers l'extérieur, des parties aériennes des plantes, en évitant l'accrochage d'insectes ou de micro-organismes. La paroi cellulaire des plantes constituée de cellulose, de pectine ou encore de lignine confère la rigidité aux cellules (Chisholm et al., 2006). Pour empêcher la pénétration de l'agent pathogène, les plantes répondent par un renforcement local de la paroi cellulaire sous le site de tentative de pénétration par formation d'une papille. Ce procédé implique le dépôt d'une matrice de callose et l'accumulation de composants tels que l'eau oxygénée, des composés phénoliques et diverses protéines et glycoprotéines possédant des propriétés hydrolytiques et antifongiques (Jayamohan and Kumudini, 2011). Afin de s'affranchir de ces défenses, de nombreux agents pathogènes produisent des enzymes lytiques pour détruire ces barrières et pénétrer dans la cellule végétale. Par exemple, certains champignons sécrètent des cutinases qui dégradent la cutine de la feuille en monomères (Boller and Felix, 2009). D'autres micro-organismes vont infecter la plante par des ouvertures naturelles telles que les stomates, ou par une blessure qui constituent également une porte d'entrée facilitée (Nümberger et al., 2004).

2.2 La réponse immunitaire induite

Lorsque les pathogènes infectent les cellules végétales, ils sont détectés par des récepteurs immunitaires, présents sur la membrane mais aussi dans les cellules, où ils constituent le système de surveillance de la plante. De manière générale, reconnaître l'agent pathogène va permettre à la plante

de résister à l'invasion par l'induction d'une réponse de défense. Par contre, si elles ne possèdent pas de récepteur approprié, l'agent pathogène ne sera pas reconnu et l'infection aura lieu.

2.2.1 Réponse immunitaire à large spectre (PTI/MTI)

Passé les barrières physiques, la première ligne de défense est rencontrée par l'agent pathogène à la surface cellulaire. Ce premier niveau de défenses induites est nommé résistance basale. Peu spécifique, il est ainsi efficace contre l'ensemble des membres d'une espèce, d'un genre ou d'une famille d'agents pathogènes. Cette première réponse est activée lors de nombreuses interactions plantes - agents pathogènes et implique des protéines réceptrices PRR qui reconnaissent des signatures microbiennes moléculaires, PAMP / MAMP (Boller and Felix, 2009). Ces récepteurs de reconnaissance PRR, pour « Pattern Recognition Receptors » de ces motifs qui sont des récepteurs transmembranaires qui portent souvent un domaine de protéine kinase (RLK receptor like kinases ou RLP receptor like proteins) (Jones and Dangl, 2006). Les PAMP sont généralement des composants essentiels de classes entières d'agents pathogènes, tels que la flagelline bactérienne (protéine du filament flagellaire) ou la chitine (polysaccharide de N-acétylglucosamine) fongique. Les plantes répondent également à des molécules endogènes libérées par l'invasion de pathogènes, tels que des fragments de paroi cellulaire ou de cuticule, appelés DAMP pour « damages associated molecular pattern ». La reconnaissance de ces motifs par les PRR conduit à l'immunité PAMP/MAMP déclenchée, PTI / MTI « PAMP / MAMP Triggered Immunity » (Boller and Felix, 2009; Chisholm et al., 2006; Jones and Dangl, 2006; Zipfel, 2014).

Selon leur « mode de vie », les pathogènes des plantes sont classés comme biotrophe, nécrotrophe ou hemibiotrophe. Il semble que la PTI soit impliquée dans les réponses contre les biotrophes, alors que la DTI « DAMPs triggered immunity » est plutôt impliquée dans la résistance contre les nécrotrophes et les insectes. La PTI et la DTI sont impliquées, à la fois dans la réponse contre les hémibiotrophes. Les pathogènes biotrophes se nourrissent de cellules vivantes c'est pour cela qu'ils évitent tout dommage, excepté au niveau de la membrane de la cellule pendant la pénétration. Les pathogènes nécrotrophes au contraire vont tout d'abord tuer la cellule hôte pour pouvoir se nourrir des débris résiduels causant le maximum de dégâts. Les hemibiotrophes possèdent une étape initiale biotrophe suivie par une étape necrotrophe (Li et al., 2013).

Les récepteurs immunitaires constitués par les PRR sont situés à la surface de la cellule. Ce sont des protéines transmembranaires RLK « receptors like kinases » avec un domaine extracellulaire de liaison au ligand de type LysM (Lysine Motif) ou LRR (Leucine Rich Repeat), un domaine transmembranaire et un domaine kinase intracellulaire. Le domaine kinase peut être absent, ces PRR sont dénommées RLP « receptors like proteins », ils fonctionnent avec les RLK pour permettre la transduction du signal. (Zipfel, 2014). Le génome du riz contient plus de 1 131 gènes codant pour des RLK et 90 pour des RLP qui peuvent être impliqués dans la signalisation cellulaire (Liu et al., 2014a).

Ces PRR vont être activés par la perception de PAMP. La flagelline, la chitine ou le peptidoglycane (PGN) font partie des PAMP les plus étudiés. La flagelline est une protéine composant le flagelle bactérien, organelle requis pour le déplacement des bactéries. Le PRR détectant la flagelline (FLS2 chez *Arabidopsis*, OsFLS2 chez le riz) est un RLK avec un domaine LRR. FLS2 détecte l'invasion bactérienne par liaison directe de la flagelline via son épitope flg22. Cette reconnaissance induit la dimérisation de FLS2 ou son heterodimerisation avec le co-récepteur BAK1 (BRI1-associated kinase 1). BAK1 est un LRR-RLK requis pour la signalisation (Chinchilla et al., 2007). Les molécules composant les membranes des pathogènes comme la chitine des champignons, ou le peptidoglycane chez les bactéries sont également reconnues par les PRR. Chez le riz, CEBiP (Chitin oligosaccharide elicitor binding protein) est le RLP majeur impliqué dans la reconnaissance de la chitine ou de ses fragments. CEBiP possède un domaine transmembranaire et deux motifs LysM extracellulaires, cependant CEBiP ne possède pas de domaine intracellulaire de type kinase permettant la transduction du signal. Ainsi la protéine CEBiP coopère avec la protéine LysM RLK OsCERK1 (chitin elicitor receptor kinase1) lors de la reconnaissance de la chitine (Akamatsu et al., 2013). Cette hétérodimérisation permet la transduction du signal en activant la PTI. De la même manière, le peptidoglycane est reconnu par des PRRs de type LysM-RLK (Liu et al., 2014a).

La cascade de signalisation activée par les différents PRR semble similaire et converge vers une même réponse immunitaire (Boller and Felix, 2009). Cette activation de la PTI se traduit par une production rapide d'espèces oxydatives (ROS), l'alcalinisation de l'apoplasme, la fermeture des stomates et le renforcement de la paroi cellulaire par un dépôt de callose (Senthil-Kumar and Mysore, 2013; Trdàj et al., 2015). Une augmentation de la concentration en calcium (Ca^{2+}) cytoplasmique survient également rapidement. Le calcium agit comme un second messager activant des protéines kinases dépendant du calcium (CDPK) conduisant à l'expression de gènes de défense (Boudsocq et al., 2010). Des gènes de défense codant des enzymes hydrolytiques (protéases, chitinases et glucanases) sont également régulés par l'action de l'acide salicylique (SA), une phytohormone qui va s'accumuler au niveau de la zone d'infection et dans les tissus adjacents conduisant à une résistance systémique acquise (SAR) (Gao et al., 2015; Giraldo and Valent, 2013). La PTI/MTI active des cascades de signalisation via la voie des MAP (mitogen-activated protein) Kinases (Boudsocq et al., 2010). Cette signalisation conduit à une reprogrammation transcriptionnelle importante réalisée par l'action de facteurs de transcription dont les plus étudiés sont ceux de type WRKY. Chez le riz, plus de 100 facteurs de transcription WRKY ont été identifiés, dont *OsWRKY30* et *OsWRKY45* qui jouent un rôle important dans la résistance du riz face au pathogène *Magnaporthe oryzae* (Liu et al., 2014a).

L'ensemble des réponses de la plante via la PTI a pour but d'empêcher l'invasion de l'hôte par le pathogène.

2.2.2 Contournement de la réponse primaire PTI/MTI, par le pathogène

Les agents pathogènes ont évolué pour supprimer ou interférer avec les réponses MTI en sécrétant des composés, tels que des effecteurs, des protéases ou des toxines favorisant l'établissement de la colonisation. Dans le cas où le pathogène est capable de contourner tous les systèmes de reconnaissance de son hôte, il parvient à se développer et à accomplir son cycle en provoquant la maladie. Cette sensibilité de la plante due aux effecteurs du parasite est qualifiée d'ETS (Effector-Triggered Susceptibility) (Jones and Dangl, 2006). Cependant, la plante est capable de réagir car elle possède d'autres récepteurs immunitaires lui permettant de mettre en place une réponse immunitaire ciblée.

2.2.3 Réponse immunitaire spécifique déclenchée par les effecteurs ETI

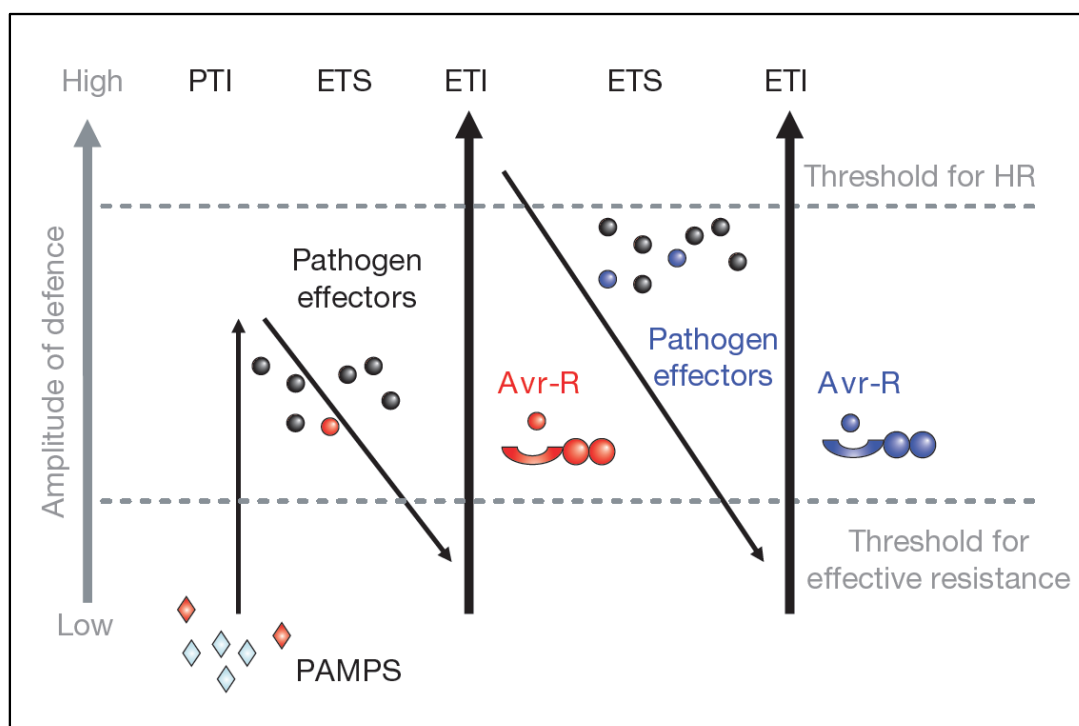


Fig i.2 Modèle en zig-zag : une réponse immunitaire induite à deux niveaux :

Dans un premier temps, les plantes reconnaissent les PAMPS (losanges rouges) via les récepteurs transmembranaires PRR (non représentés) déclenchant ainsi la PTI (PAMPS Triggered Immunity).

Certains agents pathogènes se sont adaptés et sont capables de franchir cette ligne de défense. La sécrétion d'effecteurs permet au parasite d'interférer avec la PTI et favorise sa nutrition et sa propagation. La plante devient « sensible », c'est l'ETS (Effector-Triggered Susceptibility).

Le 2^{ème} niveau de détection implique la reconnaissance d'un effecteur particulier, alors qualifié de protéine d'avirulence (Avr) (rond rouge), par une protéine de résistance R, ce qui déclenche une résistance qualifiée d'ETI (Effector-Triggered Immunity). Cette réponse conduit à de fortes réactions de défense caractérisées par la mort cellulaire ou réponse hypersensible (HR).

La pression de sélection et l'évolution poussent les agents pathogènes à perdre ou à muter l'effecteur reconnu, favorisant ainsi les souches qui portent de nouveaux effecteurs (bleus), et dont le résultat est le contournement de l'ETI. De la même manière, la sélection favorise les nouveaux allèles de protéine R de plantes capables de reconnaître ces nouveaux effecteurs, conduisant au déclenchement de l'ETI.

Adapté de (Jones and Dangl, 2006)

Dans une course aux armements incessante entre l'hôte et les attaques de micro-organismes, les plantes ont développé des récepteurs qui permettent la reconnaissance des effecteurs des agents pathogènes (Fig i.2) (Jones and Dangl, 2006). Ce sont des protéines intracellulaires multi-domaines, de type NLR (Nucleotide-Binding and Leucine-Rich Repeat domains). Elles peuvent détecter la présence ou l'activité d'effecteurs. Lorsque l'effecteur est reconnu, il est qualifié de facteur d'avirulence spécifique (Avr protéine). Cette reconnaissance permet d'initier une réponse de défense immunitaire dite « effecteur déclenchée » (ETI ou effector-triggered immunity) (Trdàj et al., 2015). Cette reconnaissance est considérée comme le deuxième niveau de détection dans le modèle en zig-zag (Fig i.2) décrit par Jones et Dangl, la première ligne de défense étant assurée par la PTI.

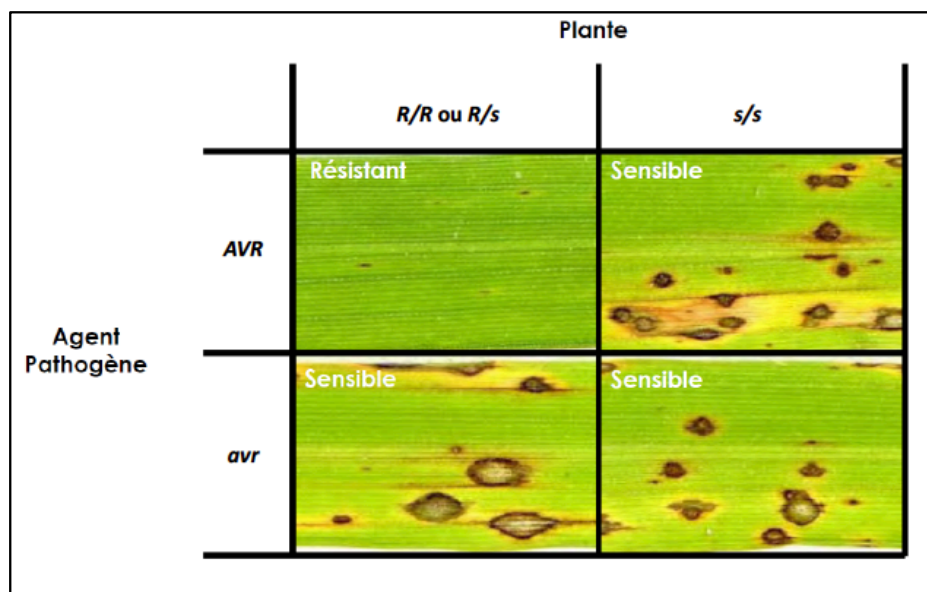


Fig i.3 Interaction compatible et incompatible, illustration de la relation gène pour gène :

L'expression concomitante des allèles dominants du gène de résistance (*R*) chez la plante et du gène la protéine effecteur d'avirulence (*AVR*) chez le pathogène entraîne une interaction incompatible et un phénotype de résistance de la plante, souvent accompagné d'une mort cellulaire programmée appelée réaction hypersensible (HR), correspondant à de petites zones de nécrose très localisées dans le panel en haut à gauche. En absence de l'allèle *R* et/ou de l'allèle *AVR* fonctionnel, la plante est sensible, la maladie se développe et l'interaction est dite compatible. (Cesari 2012)

L'ETI a pour effet une mort cellulaire programmée, qualifiée de réponse hyper-sensible (HR), qui apparaît localement au niveau des sites infectés, évitant ainsi la propagation du pathogène aux autres cellules de l'organisme hôte (Dodds and Rathjen, 2010). Cette réponse est souvent associée au système « gène pour gène ». Ce modèle a été formalisé par Flor en 1955 en se basant sur l'analyse de la virulence de différentes souches du champignon phytopathogène *Melampsora lini* face aux différentes variétés de lin (Flor, 1971). Selon ce modèle, la résistance est due à la reconnaissance des produits de deux gènes dominants entre le pathogène (gène d'avirulence *AVR*) et la plante (gène de résistance *R*). L'absence d'un des deux acteurs (gènes *AVR* ou gènes *R*) conduit à la sensibilité de la plante (Fig i.3). Cette résistance est dite « spécifique » ou « race-cultivar » puisqu'elle est efficace

uniquement chez les plantes porteuses de gènes *R* vis-à-vis de pathogènes porteurs de gènes d'avirulence. Lorsque les deux gènes sont présents et fonctionnels, l'agent pathogène ne peut pas se développer, l'interaction est dite incompatible.

Tout comme les PRR, les protéines NLR sont capables d'induire une réponse immunitaire face à différents pathogènes, comme les insectes, les champignons ou les bactéries (Rafiqi et al., 2009). Cependant, comparé à la PTI qui permet une détection à large spectre, l'ETI représente une reconnaissance spécifique entre la protéine d'avirulence du pathogène et la protéine NLR de plante. Les caractéristiques des protéines d'avirulence et des protéines de résistance seront abordées dans le paragraphe 2.3 de cette thèse. Ce type de reconnaissance engendre une coévolution fortement liée entre les deux partenaires.

2.2.4 Contournement de l'ETI par le pathogène

Pour infecter l'hôte, le pathogène doit pouvoir contourner l'ETI. Certains effecteurs sont capables de supprimer l'ETI ; par exemple, dans *Pseudomonas syringae*, l'effecteur AvrPphC supprime l'ETI déclenchée par l'effecteur AvrPphF dans certains cultivars de haricots, alors que comme son nom l'indique, AvrPphC peut causer l'avirulence sur des cultures différentes de haricots (Tsiamis et al., 2000). D'autres cas d'effecteurs bactériens, comme AvrPtoB, sont décrits comme capables d'inhiber l'ETI (Abramovitch et al., 2003). Par conséquent, il semble que certains effecteurs suppriment l'ETI déclenchée par d'autres effecteurs.

Une autre stratégie présente chez les agents pathogènes, consiste à se débarrasser des effecteurs Avr, reconnus par la plante. Ceci est seulement possible si la perte de l'effecteur Avr ne compromet pas la virulence de l'agent pathogène sur les hôtes sensibles. Pour cela, dans certains cas, la perte d'un effecteur Avr est compensée par la présence d'autres effecteurs du pathogène lui permettant de conserver sa virulence (Dong et al., 2015a). La perte d'effecteurs Avr est souvent facilitée par leur environnement génétique. En effet les gènes codants des effecteurs ne sont pas distribués de façon aléatoire à travers le génome mais semblent associés à des compartiments riches en séquences répétées et en éléments transposables (Dong et al., 2015a). Les gènes associés à ces régions ont tendance à évoluer plus rapidement que d'autres gènes et ces régions constituent donc un berceau de gènes, contribuant à l'évolution adaptative. Ce type de génome retrouvé chez les pathogènes filamenteux a été caractérisé par le modèle du « génome à deux vitesses » (Dong et al., 2015a).

Ces mécanismes ont pour conséquence que dans les populations d'un agent pathogène, la fréquence d'un gène codant un effecteur sera renforcée par la capacité de l'effecteur à promouvoir la virulence et sera réduite si l'effecteur est reconnu par le système immunitaire de la plante. Par sélection naturelle le pathogène peut donc contourner les défenses de la plante et continuer à provoquer la maladie.

2.3 Les principaux acteurs de l'ETI

2.3.1 Les protéines de résistances R

L'action d'un effecteur avirulent, si elle n'est pas contrecarrée, va conduire à l'affaiblissement ou la mort de la plante. Pour résister, les plantes ont évolué pour trouver une parade face aux pathogènes. Cette résistance n'est évidemment pas basée sur une seule stratégie, puisque les effecteurs du pathogène sont par essence diversifiés, mais elle pourra impliquer tout un panel de reconnaissance de ces effecteurs et de cascades de régulation. Cependant il existe un consensus dans la constitution de la première étape de cette reconnaissance, puisqu'elle est assurée par des récepteurs immunitaires similaires. En effet, les végétaux possèdent des gènes de résistances codant des protéines de résistances (R). Ces récepteurs immunitaires sont des protéines NLR qui assurent une surveillance et sont impliqués dans la reconnaissance d'effecteurs particuliers (Chisholm et al., 2006). Ainsi, grâce au séquençage de leurs génomes, environ 150 gènes de NLR ont été identifiés chez la plante modèle *A. thaliana* et plus de 600 gènes chez le riz (Goff et al., 2002).

2.3.1.1 Caractéristiques des NLR

Les protéines NLR de plante (Nucleotide Binding - Leucine Rich Repeat domains) ; aussi appelées NBS-LRR, NB-ARC-LRR, ou NB-LRR; comportent un domaine C-terminal riche en leucines (LRR), un domaine central de fixation d'un nucléotide (NBS ou NB-ARC) et un domaine N-terminal variable (Fig i.4).

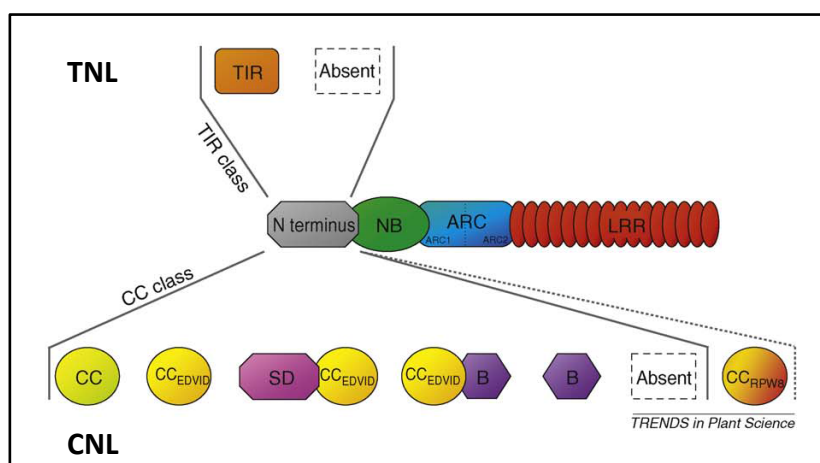


Fig i.4 Architecture des protéines NLR :

Sur ce schéma sont représentés les différents domaines constituant les protéines NLR. Deux grands groupes se distinguent au niveau du domaine N-terminal, les TNL et les CNL. Le domaine C-terminal est constitué d'une région riche en leucine (LRR). Le domaine central est un site de fixation au nucléotide (NB-ARC) (Adapté de Collier and Moffett, 2009).

Les protéines avec un domaine NB-ARC (Nucleotide Binding Adaptor Shared by Apaf-1, Resistance protein and CD-4) (Van Der Biezen and Jones, 1998) appartiennent à la super famille des protéines STAND (Signal Transduction ATPases with Numerous Domains) qui regroupe des NTPases, et sont impliquées dans des fonctions de régulation moléculaire comme l'immunité ou l'apoptose (Qi and Innes, 2013).

Les protéines NLR de plante peuvent être divisées en deux classes majeures, en fonction de leur domaine N-terminal, les CNL (CC-NB-LRR) avec un domaine coiled-coil (CC) en N-terminal et les TNL (TIR-NB-LRR) possédant un domaine N-terminal TIR (Toll/interleukin-1 receptor domain) (Meyers et al., 2003; Rafiqi et al., 2009).

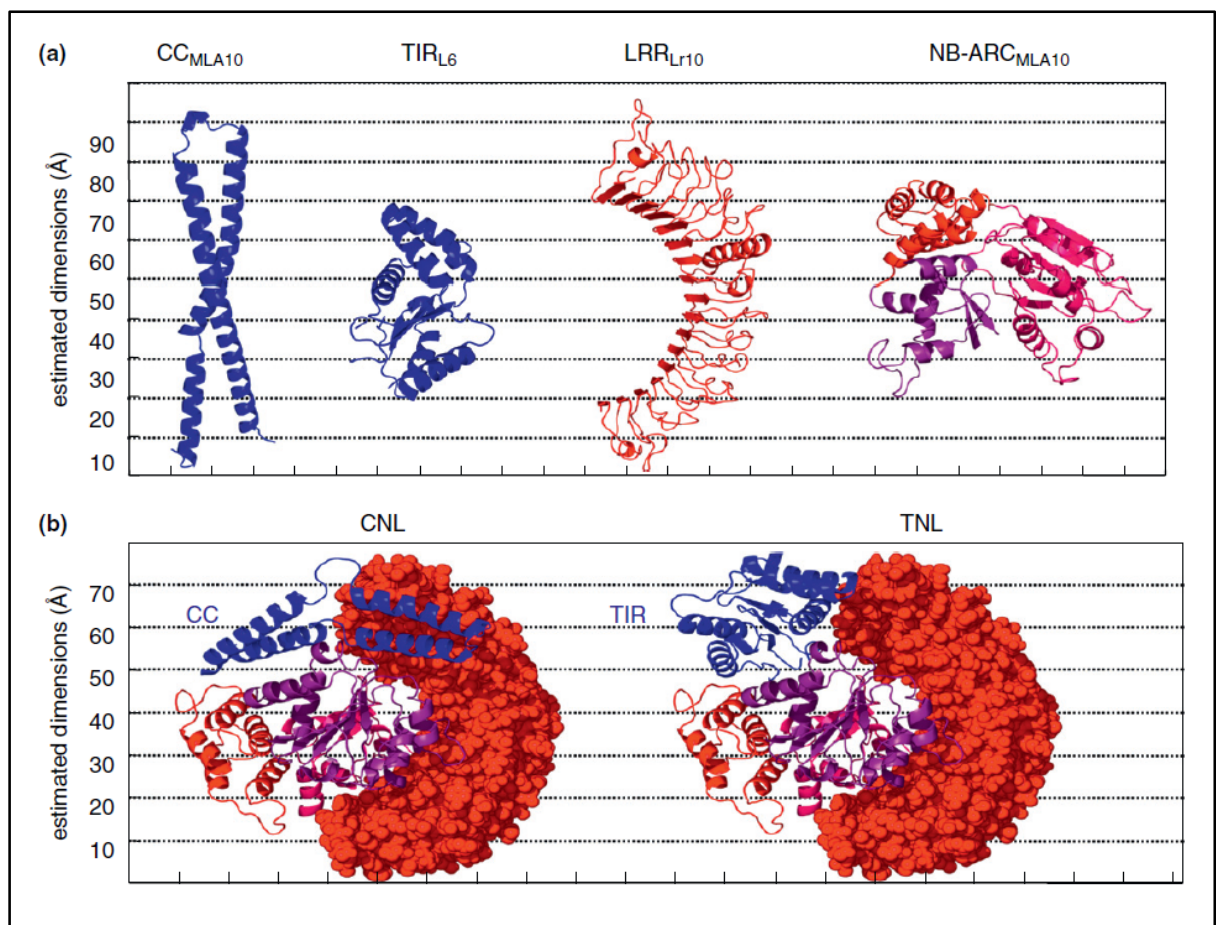


Fig 1.5 Modèles de structures des différents domaines trouvés chez les protéines NB-LRR :

a) Structure des domaines TNL et CNL avec l'estimation de leur dimension en Å.

Les modèles 3D du CC de Mla 10 et TIR de L6 sont basés sur les structures cristallines alors que les structures 3D du NB-ARC de Mla10 et du LRR de Lr10 viennent de modèles prédits.

b) Modèles hypothétiques du repliement des CNL et TNL permettant d'illustrer les distances relatives entre les domaines et leur possible orientation en conformation. Le domaine ARC2 est en violet.

Adapté de (Takken and Goverse, 2012)

Les connaissances structurales des TNL et CNL sont pour l'instant parcellaires, seules les structures de sous-domaines ont été déterminées par cristallographie ou modélisées à partir de structures homologues. L'agencement global de ces sous-domaines entre eux dans un CNL ou un TNL reste hypothétique, et est basé sur une compilation d'informations biochimiques, fonctionnelles et structurales (Fig i.5)

L'hypothèse centrale de ce modèle est que le domaine NB-ARC interagit avec la partie N-terminale du LRR, donnant à l'ensemble une conformation fermée, et ce en absence de l'effecteur. L'interface est constituée par un ensemble de résidus positivement chargés sur la surface du LRR qui interagissent avec une face du domaine NB-ARC. D'après le modèle prédit du LRR, l'extrémité C-terminale est particulièrement enrichie en résidus aromatiques et elle pourrait être impliquée dans des interactions hydrophobes avec le domaine N-terminal. La position des domaines CC et TIR, dans les CNL et les TIR, respectivement, reste fortement hypothétique, et va être notamment dépendante de l'état d'oligomérisation du récepteur (Takken and Tameling, 2009).

Outre les NLR avec une structure conventionnelle comme les CNL ou les TNL, les génomes de plantes codent pour un nombre significatif de NLR ou « NLR like protein », constituées de domaines non conventionnels ou de domaines avec un réarrangement atypique (Jacob et al., 2013). Par exemple, le NLR RRS1 chez *A. thaliana* possède en C-terminal un domaine additionnel de type WRKY (Meyers et al., 2003). Chez le riz, les NLR RGA5 et Pik-1 possèdent un domaine RATX1 (relié à la protéine de levure liant le cuivre ATX1) similaire (Cesari et al., 2013). Il est plausible qu'un événement de fusion soit intervenu entre un NB-LRR 'canonique' et un domaine additionnel, WRKY ou RATX1 (Jacob et al., 2013).

Selon l'hypothèse « Rosetta stone » de David Eisenberg (Marcotte et al., 1999), lorsque deux protéines qui sont séparées dans une espèce se retrouvent fusionnées dans une autre espèce, cette fusion serait le reflet de leur interaction, possiblement non connue, dans la première espèce. Ainsi, des domaines additionnels fusionnés au cœur du NLR pourraient contribuer à différentes fonctionnalités, comme la reconnaissance de l'effecteur, la régulation du NLR et le mécanisme de signalisation, quelle que soit leur position dans le squelette du NLR.

Les sous-domaines canoniques des NB-LRR sont décrits dans les paragraphes suivants :

- **Domaines N-terminaux TIR et CC**

Le domaine TIR (Toll/interleukin-1 receptor) est retrouvé au niveau des récepteurs immunitaires à la fois chez les animaux et chez les plantes. Chez les animaux, les TLR (Toll like receptors) correspondent à un domaine de signalisation intracellulaire qui déclenche la réponse immunitaire lors de la perception extracellulaire de PAMP (Tapping, 2009). Chez les plantes, ce domaine est retrouvé

en N-terminal des protéines NB-LRR, répertoriées comme TNL. L'autre classe, constituée de motifs « coiled coil » (CC), forme la famille des CNL. Ce type de récepteur est retrouvé notamment chez le riz où plus de 250 protéines de type CNL ont été identifiées dans son génome. De manière intéressante, le domaine TIR est généralement absent du génome de graminées (Meyers et al., 2003). Il existe également une autre classe appelée XNL, correspondant à des protéines NB-LRR possédant un domaine N-terminal n'étant ni TIR ni CC (Jacob et al., 2013).

Ces domaines N-terminaux sont généralement impliqués dans la signalisation provoquant la mort cellulaire comme cela a été démontré pour la protéine TNL L6 du lin (Bernoux et al., 2011) et la protéine CNL MLA10 d'orge (Maekawa et al., 2011b). Cependant il existe d'autres exemples, comme celui de la protéine CNL Rx de pomme de terre, pour laquelle le domaine CC n'est pas requis dans l'activation de la mort cellulaire (Rairdan et al., 2008). La fonction des domaines TIR et CC dans la perception du pathogène reste mal comprise.

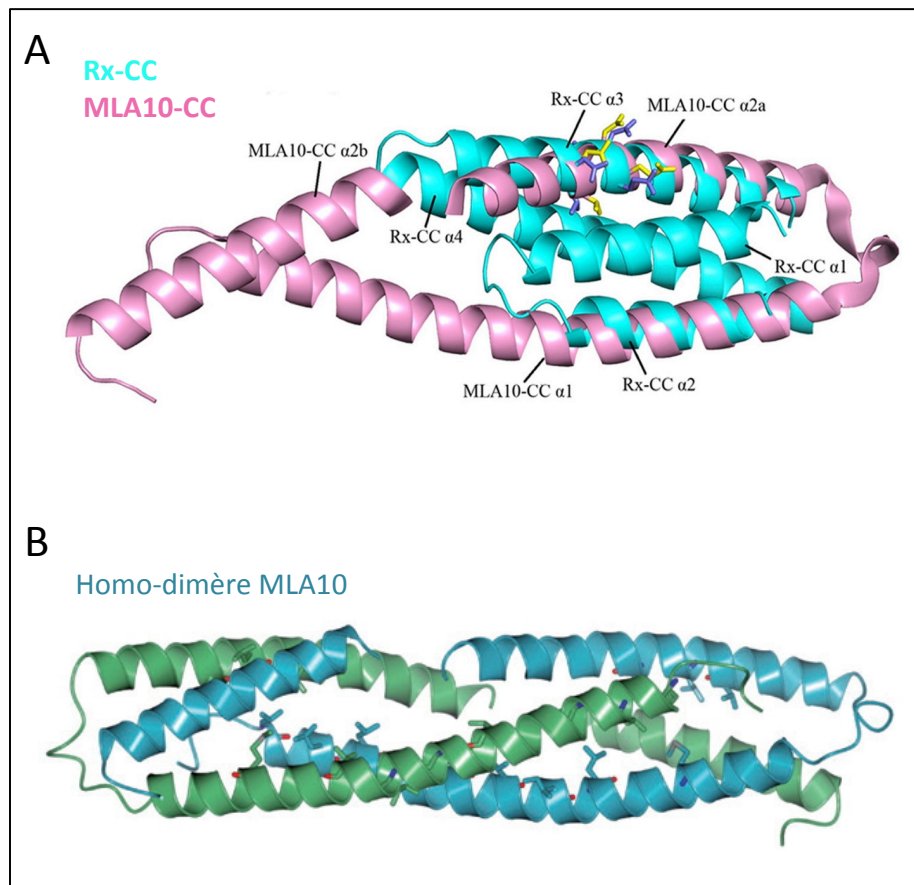


Fig i.6 Structures des domaines CC :

A. Domaines coiled-coil des protéines CNL, MLA10 et Rx:

Superposition des régions CC de Rx (cyan) et MLA10 (rose). Les chaînes latérales du motif EDVID de Rx CC et de MLA10 CC sont en gris-bleu et en jaune, respectivement. (Adapté de Hao et al., 2013)

B . Dimère de MLA10:

Les deux chaînes correspondant aux homo-dimères sont colorées en vert et bleu. Les résidus importants pour la stabilité ou la formation du dimère sont représentés en stick (Adapté de Wirthmueller et al., 2013)

Les structures des domaines Coiled-coil de Rx (Hao et al., 2013) et de MLA10 (Maekawa et al., 2011b) ont été déterminées par cristallographie. La superposition des structures Rx CC et MLA10 CC (Fig i.6) révèle l'équivalence structurale entre les hélices $\alpha 1$ et $\alpha 2$ de Rx CC et la longue hélice $\alpha 1$ de MLA10 CC, ainsi qu'entre les hélices $\alpha 3$ et $\alpha 4$ de Rx CC et les hélices $\alpha 2a$ et $\alpha 2b$ de MLA10. Les boucles connectant les hélices $\alpha 1$ - $\alpha 2$ et $\alpha 3$ - $\alpha 4$ dans Rx CC, permettent aux hélices $\alpha 1$ et $\alpha 4$ de venir en contact avec les hélices $\alpha 2$ et $\alpha 3$, par contraste avec la structure étendue de MLA10 CC. Ces structures démontrent donc une divergence certaine. Malgré l'existence des CC, repérés par leur signature caractéristique de séquence primaire, la traduction en termes structuraux apparaît pour l'instant peu claire, même si les motifs CC sont impliqués dans la formation de dimères (Fig i.6). Notamment, ces régions CC ne seraient pas obligatoirement impliquées, ou du moins pas de la même manière (ou pas du tout selon les cas), dans la reconnaissance de l'effecteur, dans l'interaction avec des co-facteurs ou encore la formation d'homo et d'hétéro-dimères de CNL. Alors que le MLA CC est un dimère constitutif, le domaine TIR de L6 dimérise lors de la perception du pathogène. La résolution des structures (Fig i.5) de domaines TIR de L6 (Bernoux et al., 2011) et CC de MLA10 (Maekawa et al., 2011b) montre que ces domaines peuvent former des homo-dimères. Des expériences de type double hybride ou de co-immunoprécipitation confirment la formation de dimères. D'autres types de NB-LRR sont caractérisés comme homo-dimères via leur domaines N-terminaux comme RPS5 chez *Arabidopsis* ou Prf chez la tomate (Takken and Goverse, 2012). Cette dimérisation est un niveau de régulation supplémentaire dans l'activation de la résistance conduisant à la signalisation de l'apoptose. Cette hypothèse est renforcée par le fait que la mutagenèse dirigée contre les acides aminés permettant la formation du dimère produit des monomères qui ne sont pas capables d'activer la signalisation conduisant à la mort cellulaire (Bernoux et al., 2011; Maekawa et al., 2011b).

• **Domaine NB-ARC**

Le domaine NB est retrouvé chez les NLR de mammifère sous le nom de NOD (nucleotide-binding oligomerization domain). Ce domaine est impliqué dans la régulation des réponses immunitaires innées et dans l'apoptose (Inohara et al., 2005). Le domaine NB-ARC joue un rôle clé dans l'activation des protéines NB-LRR, car c'est par l'activation de ce domaine qu'est transmise la détection de l'effecteur. Il est constitué de 3 sous-domaines, NB, ARC1 et ARC2 formant une poche de liaison au nucléotide qui adopte différentes conformations en fonction du nucléotide lié.

Pour l'instant, seuls des modèles, construits par homologie, du domaine NB-ARC de NB-LRR sont disponibles (Fig i.5). Ces modèles sont basés sur les structures cristallines des protéines Apaf1, CED-4 et d'autres protéines de la famille des STAND (Lukasik and Takken, 2009; Riedl et al., 2005; Takken and Goverse, 2012; Yan et al., 2005). Le sous-domaine NB est un repliement classique NTPase, ce qui place ces protéines NB-ARC dans le grand groupe des NTPases à P-loop.

Dans ce repliement, un feuillet β de 5 brins parallèles est flanqué d'hélices α . Les brins $\beta 1$ et $\beta 3$ sont associés aux motifs P-loop et Walker B, respectivement (Fig i.7). Le sous-domaine ARC1 est un

« four-helix bundle » et ARC2 adopte un repliement en « winged-helix ». La structure du NB-ARC est très compacte et à l'interface de ces trois sous-domaines se trouve le site de liaison au nucléotide.

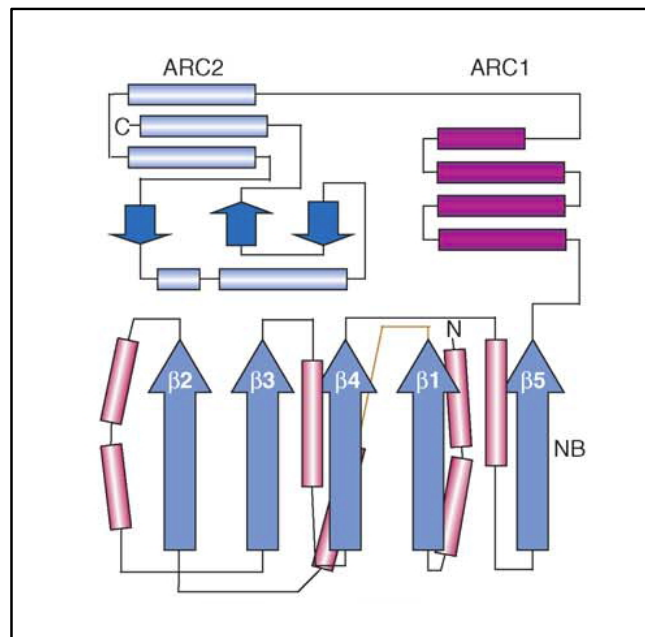


Fig i.7 Architecture du domaine NB-ARC :

Représentation du domaine NB-ARC1-ARC2 dérivé de la structure cristallographique de APAF-1. Les cylindres et les flèches correspondent respectivement aux hélices α et aux brins β (Adapté de Takken et al., 2006)

Le changement conformationnel coïncide avec l'échange de l'ADP avec de l'ATP qui stabiliserait une conformation active (Ade et al., 2007; Tameling, 2006). L'hydrolyse de l'ATP permettrait le retour à la forme inactive. Ce mécanisme semble conservé entre les différents organismes comme les bactéries ou les plantes, chez qui l'on retrouve ce type de domaine appartenant à la super-famille des STAND (Danot et al., 2009; Takken and Tameling, 2009). A cause de sa capacité à osciller entre la forme active, liée à l'ATP, et inactive liée à l'ADP, ce domaine est considéré comme un interrupteur moléculaire (Fig i.8). Des motifs structuraux conservés ont été identifiés dans ces domaine NB-ARC comme le motif P-loop (ou Walker A), GLPL, RNBL-B, ou encore MHDV qui sont essentiels à l'activation et à la régulation des NB-LRR (Meyers et al., 2003; Tameling, 2006). Des mutations dans ces motifs induisent la dérégulation pouvant conduire à l'auto-activation des NB-LRR provoquant une réponse hypersensible en l'absence d'effecteurs de pathogènes (Ade et al., 2007; Maekawa et al., 2011b; Tameling, 2006).

• **Domaine LRR**

Dans les NLR, le domaine LRR est considéré comme un domaine senseur. Il est impliqué dans les interactions protéines-protéines et c'est par ce domaine qu'est acquise la spécificité de

reconnaissance de nombreuses protéines NLR (Dodds et al., 2001; Ellis et al., 1999; Padmanabhan et al., 2009; Rafiqi et al., 2009). Situé en C-terminal des protéines NLR, ce domaine est très polymorphe. La plupart des variabilités d'acides aminés se produisant dans ce domaine sont dues à une sélection positive, caractérisée dans la plupart des gènes *R* par des rapports Ka/Ks (Ka: substitution non-synonyme divisée par Ks: substitution synonyme) élevés, de fréquentes recombinaisons, une variation du nombre de copies et une divergence de séquence élevée entre les espèces (Chen et al., 2010).

L'une des fonctions du domaine LRR est de reconnaître la présence de l'effecteur de pathogène par liaison directe ou indirecte (Caplan et al., 2008). Le premier cas de reconnaissance directe a été décrit entre la protéine NLR Pi-ta du riz et l'effecteur de *Magnaporthe oryzae* AvrPi-ta (Jia et al., 2000). L'étude de variants de Pi-ta montre que la spécificité implique des résidus prédits exposés à la surface du domaine LRR, et que des mutations dans cette région provoquent la perte de l'interaction. Le nombre inhabituellement faible de polymorphismes dans le domaine LRR des allèles résistants, suggère une sélection favorisant la mutation avantageuse (Huang et al., 2014). Dans le pathosystème lin - rouille du lin, les domaines LRR de protéines NLR subissent également une forte pression de sélection au niveau des régions impliquées dans la spécificité de reconnaissance (Ellis et al., 1999). Des expériences d'échange de domaines montrent que le changement de seulement six acides aminés d'un domaine LRR peut lui conférer une nouvelle spécificité (Dodds et al., 2001) ce qui indique que le polymorphisme du LRR, à lui seul, peut conduire à un changement de spécificité (Padmanabhan et al., 2009). Les effecteurs peuvent également être reconnus de façon indirecte en association avec une protéine additionnelle (« host factor ») qui a été modifiée par l'effecteur (Caplan et al., 2008). Dans la plupart des cas, cette reconnaissance indirecte fait intervenir en plus du LRR le domaine N terminal TIR ou CC.

En plus de son rôle de senseur, le domaine LRR régule l'activité du récepteur immunitaire par des interactions intramoléculaires avec les autres domaines. L'interaction entre le LRR et le sous-domaine ARC2 régule la transition entre l'état actif et inactif (Fig i.8) (Takken and Tameling, 2009). La première fonction du domaine LRR consiste donc à maintenir la protéine NLR dans une conformation inactive en l'absence d'effecteur (Rairdan and Moffett, 2006). Il a été mis en évidence que des mutations dans la partie C-terminale du domaine LRR de la protéine NLR Rx de la pomme de terre provoque une auto-activation en l'absence de la protéine virale de capsid (CP) du virus de la pomme de terre (PVX) (Lukasik and Takken, 2009). Le mécanisme par lequel le domaine LRR traduit la perception du pathogène en activation semble être plus complexe qu'une simple levée de répression qui pourrait découler des mutations introduites dans le LRR. En effet, en absence de ces domaines LRR, les études n'ont pas démontré une activation constitutive des NLR (Rairdan and Moffett, 2006). La reconnaissance d'un effecteur via le LRR entraîne un changement d'interface entre le LRR et ARC libérant ainsi le domaine NBS pour l'échange de nucléotides (Fig i.8). Ce réarrangement entrainerait également la libération du domaine TIR /CC impliqué dans la signalisation (Takken and Tameling, 2009). La dualité fonctionnelle du domaine LRR passe donc par sa capacité à reconnaître

spécifiquement l'effecteur dans un environnement où le LRR est impliqué également dans des interactions intramoléculaires avec les autres domaines au sein du récepteur immunitaire.

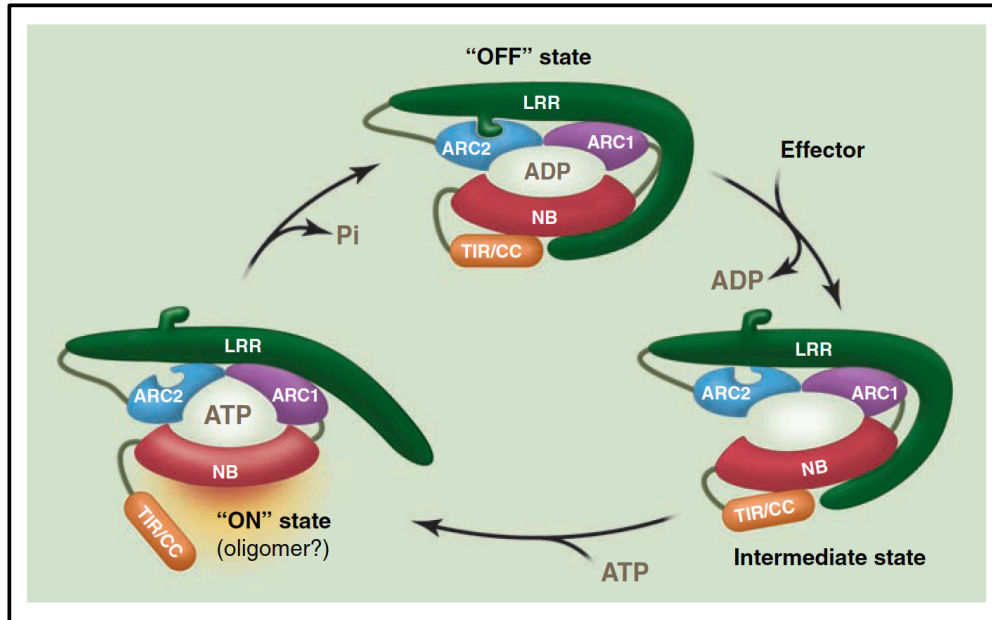


Fig. i.8 Modèle d'activation des protéines R :

En l'absence de pathogène, la protéine NB-LRR est auto-inhibée, la liaison à l'ADP stabilisée par le domaine LRR maintient cet état inactif « OFF ». La perception de l'effecteur par la partie C-terminale du domaine LRR change l'interface entre la partie N-terminale et le sous-domaine ARC2, générant une conformation plus ouverte de la protéine R permettant l'échange du nucléotide. L'échange ADP/ATP induit un second changement conformationnel, altérant les interactions entre les domaines NB-ARC central, le N-terminal du TIR/CC et le C-terminal du LRR résultant à l'état actif « ON ». L'hydrolyse de l'ATP initie le retour à l'état inactif.

Adapté de (Takken and Tameling, 2009)

Les domaines LRR sont présents dans de nombreuses protéines eucaryotes, procaryotes et virales dont les structures ont été résolues par cristallographie (Fig i.9) (Di Matteo et al., 2003; Kobe and Deisenhofer, 1993; Liu et al., 2013). Même si des différences existent entre ces structures de LRR, elles partagent un motif structural caractérisé par une région conservée en résidus leucines hydrophobes. La description du domaine est basée sur la structure cristalline de l'inhibiteur de ribonucléase RI (Kobe and Deisenhofer, 1993). Le motif riche en leucine (LRR) est généralement constitué de 20 à 29 résidus avec un segment de 11 résidus conservés formant une séquence consensus $LxxLxLxx^N/C xL$; x pouvant être n'importe quel acide aminé, les positions L (leucine) pouvant également être des valines, isoleucines ou phénylalanines (Kobe and Kajava, 2001). Ce motif correspond à une unité structurale comprenant un brin beta et une hélice alpha connectés par une boucle (Fig i.9). Les unités structurales sont disposées de façon à ce que les brins et les hélices soient parallèles à un axe commun, donnant une molécule en forme de fer à cheval où les brins beta forment un feuillet parallèle à l'intérieur et les hélices alpha sont à la circonférence extérieure du fer à cheval. La séquence conservée $LxxLxLxx^N/C xL$ correspond au segment entourant le feuillet beta. Il

existe une variabilité structurale au niveau des régions inter-brin, et les zones d'hélice peuvent également être plus courtes (Kobe and Kajava, 2001).

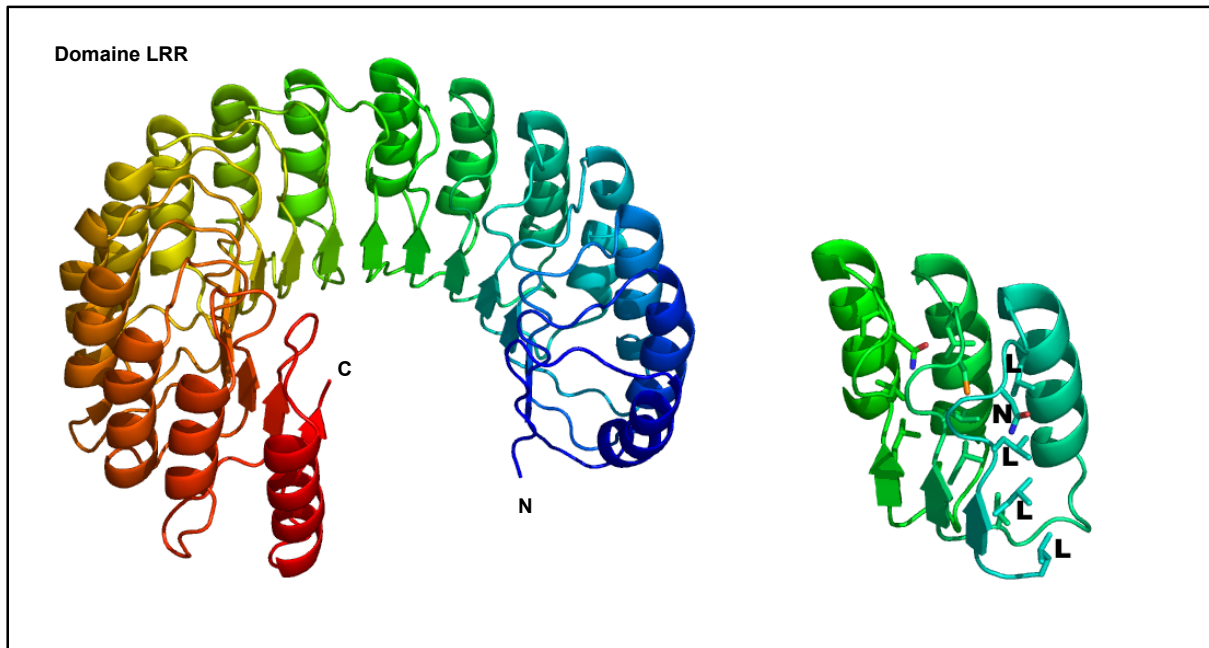


Fig.i.9 Structure du LRR de RI (pdb 2BNH) :

Les extrémités N et C terminales sont respectivement colorées en bleu et rouge. A droite trois répétitions ont été extraites. Les chaînes latérales et les positions des résidus leucine et asparagine sont indiquées sur le premier motif.

Les analyses du génome d'*Arabidopsis* ont révélé que les protéines NLR possèdent au moins 14 motifs LRR avec une région répétée d'environ 24 résidus (Meyers et al., 2003). La variabilité des domaines LRR, mise en évidence entre autres chez *Arabidopsis*, montre que la conservation de certains motifs n'est pas absolue. Les domaines LRR sont capables de tolérer des duplications ou des délétions entières de motifs répétés, ce qui leur permet une forte diversification et d'évoluer vers de nouvelles spécificités d'interaction (Padmanabhan et al., 2009). Cette sélection positive est observée sur les résidus exposés au solvant présents sur la face concave formée par les brins beta et les boucles adjacentes représentant les surfaces d'interaction les plus communes des protéines LRR (Kobe and Kajava, 2001; Padmanabhan et al., 2009).

2.3.2 Les effecteurs

Les pathogènes réussissent à perturber les défenses des plantes et les processus cellulaires via le déploiement d'une gamme remarquablement diversifiée de protéines effecteurs. Ces protéines sont les produits des gènes d'avirulence (*Avr*). Certains effecteurs peuvent être sécrétés avant même la

pénétration de l'hôte, beaucoup entrent dans les cellules végétales où ils ciblent divers compartiments cellulaires et d'autres se déplacent dans des cellules voisines afin de les préparer avant l'invasion.

2.3.2.1 Caractéristiques des effecteurs

Les effecteurs sont caractéristiques d'un type de pathogène et sont essentiels à l'infection.

Certaines bactéries pathogènes comme *Pseudomonas syringae*, *Ralstonia solanacearum* ou différentes espèces du genre *Xanthomonas* ont des systèmes de sécrétion de type III (TTSS: type three secretion system) utilisés pour injecter directement des effecteurs dans le cytoplasme des cellules hôtes. Certains de ces effecteurs sont reconnus chez la plante et peuvent donc déclencher des résistances, tels les effecteurs de *P. syringae* AvrRpt2, AvrRpm1 ou encore AvrPto qui induisent respectivement la résistance chez les plantes *Arabidopsis thaliana* et la tomate (Belkhadir et al., 2004; Xing et al., 2007). La sécrétion d'AvrPto dans le cytoplasme favorise la virulence chez les plantes sensibles mais déclenche une réaction de résistance chez les plants de tomates portant Pto, une serine/thréonine protéine kinase et Prf, une protéine de type NB-LRR interagissant avec Pto. L'importance de ces effecteurs est illustrée par l'observation de souches mutantes du TTSS qui se retrouvent incapables de surmonter la PTI et d'infecter les plantes (Hauck et al., 2003). La structure cristalline d'AvrPto (ensemble de 4 hélices α) en association avec Pto a été résolue par Xing et al., 2007. La protéine Pto possède un domaine kinase situé dans la boucle P+1. L'interaction entre AvrPto et Pto est arbitrée par une phosphorylation stabilisant la boucle P+1 et une seconde boucle dans Pto, ce qui régule négativement la réponse de défense induite par Prf en absence d'AvrPto. AvrPto agit donc comme un inhibiteur de la kinase Pto en interagissant avec les boucles de Pto (Xing et al., 2007). D'autres pathogènes comme les champignons ou les oomycètes ne possèdent pas de TTSS permettant de délivrer directement leurs effecteurs dans le cytoplasme. Pendant l'infection, ces parasites s'engagent dans des interactions biotrophes complètes ou partielles qui les amènent à développer des relations alimentaires avec les cellules vivantes de leur hôte par des associations étroites avec les tissus végétaux (Kamoun, 2007). Ceci est possible grâce à des structures spécialisées, tels que les hyphes, qui explorent l'espace extracellulaire (apoplaste), ou l'haustoria (notamment chez les oomycètes) (Fig i.10) et les hyphes invasives (Fig i.17) qui pénètrent dans les cavités de la cellule hôte et s'invaginent dans la membrane plasmique de l'hôte (Petre and Kamoun, 2014). Pour promouvoir l'infection, ces pathogènes secrètent tout un arsenal de protéines effecteurs par le biais de ces structures spécialisées. Chez les hemibiotrophes, comme les espèces du genre *Phytophthora* ou *Magnapotha oryzae*, différents effecteurs peuvent avoir des effets opposés. Certains, exprimés pendant la phase biotrophe précoce, peuvent inhiber la mort de la cellule hôte tandis que d'autres, exprimés pendant le stade final nécrotrophe de la maladie, agissent comme une toxine et tuent les cellules végétales (Giraldo and Valent, 2013; Koeck et al., 2011).

Souvent des effecteurs de biotrophie agissent au niveau du cytoplasme de la cellule hôte. Si leur mode de sécrétion impliquant l'hyphe invasive (Giraldo et al., 2013) ou l'haustoria (Petre and Kamoun, 2014) semble bien compris, la question de leur translocation dans le cytoplasme de l'hôte reste encore une grande controverse (Ve et al., 2013).

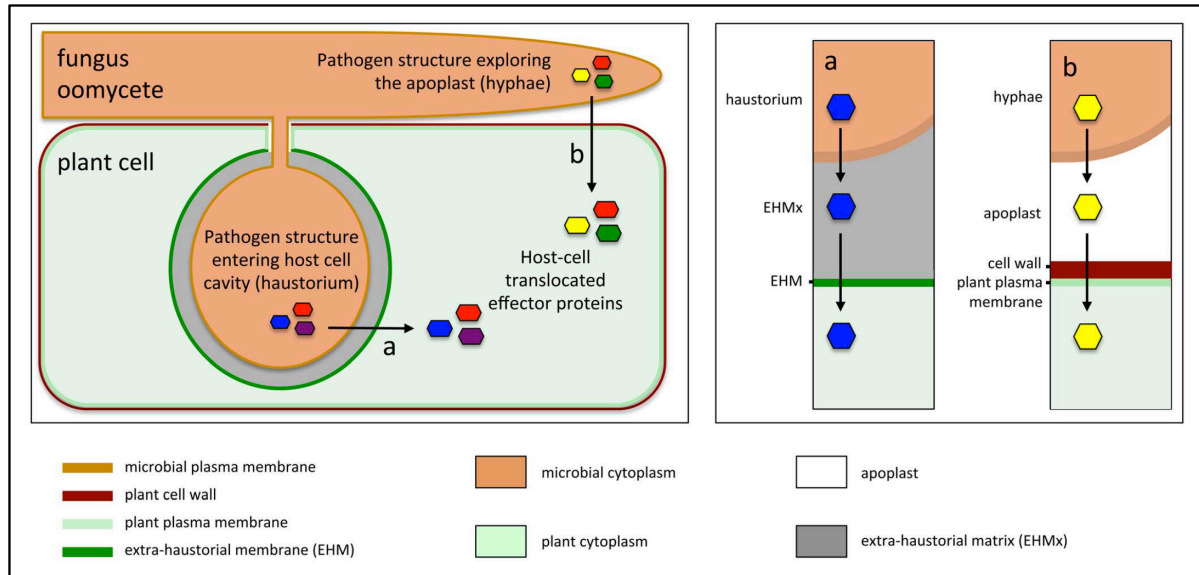


Fig i.10 Structures permettant la sécrétion des effecteurs chez les Champignons et les Oomycètes : Panneau de gauche : Les oomycètes et les champignons parasitent les plantes par l'intermédiaire de structures particulières à l'infection comme l'hyphe extracellulaire ou l'hyphe invasive et l'haustoria qui pénètrent la cellule hôte et s'invaginent dans la membrane plasmique. L'haustoria (a) et l'hyphe (b) secrètent des effecteurs qui sont transloqués dans le cytoplasme par un mécanisme inconnu. Panneau de droite : Les effecteurs sécrétés par l'haustoria (a) et l'hyphe (b) traversent différents interfaces biologiques (matrice extra-haustoriale EHMx/ membrane extra-haustoriale EHM) pour les effecteurs sécrétés par l'haustoria, et l'apoplaste, la membrane de la cellule de plante, la membrane plasmique pour les effecteurs sécrétés par l'hyphe. Adapté de (Petre et Kamoun, 2014)

Chez les oomycètes, une classe d'effecteurs correspondant à la famille RxLR possède un signal de sécrétion suivi d'un domaine N-terminal conservé défini par la séquence consensus RxLR (Arg-x-Leu-Arg, ou x correspond à n'importe quel acide aminé) accompagné dans certain cas par un motif dEER (souvent Asp, Glu-Glu-Arg) (Win et al., 2012). Ce domaine RxLR est requis pour cibler l'effecteur vers l'haustoria et pour la translocation dans les cellules de plantes. La région C-terminale des effecteurs RxLR est celle où l'on retrouve le plus de polymorphismes, ce qui est cohérent avec la vision que l'activité de l'effecteur réside dans cette région (Boutemy et al., 2011; Schornack et al., 2009). Ces effecteurs RxLR possèdent également, au niveau du domaine C-terminal, des séquences conservées (Trp, Tyr et Leu) répétées en tandems (Win et al., 2012).

L'analyse structurale des effecteurs de type RxLR est en plein essor. Ainsi l'élucidation des structures des effecteurs ATR1 (de *H. Arabidopsisidis*) (Chou et al., 2011), PexRD2 (de *P. infestans*), d'Avr3a11 et d'Avr3a4 (homologues de *Phytophthora capsici*) (Boutemy et al., 2011; Yaeno et al., 2011) montre que malgré une identité de séquence inférieure à 20% ces structures possèdent un repliement compact conservé formé de quatre hélices (Fig i.11).

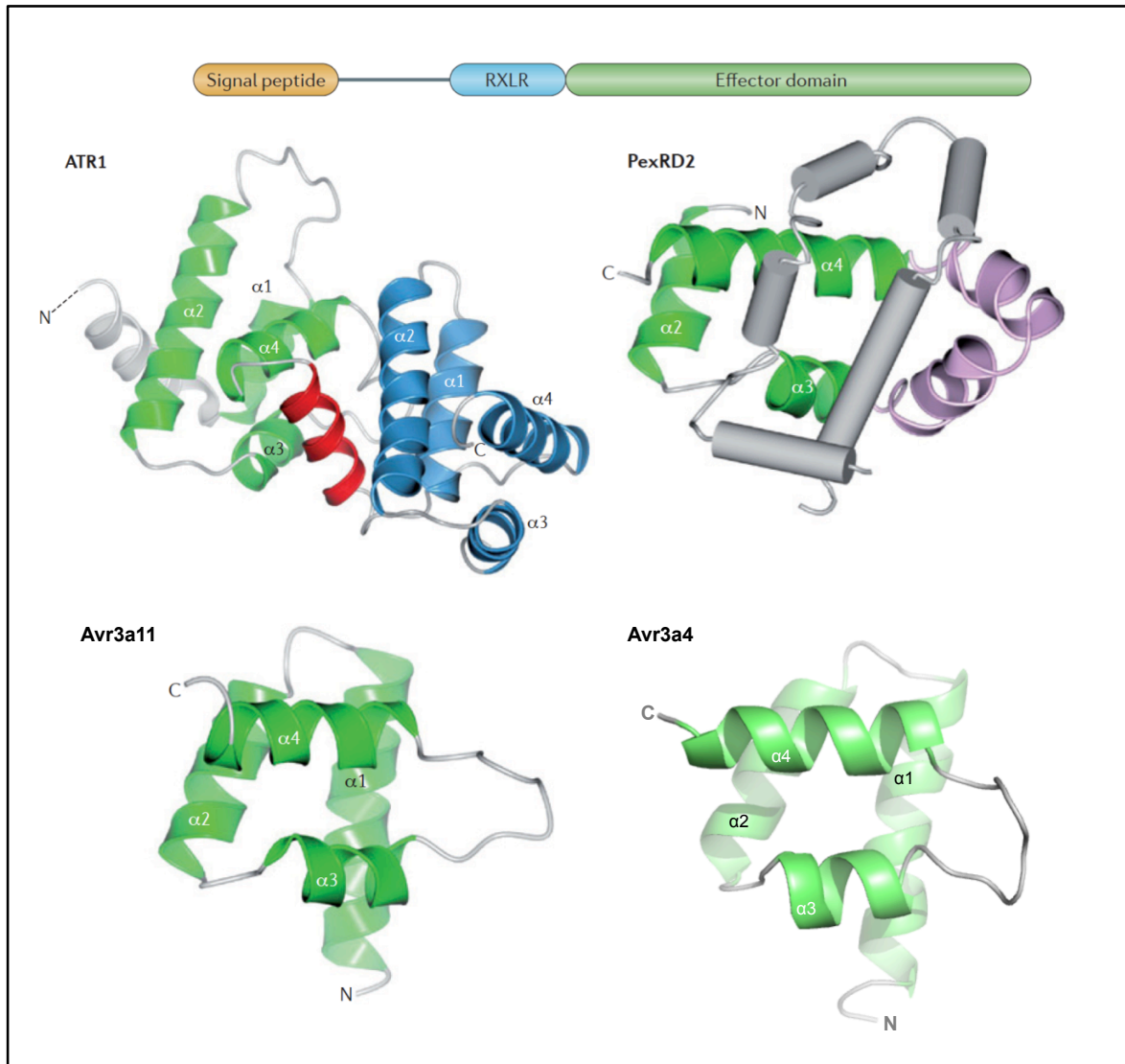


Fig i.11 Structures d'effecteurs de type RxLR :

La partie haute de la figure montre une représentation schématique d'un effecteur RxLR, avec le peptide signal de sécrétion, le domaine RxLR suivi du domaine effecteur.

Les hélices α contenant les domaines conservés WY sont colorées en vert ($\alpha 2$, $\alpha 3$ et $\alpha 4$). ATR1, Avr3a11 et Avr3a4 possèdent une hélice α supplémentaire ($\alpha 1$) en N-terminal. ATR1 possède deux domaines WY en tandem (hélices vertes et bleues) séparés par un lien coloré en rouge. PexRD2 est un homo-dimère, l'un des monomères étant représenté par les cylindres gris ; les hélices colorées en violet n'appartiennent pas au domaine WY.

(Adapté de Wirthmueller et al., 2013)

De façon intéressante, les motifs en tandem appelés motifs W et Y, se retrouvent conservés au niveau des trois hélices $\alpha 2$, $\alpha 3$ et $\alpha 4$. Une des hypothèses envisagées serait que la stabilité

moléculaire de ce repliement, combinée à une forte variation de séquence au niveau des résidus de surface, permettrait une évolution des effecteurs dans le maintien de leur activité de virulence tout en échappant à la reconnaissance par le système immunitaire de la plante (Boutemy et al., 2011; Win et al., 2012).

2.3.3 Les modèles de reconnaissance entre effecteur et protéine R

Les mécanismes de reconnaissance des effecteurs par les NLR ont été largement étudiés aux cours des dernières années avec pour trame la relation gène-pour-gène proposée par Flor en 1971 (Flor, 1971) pour expliquer l'interaction entre la plante (le lin) et son pathogène (*Melampsora lini*). Actuellement, il est communément admis que la notion de résistance est liée à la reconnaissance spécifique entre le produit du gène *R* de l'hôte et celui du gène *Avr* qui lui correspond.

Les NLR de plante ont deux modes principaux de reconnaissance de l'effecteur : un mode de reconnaissance directe et un mode de reconnaissance indirecte (Fig i.12). Dans les deux cas, les NLR de plante gardent une conformation inactive par des interactions intra ou intermoléculaires en absence de leurs effecteurs correspondants (Takken and Goverse, 2012).

Dans le cas de la reconnaissance directe, un effecteur est détecté par des interactions physiques avec sa protéine NLR correspondante. Ce mode de reconnaissance directe a été mis en évidence pour différents couples de protéines R et Avr, comme par exemple pour Pi-ta chez le riz et AvrPi-ta chez *M. grisea* (Jia et al., 2000), pour la protéine RRS1 chez *A. thaliana* avec l'effecteur PopP2 de *R. solanacearum* (Deslandes et al., 2003) ou encore les protéines de lin L5, L6 et L7 avec AvrL567 de la rouille du lin (Wang et al., 2007). En général le domaine du NLR impliqué dans l'interaction est le domaine LRR, cependant dans certains cas la reconnaissance directe peut se faire par des domaines particuliers comme c'est le cas pour Pik-1 (avec un domaine additionnel HMA : Heavy Metal Associated domain) et l'effecteur Avr-Pik (Maqbool et al., 2015) ou encore pour la protéine RGA5-A qui possède un domaine RATX1 (autre nom du domaine HMA) impliquée dans la reconnaissance directe de deux effecteurs de *M. oryzae*, Avr1-CO39 et Avr-Pia (Cesari et al., 2013).

Ces deux derniers exemples pourraient se rapprocher d'un mode de reconnaissance indirecte dans lequel la protéine leurre est directement intégrée dans le récepteur immunitaire. Ainsi la distinction entre mode direct et indirect, et surtout les variations possibles dans le mode de reconnaissance indirecte n'est qu'une façon arbitraire de classer les réseaux d'interactions, souvent complexes, impliqués dans cette phase initiale de la réponse immunitaire. Cette distinction pouvant évoluer en fonction des connaissances acquises pour un système donné (Wu et al., 2015).

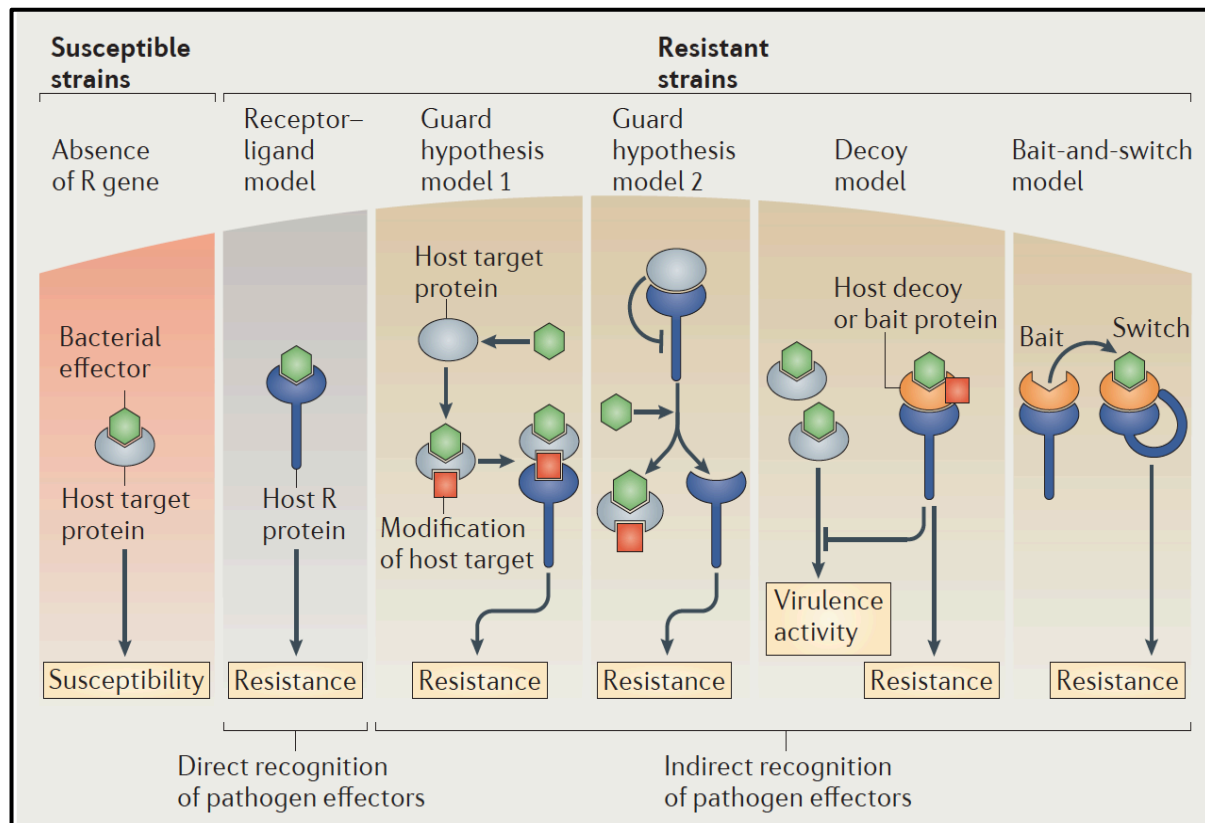


Fig i.12 Modèles de détection de l'effeteur par les protéines NB-LRR :

Lignées sensibles : En l'absence du gène de résistance R, l'effeteur va pouvoir s'associer à sa cible de virulence, la plante devient sensible à l'infection.

Lignées résistantes : La détection de l'effeteur peut se faire par deux types de détection via les protéines R :

- Une reconnaissance **directe** de type récepteur – ligand.
- Une reconnaissance **indirecte** selon plusieurs hypothèses:

Le modèle de garde : Dans ce modèle, certaines cibles cellulaires sont « surveillées » par des protéines R spécifiques, la réponse immunitaire sera induite s'il y a perturbation de la protéine cible. Dans le modèle 1, l'interaction entre la cible de l'hôte et l'effeteur conduit à la modification de la cible. Cette modification est reconnue par la protéine R ce qui déclenche la réponse immunitaire. Dans un modèle alternatif (modèle 2), la protéine de garde R et la cible hôte sont physiquement en interaction et la réponse immunitaire est réprimée. Pendant l'infection, la modification de la cible par l'effeteur libère la protéine R qui va s'engager dans des voies de signalisation permettant l'induction des réponses de défense.

Le modèle du leurre : Ce modèle suggère que pendant l'évolution, l'hôte acquière des protéines qui sont des leurres, semblables aux cibles de virulence des effecteurs. Ces protéines fonctionnent comme des sentinelles guettant la présence d'effecteurs. Ces leurres pourraient provenir d'événements de duplication génique ou pourraient être des variants d'épissage de la cible originale. Cependant, ces leurres sont généralement supposés n'avoir aucune autre fonction dans la cellule que d'inhiber la virulence de l'effeteur et de déclencher la résistance.

Le modèle de l'appât « Bait and Switch » : Dans ce cas l'effeteur n'est pas détecté après l'interaction avec sa cible de virulence mais plutôt « appâtée » par une protéine leurre liée à une protéine R. Les protéines R fonctionnent comme des commutateurs pour induire la signalisation une fois que l'effeteur est lié.

Adapté de (Stuart et al, 2013)

Dans le cas générique de la reconnaissance indirecte de l'effeteur, le NLR va percevoir sa présence en détectant des modifications liées à l'activité de l'effeteur sur les protéines-cibles. Plusieurs variantes de ce modèle sont décrites dans la figure i.12. Dans le modèle de garde ou du

leurre, une protéine cible (ou hôte) de la plante est modifiée par une activité enzymatique liée à la présence de l'effecteur. C'est cette protéine de l'hôte modifiée qui sera reconnue par le NLR. Par exemple, les effecteurs AvrRpm1 et AvrB augmentent l'expression de la protéine-kinase d'*Arabidopsis* RIPK qui phosphoryle la protéine de garde RIN4. Cette modification de RIN4 est reconnue par le NLR RPM1 permettant son activation (Chung et al., 2011; Khan et al., 2016; Mackey et al., 2002). La distinction entre un garde et un leurre se fait par la fonction de la protéine hôte dans le système immunitaire. Le garde possède une fonction dans l'immunité alors que le leurre n'en possède pas. Le leurre est en fait une protéine hôte similaire à une protéine ayant une fonction dans le système immunitaire certainement acquise lors de l'évolution. Supprimer un garde aura un impact sur la virulence de l'agent pathogène alors que supprimer un leurre n'en aura aucun. La distinction entre un garde et un leurre est souvent difficile, notamment *in vivo* (van der Hoorn and Kamoun, 2008; Stuart et al., 2013). Un exemple de leurre est celui de la pseudokinase ZED1, acétylée par l'effecteur HopZ1a et qui est alors reconnue par le NLR Zar1 (Lewis et al., 2013). Le modèle de détection indirecte avec un appât, n'implique pas de modification de la protéine hôte par l'activité de l'effecteur. Dans ce modèle l'effecteur reconnaît une protéine appât ou un cofacteur lié au NLR (Collier and Moffett, 2009), cette reconnaissance permettant l'activation de l'ETI par l'induction des réponses de défense. Dans la littérature, les exemples de détection indirecte par la modification de protéines hôte de la plante par l'effecteur sont les plus nombreux.

2.3.4 Signalisation en aval de la reconnaissance de l'effecteur

Les mécanismes moléculaires par lesquels les protéines NLR activent le système de défense après la reconnaissance des effecteurs restent largement méconnus (Jacob et al., 2013). Les voies de signalisation impliquées dans l'ETI rejoignent celles de la PTI, comme la production de ROS, le renforcement de la paroi cellulaire, la synthèse de molécules antimicrobiennes et l'activation de gènes de défense. Ces réseaux de signalisation sont régulés par des cascades de MAPK, des facteurs de transcription (type WRKY), des hormones végétales, des modifications épigénétiques ou encore des petits ARN (Liu et al., 2014a). Par exemple, au niveau des cascades impliquant les MAP Kinases, la protéine du riz OsMAPK5 réprime la résistance à un large panel de pathogènes infectant le riz (Xiong and Yang, 2003) ou encore OsMAPK6 qui est nécessaire à la transduction du signal impliquant le NLR Pit qui interagit avec la protéine OsRac1 (petite protéine G de la superfamille Ras) qui elle-même interagit avec un complexe comprenant quatre chaperonnes RAR1-HSP90-STG1-Hop en plus de la OsMAPK6 (Kawano et al., 2010). Certains NLR sont capables de réguler directement la transcription comme par exemple le récepteur CNL Pb1 chez le riz qui régule l'ETI en interagissant, par l'intermédiaire de son domaine CC avec le facteur de transcription de type WRKY OsWRKY45, impliqué dans une activité transcriptionnelle de défense conduisant à la résistance (Cui et al., 2015). Les actions des NLR dans le déclenchement de la réponse immunitaire ne sont pas localisées seulement dans le noyau mais aussi dans différents compartiments de la cellule. Ainsi ces NLR

activeraient des cascades de MAP-Kinases impliquant la phosphorylation de facteurs de transcription de type WRKY menant à une reprogrammation permettant la résistance. D'autres études sont nécessaires pour élucider les mécanismes par lesquels les NLR activent les MAPK et leur substrats et pour comprendre comment les MAPK se déplacent dans les différents compartiments cellulaires (Cui et al., 2015).

B. Le pathosystème riz - Magnaporthe oryzae

1. Le riz

Le riz est une céréale monocotylédone de la famille des Poacées (ou graminées). Cette famille regroupe d'autres céréales comme le blé, le seigle, l'orge ou le millet (Goff et al., 2002). Le genre *Oryza* compte au moins 21 espèces différentes qui peuvent être classées en deux types, les riz sauvages et les riz domestiqués. Seulement deux espèces de riz sont cultivées *Oryza glaberrima* (Afrique occidentale) et *Oryza sativa* (Asie), leur domestication est très ancienne, et remonte à plus de 6000 avant JC (Fuller, 2011). Les croisements entre espèces sauvages et cultivées, donnent au riz une diversité génétique considérable. L'espèce *O. sativa* correspond au riz commun asiatique majoritairement cultivé, elle est divisée en trois sous espèces *O. sativa indica*, *O. sativa japonica* et *O. sativa javanica* aussi appelée *tropical japonica* (Sweeney and McCouch, 2007). Le riz cultivé présente une grande diversité de formes et le nombre de cultivars (ou variétés), diversité qui provient de croisements naturels d'*O. sativa* avec des formes sauvages d'*O. rufipogon* ou de croisements *intra-sativa* combinés à la sélection naturelle et humaine depuis la domestication (Khush, 1987).

Bien avant la maîtrise des outils de génétique moderne, l'homme a donc sélectionné des variétés de riz en fonction de leur morphologie. Le plant de riz présente des racines, une tige (ronde et creuse, subdivisée en entre-noeuds par des nœuds), des feuilles (plates, longues et fines), et une panicule (partie terminale portant les grains). La sélection s'est faite sur les variétés de riz les plus résistantes aux maladies et présentant une panicule portant le plus de grains avec un meilleur rendement.

Le riz est une plante herbacée annuelle, diploïde avec 12 chromosomes et autogame (autofécondation). Comparé aux autres céréales, la taille du génome du riz est petite 460 Mega-paires de bases (Mb) contre 2500 Mb pour le Maïs (Wong et al. 2002). Cela en fait un modèle de choix pour développer des outils de biologie moléculaire qui faciliteront l'étude des céréales ayant un génome plus complexe.

Le riz est cultivé dans 114 pays répartis sur tous les continents, de l'Asie du Sud-est à l'Afrique en passant par l'Amérique et l'Europe (Skamnioti and Gurr, 2009). Cela représente 163 millions d'hectares cultivés avec une production de 491 millions de tonnes (Nachimuthu et al., 2015). L'Asie reste le continent principal pour la culture et la consommation du riz où 90% des récoltes sont produites et consommées. Les grands pays producteurs et consommateurs sont l'Inde, la Chine, l'Indonésie, le Bangladesh, la Thaïlande et le Vietnam. Le riz représente la première céréale pour l'alimentation humaine et constitue aujourd'hui la base de la nourriture de trois milliards de personnes, soit presque la moitié de la population mondiale. C'est une des céréales les moins échangées avec moins de 4% de la production mondiale. Les principaux producteurs sont des pays en développement pour lesquels la culture est essentielle pour alimenter leur population. À l'horizon 2030, au vue de la croissance démographique dans cette région du monde, la production de riz devra augmenter de 38% selon l'institut de recherche international pour la sécurité alimentaire (Wilson and Talbot, 2009).

Même si le riz peut pousser dans des environnements forts divers, sa culture nécessite un climat chaud et humide. Les changements climatiques comme le réchauffement sont des facteurs environnementaux qui impactent le rendement de cette culture. Cependant une des causes principales de la perte de culture provient des maladies contractées par la plante. En effet même si chez les plantes la maladie reste l'exception, lorsqu'elle s'installe elle est dévastatrice. Ces différents stress, à la fois biotiques et abiotiques, entraîne un fluctuation des prix avec de fortes augmentations notamment en cas d'épidémie (Azizi et al., 2014). De ce fait, les aléas des récoltes peuvent obliger un pays potentiellement exportateur à devenir ponctuellement importateur et donc à changer de politique commerciale.

Dans les années 1960, de nouvelles variétés de riz sont apparues sur le marché. Elles avaient un meilleur rendement, étaient plus résistantes aux intempéries, mais nécessitaient de forts apports d'engrais et de pesticides. Connue sous le nom de « Révolution Verte » ces nouvelles variétés, dont beaucoup ont été créées par l'IRRI (The International rice research institute), avaient pour but d'éliminer les famines et de rendre les pays émergents autosuffisants sur le plan alimentaire (Skamnioti and Gurr, 2009).

Avec plus de 900 millions de personnes sous alimentées dans le monde, la malnutrition est encore très présente notamment en Afrique sub-saharienne et en Asie, où elle provoque chaque année la mort de 3 à 5 millions d'enfants. À elles deux, la malnutrition et la famine, tuent plus que le sida, le paludisme et la tuberculose réunis. Le manque d'accès à une nourriture de base comme le riz participe à l'insécurité alimentaire des populations les plus vulnérables (source Unicef).

L'augmentation du rendement des cultures n'en reste pas moins un attrait économique très important d'où l'implication de grosses multinationales comme Monsanto dans le contrôle des semences. Même si ces mesures ont permis un meilleur rendement de production du riz, les conséquences sur l'environnement sont néfastes (pollutions, appauvrissement des sols, renforcement de la résistance des pathogènes).

A présent, les nouveaux challenges des biotechnologies se focalisent sur l'augmentation du rendement, tout en limitant les apports d'engrais, de pesticides et d'eau pour les cultures, ce qui

permettrait la mise en place de pratiques culturales plus performantes sans porter atteinte à l'équilibre environnemental et à la santé des agriculteurs.

La diversité génétique du riz représente une des clefs pour l'élaboration de nouvelles variétés. Les différents caractères du riz comme le rendement, la qualité ou encore la résistance aux stress biotique et abiotique, sont directement liés à différents gènes (Nachimuthu et al., 2015). La caractérisation du génome du riz est à l'heure actuelle très étudiée, mais l'enjeu est d'attribuer une fonction aux 40 000 à 50 000 gènes identifiés chez *Oryza sativa*. De nombreux gènes majeurs et QTLs (Quantitative Trait Loci) sélectionnés pour leurs traits agronomiques ont été identifiés et cartographiés grâce à une très bonne connaissance du génome et des marqueurs moléculaires (Ge et al., 1999; Huang et al., 2015).

L'un des problèmes majeurs de la riziculture reste la sensibilité du riz aux agents pathogènes et notamment aux champignons phytopathogènes. La maladie principale du riz est la pyriculariose qui est causée par le champignon *M. oryzae* (Fig i.13). Ce phytopathogène constitue une menace importante pour les cultures ainsi que pour la sécurité alimentaire mondiale (Caracuel-Rios and Talbot, 2007).



Fig i.13 Symptômes de la pyriculariose provoqué par *Magnaporthe. oryzae* :

- A. Champ de riz avec en premier plan une parcelle détruite par le champignon (photo D.Thareau)
 - B. Symptôme sur la feuille. L'apparition de ces lésions en forme d'œil avec un contour plus foncé est caractéristique de la maladie.
 - C. Panicule infectée conduisant au dessèchement des grains.
 - D. Grandes lésions d'où les spores sont libérées au niveau de la partie blanche.
- Adapté de (Dean et al., 2012; Wilson and Talbot, 2009)

2. La pyriculariose

Le champignon *M. oryzae* s'attaque à toutes les parties aériennes de la plante. Le champignon infecte souvent les nœuds de la tige du riz ce qui peut conduire à la perte totale des grains. En effet, en cas d'infection, la panicule portant les graines n'a plus d'apport en nutriments provenant de la plante car les tissus contaminés de la tige en amont se nécrosent. Dans les premiers stades de l'infection, des lésions nécrotiques en forme d'œil, apparaissent au niveau des feuilles. Le résultat est une diminution de la capacité photosynthétique de la plante, suivi par un dessèchement aussi bien des feuilles que des grains. En général, les pertes sont estimées entre 10 et 30% de la récolte, et une épidémie peut être dévastatrice et détruire 100% de la production (Dean et al., 2012; Liu et al., 2014)

Le champignon *M. oryzae* fait partie des espèces complexes à combattre car il peut s'attaquer à plusieurs variétés de plantes et de céréales comme l'orge, le blé et le millet. Les méthodes de lutte contre la maladie impliquent des traitements par fongicides, des pratiques culturales particulières et l'utilisation de variétés qui portent des résistances génétiques (Pennisi, 2010). Cependant, les fongicides spécifiques à la pyriculariose se révèlent peu efficaces lors d'épidémies et trop coûteux pour beaucoup de riziculteurs. Le moyen le plus économique et écologique de contrer ce pathogène est d'utiliser des variétés de riz résistantes. Cependant, souvent, le champignon surmonte rapidement la résistance et les variétés redeviennent sensibles après deux à trois ans (Dean et al., 2012).

La connaissance de la biologie, de la diversité génétique et de la capacité d'adaptation de cet agent pathogène est essentielle pour pouvoir développer de nouvelles stratégies durables et efficaces pour lutter contre les maladies fongiques.

Historiquement le champignon pathogène causant la pyriculariose était connu sous le nom de *M. grisea*. Depuis 2002, deux espèces distinctes dans le genre *Magnaporthe* ont été mises en évidence (Couch and Kohn, 2002), *M. grisea* pathogène de *Digitaria* (mauvaise herbe appartenant à la famille des graminées) et *M. oryzae* pathogène du riz et d'autres graminées (Wilson and Talbot, 2009).

Le champignon *M. oryzae* est un champignon ascomycète, haploïde, filamenteux. Sa reproduction peut être sexuée ou asexuée (Fig i.14.A).

2.1 Cycle infectieux de *Magnaporthe oryzae*

M. oryzae est un pathogène hémibiotrophe, ainsi, pendant la phase précoce de l'infection, il se nourrit sur des cellules de plantes vivantes et doit donc les maintenir en vie (biotrophie). Ensuite (4-5 jours après le début de l'infection), il change de stratégie et tue les cellules de riz, pendant les phases tardives de l'infection, se nourrissant alors du tissu mort (nécrotrophie).

L'infection foliaire est initiée par l'adhésion d'une spore (conidie) de *M. oryzae* sur le riz (Fig i.14.B). La conidie ainsi ancrée solidement à la surface hydrophobe de la feuille peut germer en quelques heures lors de conditions favorables (température autour de 25°C et hygrométrie > 90%). Un tube germinatif se forme puis se différencie en une structure appelée appressorium (entre 2 et 20 heures après l'adhésion). Au cours de sa maturation, l'appressorium forme une paroi particulière, riche en chitine et doté d'une couche de mélanine, le rendant imperméable aux solutés mais perméable à l'eau (Wilson and Talbot, 2009). Cette structure spécialisée va générer une pression de turgescence énorme (jusqu'à 8MPa, ou 80 bar) qui va rompre la cuticule de la feuille, par la pression physique qu'elle exerce, et ainsi permettre l'invasion des tissus par un hyphe de pénétration (20-28h) (Dean et al., 2005)

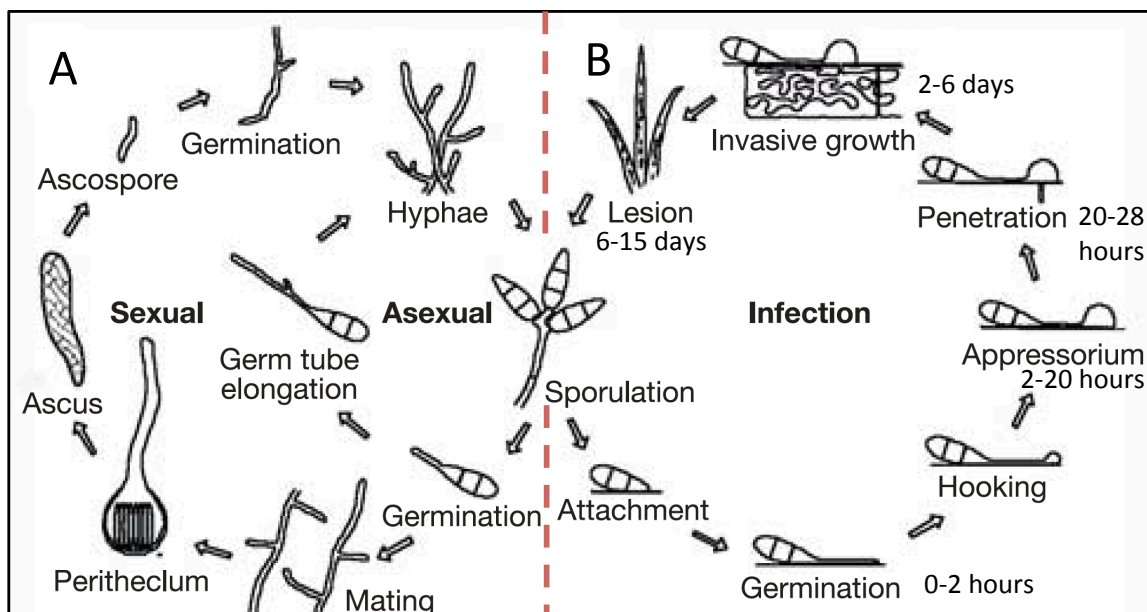


Fig i.14 Cycle de vie de *M. oryzae* :

A. Cycle de reproduction

B. Cycle d'infection :

Le cycle infectieux débute par l'adhésion d'une spore à la surface de la feuille de riz. La spore germe et l'hyphe de germination se différencie en appressorium mélanisé qui permet au champignon de s'insérer dans la feuille. L'hyphe invasif pénètre dans la première cellule épidermique avant d'envahir les cellules avoisinantes. Le champignon prolifère dans la plante (2 à 6 jours après l'adhésion) avant de former des lésions sporulantes (6 à 15 jours après infection).

Adapté de (Dean et al., 2005; Ribot et al., 2008)

A ce stade, si le champignon est « détecté » par la plante, par la reconnaissance des effecteurs, les cellules en contact avec le champignon réagissent par une réponse immunitaire, appelée réponse hypersensible, qui conduit à la mort des cellules de plante envahies par le champignon, ce qui évite la propagation du champignon et conduit à la résistance de la plante.

Dans le cas contraire, l'invasion des tissus végétaux se produit au moyen d'un hyphe invasif qui se ramifie en bulbes (Fig i.16). Ces structures sont invaginées dans la membrane plasmique de la cellule de riz. Par ses ramifications le champignon va envahir les cellules épidermiques et mésophylliques adjacentes en passant par les plasmodesmes (Fig i.15, Fig i.16) (microcanaux intercellulaires).

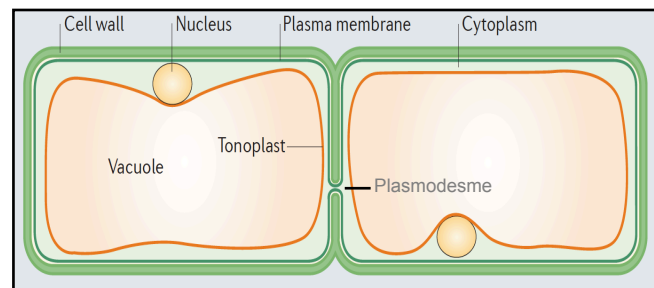


Fig i.15 Description d'une cellule végétale :

La cellule végétale possède deux compartiments, le cytoplasme du côté interne de la membrane plasmique et le compartiment apoplasmique (également connu comme compartiment extracellulaire ou paroi cellulaire) du côté externe de la membrane plasmique.

Les plasmodesmes (canaux microscopiques) permettent le transport de petites molécules ou de protéines entre les cellules végétales, et relient les cytoplasmes de la plupart des cellules dans les tissus végétaux. Les cellules de garde (constituées par les stomates) manquent totalement au niveau des plasmodesmes. Les cellules végétales contiennent généralement une grande vacuole séparée du cytoplasme par une membrane vacuolaire, connue sous le nom tonoplaste.

Adapté de (Giraldo and Valent, 2013)

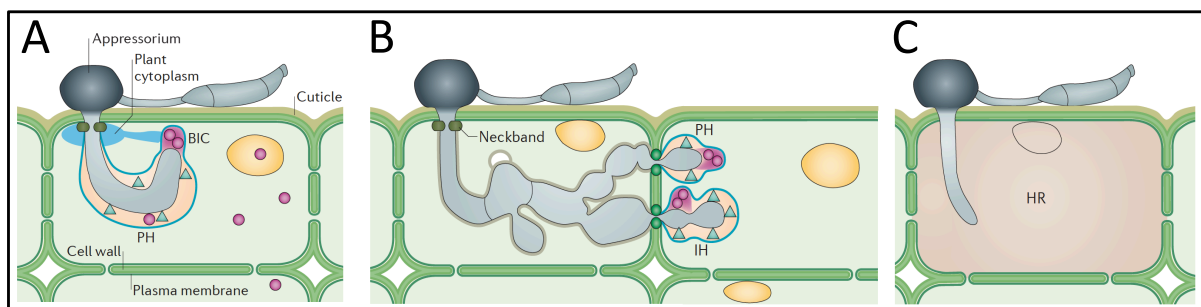


Fig i.16 Schéma de l'infection d'une cellule végétale par *M. oryzae* :

L'hyphe infectieuse du champignon pénètre la cellule végétale via l'apressorium.

A. Formation du complexe biotrophe interfacial (BIC) à l'extrémité de l'hyphe permettant la translocation dans le cytoplasme d'effecteurs favorisant l'infection.

B. Ramification du champignon en hyphes secondaires permettant la colonisation de la cellule par les plasmodesmes.

C. Réponse hypersensible de la cellule. La cellule ayant reconnu le pathogène meurt pour éviter la propagation.

Adapté de (Giraldo et al., 2013)

C'est dans un second temps (4-7 jours après l'infection) que le champignon devient nécrotrophe (Fig i.16). Lors de cette phase, des toxines sont sécrétées par le pathogène, qui ont pour effet d'induire la mort de la cellule hôte (Giraldo and Valent, 2013; Koeck et al., 2011). Les lésions

caractéristiques de la maladie apparaissent entre 4 et 5 jours après l'infection (Fig i.13D). Les spores de champignon vont se répandre à partir de ces lésions dans des conditions de forte humidité (7-10 jours), permettant à la maladie de se propager rapidement aux plants de riz adjacents par le vent et les gouttes de rosée (Wilson and Talbot, 2009).

2.2 Les effecteurs du pathogène *M. oryzae*

Les effecteurs jouent un rôle important lors du processus d'infection, notamment en favorisant la virulence du champignon. Ils sont également la cible des récepteurs NLR du riz, permettant ainsi la résistance de la plante.

• *Effecteur de type Avr*

Le génome de *M. oryzae* est d'environ 40Mb avec 7 chromosomes (Dean et al., 2005). La petite taille de ce génome a tendance à favoriser une grande plasticité génomique, ce qui représente un atout puissant pour lutter contre la pression de sélection de l'hôte (Zhang et al., 2015). Dans les interactions riz-*M. oryzae*, plus de 40 gènes *Avr* (Zhang and Xu, 2014) ont été identifiés et 9 ont été clonés (Lo Presti et al., 2015). Les protéines *Avr* appartiennent à une classe spéciale d'effecteurs codés par les gènes d'avirulence, pouvant être reconnus par leurs protéines R correspondantes et conduisant à une reconnaissance spécifique (Zhang and Xu, 2014). Parmi les gènes *Avr* qui ont été clonés, tous sauf *ACE1* (Avirulence Conferring Enzyme1) codent des protéines secrétées, exprimées au niveau de l'hyphe invasif. *ACE1* est spécifiquement exprimé dans l'appressorium et il code une protéine hybride intracellulaire de type PKS-NRPS (polyketine synthase et la non-ribosomal peptide synthetase) impliquée dans la production d'un métabolite secondaire. Ce dernier serait reconnu par la protéine de résistance Pi33 permettant la signalisation conduisant à la réponse hypersensible (Bohnert, 2004).

Avr-Piz-t et *Avr-Pita1* sont deux autres protéines *Avr* avec des fonctions biochimiques connues. Une des fonctions d'*Avr-Piz-t* est de supprimer la PTI en inhibant l'activité ubiquitine ligase RING E3 de la protéine APIP6 (*Avr-Piz-t*-Interacting Protein 6) (Park et al., 2012). *AvrPi-ta1* code pour une possible métallo-protéase à zinc (Jia et al., 2000; Khang et al., 2008) et elle appartient à une famille de gènes *AvrPi-ta* (gènes *AvrPi-ta1*, *AvrPi-ta2* et *Avr-Pi-ta3*). D'une manière générale, la cible de virulence des effecteurs reste inconnue, et ce parce que les effecteurs *Avr* sont identifiés par l'intermédiaire de leur protéine de résistance associée, également appelée cible d'avirulence. Cependant, leur fonction première n'est certainement pas celle de se faire reconnaître par la plante. La famille de gènes *PWL* (Pathogenicity toward Weeping Lovegrass) comporte 4 gènes (*PWL1*, *PWL2*, *PWL3*, et *PWL4*) (Schneider et al., 2010). Les effecteurs *Pwl* sont des petites protéines riches en glycines dont la seule fonction connue est celle d'avirulence (Schneider et al., 2010; Zhang and Xu, 2014). L'effecteur *Avr1-CO39* est également exprimé au niveau de l'hyphe invasif, sa reconnaissance

par les produits du gène *Pi-CO39* déclenche la réponse hypersensible (Ribot et al., 2013). D'autres effecteurs comme Avr-Pia, Avr-Pik/km/kp, et Avr-Pii ont été identifiés. Avr-Pia interagit directement avec le récepteur RGA5-A qui interagit également avec Avr1-CO39 (Cesari et al., 2013).

- **Effecteur de type non-Avr**

M. oryzae possède d'autres effecteurs, qualifiés de non-Avr. L'un des effecteurs appartenant à cette classe le mieux caractérisé est Slp1 (secreted LysM protein 1). Cette protéine est retrouvée au niveau de l'hyphe invasif et de sa membrane. Slp1 n'est pas nécessaire pour la pénétration de l'appressorium mais est importante pour le développement de l'hyphe invasif dans la plante. La fonction de Slp1 est d'éviter la reconnaissance de la chitine par CEBiP en se liant aux oligosaccharides de chitine provenant de la paroi cellulaire du champignon. Ainsi Slp1 supprime la réponse immunitaire non-spécifique (PTI) de la plante induite par la reconnaissance de la chitine.

Des analyses transcriptomiques ont identifié un grand nombre de protéines sécrétées à faible poids moléculaire, exprimées spécifiquement ou fortement pendant la phase biotrophe de l'invasion. Certains de ces effecteurs candidats ont été référencés comme des BAS (biotrophy associated secreted protein) dont 4 ont été caractérisés pour leur localisation au cours de l'infection (Mosquera et al., 2009; Zhang and Xu, 2014).

Un autre effecteur MC69 a été identifié par une approche de mutagénèse systématique de gènes codant des petites protéines sécrétées pendant l'infection. Un mutant *mc69* montre une réduction importante des symptômes de la pyriculariose indiquant l'importance de MC69 dans la pathogénicité de *M. oryzae*. Le gène *MC69* code une protéine de 54 acides aminés. Cette protéine est essentielle au développement de l'hyphe invasif et n'est pas transloquée dans le cytoplasme de l'hôte (Saitoh et al., 2012).

- **Localisation, sécrétion et translocation**

Les effecteurs de *M.oryzae* peuvent être divisés en deux types différents, basés sur leur localisation dans les cellules de plantes (Giraldo et al., 2013), les effecteurs cytoplasmiques et les effecteurs extracellulaires.

Il a été observé que des effecteurs cytoplasmiques, tel que Avr-Pita, Pwl1, Pwl2, Bas2, et Avr-Piz-t s'accumulent préférentiellement au niveau du BIC (biotrophic interfacial complexe). Le BIC est une structure ou protubérance membranaire, dont l'origine est végétale (cf Giraldo et al., 2013), qui se développe en surface de la membrane plasmique à l'extrémité du premier hyphe invasif. Il a été proposé que dans chaque nouvelle cellule infectée, les effecteurs sont d'abord sécrétés au niveau du BIC et sont ensuite internalisés dans le cytoplasme de la cellule hôte. Le BIC persiste après la différenciation en hyphes secondaires et le champignon continue à sécréter des effecteurs via le BIC

lors de l'invasion des cellules voisines (Khang et al., 2010). Il a été suggéré que le mécanisme de sécrétion au niveau du BIC implique le complexe d'exocytose et les protéines SNARE (Soluble N-éthylmaleimide-sensitive-factor Attachment protein REceptor) (Fig i.17). L'exocytose est un complexe octamérique, comprenant entre autre les protéines Exo70 et Sec5, impliquées dans la régulation spatio-temporelle du trafic membranaire (notamment la polarité). Les protéines SNARE sont impliquées dans l'attachement et la fusion des vésicules au niveau de la membrane plasmique (Giraldo et al., 2013).

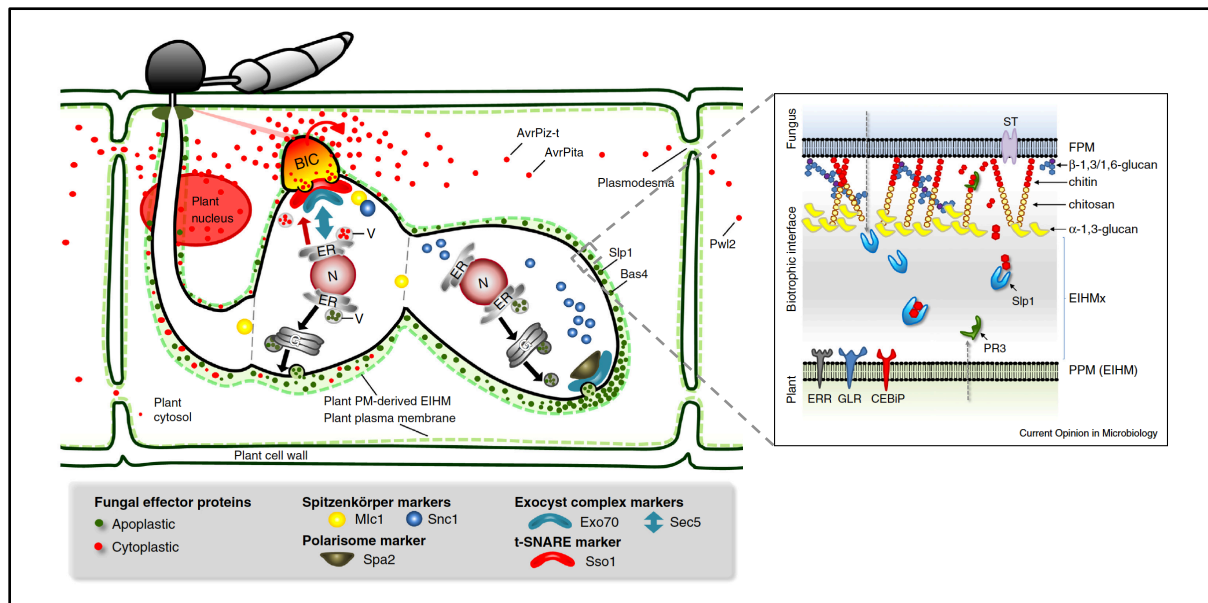


Fig i.17 Modèle de sécrétion d'effecteurs par *M. oryzae* :

Cette figure représente le bulbe invasif formé par le champignon à l'intérieur d'une cellule de riz.

Les effecteurs cytoplasmiques (en rouge) suivent la voie de sécrétion impliquant l'exocytose et les protéines SNARE via le BIC (en orange).

Les effecteurs apoplastiques (en vert) suivent la voie classique de sécrétion par le système de Golgi.

Le zoom sur la droite représente l'interface biotrophe entre le champignon et la plante.

Abréviations : FPM (Fungal Plasma Membrane) ; EIHMx (Extra Invasive Hyphae Matrix) ; EIHM (Extra Invasive Hyphae Membrane) PPM (Plant Plasma Membrane)

(Adapté Oliveira-Garcia and Valent, 2015)

Les effecteurs extracellulaires ou apoplastiques, tels que Bas4 et Slp1, ne sont pas ou peu associés avec le BIC. Ils sont sécrétés via le réticulum endoplasmique, puis le système de sécrétion de l'appareil de Golgi (Giraldo et al., 2013). Ce mécanisme est inhibé par la Bréfeldine A, une petite molécule qui cible des petites protéines G et les facteurs d'échange (GTP/GDP) associés, qui est un réseau régulant le transit des vésicules au niveau du Golgi. Ce mécanisme nécessite l'intégrité du cytosquelette fongique comme le montrent des expériences avec des inhibiteurs de la formation de microtubules (Giraldo et al., 2013). Après sécrétion, ces effecteurs sont dispersés dans l'interface entre la membrane plasmique fongique et la membrane plasmique végétale. A l'heure actuelle, les

mécanismes permettant aux effecteurs de l'apoplasme d'être reconnus par des récepteurs et les processus de translocation dans le cytoplasme de la cellule hôte sont encore mal compris.

Finalement, quel que soit le type d'effecteurs 'cytoplasmique' ou ceux qui transitent par l'apoplasme, aucun motif, séquentiel ou de repliement de ces protéines, qui pourrait être impliqué dans la localisation cellulaire après sécrétion n'a été identifié. C'est pourquoi il est impossible de prédire actuellement à partir de la séquence primaire si un effecteur est de type 'apoplastique' ou 'cytoplasmique' (Sperschneider et al., 2015; Zhang and Xu, 2014).

- **Modifications post-traductionnelles**

Les études récentes montrent que les effecteurs subissent des modifications après traduction qui sont nécessaires pour les convertir dans leur forme active. Par exemple, Slp1, qui ne lie la chitine qu'une fois N-glycosylé, ou encore l'effecteur de *M. oryzae* Bas4 qui est également N-glycosylé (Chen et al., 2014). Plusieurs effecteurs, comme Avr4 et Avr9 de *C. fulvum* et SnToxA de *S. nodorum*, doivent être modifiés pour devenir actifs. Ces protéines sont synthétisées sous forme de précurseurs, et après clivage du peptide signal, des modifications additionnelles peuvent également intervenir, mais pour l'instant peu de connaissances sont disponibles (Stergiopoulos and de Wit, 2009). D'autres effecteurs comme ceux appartenant à la classe des SCR (small-cystein-rich) possèdent un peptide signal, nécessaire pour l'adressage au réticulum endoplasmique, et qui sera clivé par des signal-peptidases. D'autres processus de maturation pourront survenir dans les vésicules du Golgi pendant le transit vers le pôle d'exocytose.

2.3 Les protéines R chez le riz

L'analyse du génome du riz (466Mbp) prédit 456 gènes de NLR dont 274 CNL, 0 TNL et 182 XNL (X correspondant à un domaine N-terminal autre que TIR ou CC) (Jacob et al., 2013). Une centaine de gènes *R* impliqués dans la résistance à la pyriculariose causée par *M. oryzae* ont été identifiés et 23 d'entre eux ont été clonés (table i.1) (Liu et al., 2014). La majorité des gènes *R* code pour des protéines de type CNL, excepté *Pi-d2* qui code pour une RLK (Chen et al., 2006).

Les études d'interaction moléculaire se sont concentrées sur cinq paires de gènes de résistance *R* chez le riz et d'avirulence *Avr* chez le champignon (*Pi-ta/AvrPi-ta*, *Pik/Avr-Pik*, *Piz-t/AvrPiz-t*, *Pia/Avr-Pia* et *Pi-Co39/Avr1-CO39*) (Liu et al., 2014).

Dans les cas de *Pi-ta/AvrPi-ta* et *Piz-t/AvrPiz-t*, la reconnaissance d'un seul *Avr* se fait par un seul produit de gène *R* dominant. L'interaction entre *Pita* et *AvrPi-ta* fut la première interaction directe mise en évidence entre un effecteur *Avr* de champignon et une protéine *R* de plante. L'interaction physique entre les partenaires est abolie par une seule substitution d'acide aminé dans la région LRR de *Pita*,

ce qui suggère un rôle essentiel du domaine LRR de Pita dans la reconnaissance de l'AvrPi-ta (Jia et al., 2000). Même si l'interaction directe n'a pas encore été détectée, l'expression d'AvrPiz-t et de Piz-t dans des plants de riz transgéniques conduit à la mort cellulaire ce qui montre qu'il existe une reconnaissance entre les produits de cette paire de gènes (Liu et al., 2014).

Dans d'autres cas, deux protéines NLR sont requises pour l'activation de la résistance, (Pik-1/Pik-2, Pi5-1 /Pi5-2, RGA4/ RGA5) (Cesari et al., 2014a; Ashikawa et al., 2008; Lee et al., 2009; Okuyama et al., 2011). Aucune de ces protéines NLR lorsqu'elles sont seules n'est capable d'activer la résistance en présence de son effecteur Avr correspondant.

Resistance genes		Avirulence effectors		
R gene	Encoding protein	Avr gene	Encoding protein	Pathogen
<i>Pib</i>	NB-LRR	ND	Unknown	<i>Magnaporthe oryzae</i>
<i>Pi-ta</i>	NB-LRR	<i>AvrPi-ta</i>	224 AA secreted protein	
<i>Pi9</i>	NB-LRR	ND	Unknown	
<i>Pi2</i>	NB-LRR	ND	Unknown	
<i>Piz-t</i>	NB-LRR	<i>AvrPiz-t</i>	108 AA secreted protein	
<i>Pi-d2</i>	B-lectin RLK	ND	Unknown	
<i>Pi33^c</i>	Unknown	ACE1	Polyketide synthase	
<i>Pi^f</i>	Unknown	<i>AvrPij</i>	70 AA secreted protein	
<i>Pi36</i>	NB-LRR	ND	Unknown	
<i>Pi37</i>	NB-LRR	ND	Unknown	
<i>Pikm^a</i>	NB-LRR	<i>Avr-Pik/km/kp</i>	113 AA secreted protein, five alleles (A-E)	
<i>Pit</i>	NB-LRR	ND	Unknown	
<i>Pi5^a</i>	NB-LRR	ND	Unknown	
<i>Pid3</i>	NB-LRR	ND	Unknown	
<i>Pid3-A4</i>	NB-LRR	ND	Unknown	
<i>Pi54</i>	NB-LRR	ND	Unknown	
<i>Pish</i>	NB-LRR	ND	Unknown	
<i>Pik</i>	NB-LRR	<i>Avr-Pik/km/kp</i>	113 AA secreted protein, five alleles (A-E)	
<i>Pikp</i>	NB-LRR	<i>Avr-Pik/km/kp</i>	113 AA secreted protein, five alleles (A-E)	
<i>Pia^{a,b}</i>	NB-LRR	<i>Avr-Pia</i>	85 AA secreted protein	
<i>Pi-CO39^{a,b}</i>	NB-LRR	<i>Avr1-CO39</i>	89 AA secreted protein	
<i>Pi25</i>	NB-LRR	ND	Unknown	
<i>Pi1</i>	NB-LRR	ND	Unknown	
<i>pi21</i>	Proline-containing protein	ND	Unknown	
<i>Pb1</i>	NB-LRR	ND	Unknown	
ND	Unknown	PWL1	Unknown	
ND	Unknown	PWL2	145 AA secreted protein	

Table i.1 Gène de résistance cloné chez le riz et effecteur avirulent de *M. oryzae* correspondant :

^a La fonction de ces trois gènes R nécessite deux protéines NB-LRR

^b Ces deux gènes R partagent le même locus du gène NB-LRR

^c Le gène n'a pas encore été cloné

Abréviations : AA acides aminés, NB-LRR nucleotide binding leucine rich repeat, ND non déterminé, RLK receptor-like kinase.

(Adapté de Liu et al., 2014)

Le locus de résistance Pik possède au moins 6 allèles (*Pik*, *Pikm*, *Pikp*, *Piks*, *Pikh*, et *Pi1*) ce qui lui donne différentes spécificités (Kanzaki et al., 2012; Liu et al., 2014). L'effecteur correspondant, AvrPik, est également très polymorphe avec au moins 5 allèles différents (*Avr-Pik-A*, *B*, *C*, *D*, et *E*). Des tests en double hybride dans la levure montrent que chaque protéine Avr interagit avec au moins

un ou deux allèles du NLR Pik-1 (exemple : Pikp-1 avec Avr-Pik-D; Pik-1 avec Avr-Pik-D et -E; et Pikm-1 avec Avr-Pik-A, -D, et -E). Cela suggère que différents allèles d'AvrPik sont reconnus et interagissent avec différents allèles de Pik (Kanzaki et al., 2012). La raison pour laquelle deux protéines de résistance sont nécessaires pour l'activation de la signalisation n'est pas connue. Il est possible que l'une des protéines R soit impliquée dans la reconnaissance de l'effecteur et que l'autre protéine R soit impliquée dans la transduction du signal conduisant à la réponse de résistance (Liu et al., 2014).

Le locus de résistance *Pia* consiste en deux gènes adjacents de type NLR (RGA4 et RGA5) qui sont orientés dans des directions opposées (Okuyama et al., 2011). L'expression transitoire de l'effecteur Avr-Pia et de Pia, dans des protoplastes de riz, cause une mort cellulaire rapide ce qui indique une reconnaissance spécifique entre Avr-Pia et Pia dans les cellules de riz (Okuyama et al., 2011). Curieusement, les protéines RGA4 et RGA5 sont retrouvées au niveau du locus de résistance Pi-CO39. Les mutants ne possédant pas RGA4 deviennent sensibles aux souches de *M. oryzae* possédant l'effecteur Avr1-CO39, et des lignées de riz transgéniques exprimant à la fois RGA4 et RGA5 deviennent résistantes. Ces résultats suggèrent que la paire de gènes *RGA4* et *RGA5* est responsable du phénotype de résistance associé au locus *Pi-CO39* (Cesari et al., 2013). *RGA5* possède deux transcrits provenant d'épissage alternatif, RGA5-A et RGA5-B, et les effecteurs Avr-Pia et Avr1-CO39, qui ne partagent aucune homologie de séquence, interagissent tous les deux de façon directe avec RGA5-A mais pas avec RGA5-B. Plus précisément ces deux effecteurs interagissent au niveau d'une petite région C terminale de RGA5-A, homologue à un domaine de liaison au cuivre chez *Saccharomyces cerevisiae* (ATX1) d'où les termes RATX1 ou encore HMA (Heavy Metal Associated domain) (Cesari et al., 2013). Ce domaine est également retrouvé dans une autre protéine R de riz, Pik-1, suggérant que ce domaine additionnel pourrait intervenir dans la reconnaissance de divers effecteurs. Le mécanisme par lequel RGA5 interagit avec les effecteurs Avr reste encore inconnu (Liu et al., 2014).

La plupart des gènes R (22 sur 23) de résistance à la pyriculariose sont dominants. L'exception est pi21, un gène récessif donnant une résistance durable à la maladie et qui est utilisé avec succès dans les cultures. Le gène Pi21 code pour une protéine riche en proline, qui constitue un domaine probable de liaison aux métaux lourds, pouvant être impliqué dans des interactions protéine – protéine (Fukuoka et al., 2009).

3. Projet de thèse

A l'heure actuelle, la pyriculariose provoquée par le champignon *M. oryzae* constitue l'une des principales maladies du riz, avec des épidémies récurrentes et très destructrices qui touchent toutes les régions rizicoles du monde. L'étude de cette maladie représente des enjeux économique et humanitaire importants notamment au niveau de la sécurité alimentaire. En outre, le statut de céréale modèle du riz, et le statut de champignon parasite modèle de *M. oryzae*, font de l'interaction entre ces deux partenaires un modèle privilégié pour l'étude du contrôle génétique et moléculaire des relations hôte-pathogène. Les mécanismes de reconnaissance entre effecteurs et protéines NLR sont encore mal compris. Dans cette optique, l'objectif de cette thèse est une approche d'étude structure-fonction des effecteurs de *M. oryzae* et de leurs protéines R associées.

Au commencement de cette thèse, aucune structure d'effecteurs de *M.oryzae* n'est disponible dans la PDB (protein data bank). La séquence primaire de ces Avr est très divergente et ne correspond à rien de connu. Mon projet de thèse consiste, dans un premier temps, à déterminer la structure tridimensionnelle des effecteurs Avr1-CO39 et Avr-Pia de *M. Oryzae* par Résonance Magnétique Nucléaire. Ces effecteurs ont été caractérisés par nos collaborateurs du BGPI (Biologie et Génétique des Interactions Plantes Pathogènes) de Montpellier comme interagissant avec la même protéine NLR RGA5-A *in planta*. L'identification du domaine RAXT1, également appelé HMA, impliqué dans l'interaction avec les effecteurs Avr1-CO39 et Avr-Pia est la première démonstration de double spécificité pour une même paire de NLR. L'identification des acides aminés clés impliqués dans ces interactions est d'une grande importance pour comprendre le mécanisme de reconnaissance d'un point de vue moléculaire et sera abordée dans la deuxième partie de cette thèse.

Pour mener à bien ce projet, plusieurs techniques ont été utilisées. Toute étude structurale nécessite une grande quantité de protéine d'intérêt pure, et dans le cas d'une étude par RMN, marquée par du ¹⁵N et du ¹³C. La biochimie représente donc une étape cruciale et incontournable dans cette étude. Du fait de leur faible poids moléculaire, d'environ une dizaine de kilo Dalton, la structure des protéines Avr a été étudiée par RMN liquide. La détermination des structures finales a été réalisée par divers logiciels de modélisation. Les résultats de cette première partie sont abordés dans le chapitre 1 de cette thèse. La caractérisation de l'interaction a été réalisée par différentes méthodes de biophysiques et a permis une cartographie de la surface d'interaction d'Avr-Pia avec RAXT1. Cette étude est complétée par des résultats obtenus *in planta* par nos collaborateurs du BGPI. L'ensemble de ces données est décrit dans le chapitre 2.

Les résultats des différents chapitres sont ensuite discutés dans une dernière partie, permettant également de tirer le bilan de cette étude et les perspectives envisageables.

CHAPITRE 1

Analyse structurale d'effecteurs de *Magnaporthe oryzae*

1. Introduction

La détermination de la structure d'une protéine fournit des informations variées souvent reliées à la fonction de la protéine, notamment en termes de surface potentielle d'une interaction ou encore de site de liaison d'un ligand. Dans le cas où la structure ne fournit pas de lien direct avec la fonction, il est crucial d'étudier le contexte biologique de l'activité de la protéine. L'analyse de la structure des effecteurs de *M. oryzae* est un challenge notable puisqu'au départ du projet aucune donnée structurale pour ces effecteurs n'est connue dans la littérature.

Protéines	Origine	Fonction	Repliement	PDB
NLP	<i>P. aphanidermatum</i>	Toxine cytotolytique	Toxine	3GNU
AvrPtoB250-359-BAK1 kinase domain	<i>P. syringae</i> et <i>A. thaliana</i> , respectivement	Inhibiteur de kinase	4-helix bundle-kinase	3TL8
AvrPtoB121-205-Pto	<i>P. syringae</i> et <i>S. pimpinellifolium</i> , respectivement	Inhibiteur de kinase	4-helix bundle-kinase	3HGK
AvrPtoB436-553	<i>P. syringae</i>	Ubiquitine ligase	4-helix bundle	2FD4
AvrPto	<i>P. syringae</i>	Inhibiteur de kinase	3-helix bundle	1R5E
AvrPto-Pto	<i>P. syringae</i> et <i>S. pimpinellifolium</i> , respectivement	Inhibiteur de kinase	3-helix bundle-kinase	2QKW
HopU1	<i>P. syringae</i>	mADP-RT	Cholera toxin-like	3U0J
AvrPphF	<i>P. syringae</i>	Inconnue (mADP-RT?)	Diphtheria toxin-like	1S28
AvrPphB (HopAR1)	<i>P. syringae</i>	Protéase	Papain-like protease	1UKF
AvrB	<i>P. syringae</i>	Inconnue	Fido	1NH1
AvrB-ADP	<i>P. syringae</i>	Inconnue	Fido	2NUN
AvrRps4C	<i>P. syringae</i>	Inconnue	CC	4B6X
AVR3a11	<i>P. capsici</i>	Inconnue	WY	3ZR8
PexRD2	<i>P. infestans</i>	Inconnue	WY	3ZRG
AVR3a4	<i>P. capsici</i>	Inconnue	WY	2LC2
ATR1	<i>H. arabidopsidis</i>	Inconnue	WY	3RMR
ATR13	<i>H. arabidopsidis</i>	Inconnue	α -helical	2LAI
AvrL567-D	<i>M. lini</i>	Inconnue	β -sandwich	2QVT
AvrL567-A	<i>M. lini</i>	Inconnue	β -sandwich	2OPC
AvrPiz-t	<i>M. oryzae</i>	Inconnue	β -sandwich	2LW6

Tableau 1.1 Liste d'effecteurs dont la structure a été déterminée. (Adapté de Wirthmueller et al., 2013)

L'intérêt pour l'approche structurale est en plein essor, depuis la résolution des structures d'effecteurs oomycètes de type RxLR (Boutemy et al., 2011; Win et al., 2012). L'analyse de ces structures a permis d'identifier un domaine conservé de type WY-four-helix-bundle permet de stabiliser le cœur hydrophobe de ces protéines. La conservation de ce domaine WY a été mis en évidence par une analyse des séquences de protéomes d'oomycètes, avec 42% des effecteurs d'oomycètes de type RxLR contenant probablement un domaine WY (Boutemy et al., 2011 ; Win et al., 2012). Ces dernières années montrent un fort intérêt pour l'investigation des structures des effecteurs d'virulence (Avr) (Cf tableau 1.1) ainsi que de leurs récepteurs NLR associés. Une des questions qui reste encore peu abordée dans l'étude des effecteurs de pathogènes fongiques est de comprendre les mécanismes associés à leur virulence et les cibles potentielles dans la cellule hôte. Dans ce cadre, la connaissance des cibles de ces effecteurs est essentielle pour comprendre la fonction en relation avec la structure (Wirthmueller et al., 2013).

1.1 Les effecteurs de *M. oryzae*

L'analyse du secretome de *M. oryzae* prédit des milliers de protéines (~1700 protéines sécrétées d'après Dong et al., 2015) et seulement quelques unes ont une fonction connue (Liu et al., 2014). Un tri basé sur la séquence primaire seule est difficile mais révèle l'existence d'une classe d'effecteurs caractérisée par une petite taille (souvent moins de 200 acides aminée) et la présence de plusieurs résidus cystéines. La très grande diversité des séquences de ces effecteurs potentiels renforce la difficulté à extraire une information pertinente quand à l'existence de familles d'effecteurs. Cependant, la classification de ces effecteurs pourrait fournir des éléments importants pour comprendre leur virulence ou encore la spécificité hôte-pathogène dans le cadre du phénomène d'virulence. A l'heure actuelle seulement 8 gènes d'effecteurs d'virulence (Avr) ont été clonés.

Les récentes études menées sur le système riz-*M. oryzae* ont permis de mettre en évidence la reconnaissance des effecteurs AVR1-CO39 et AVR-Pia par la paire de protéines NB-LRR RGA4 et RAG5 (Cesari et al., 2013) par une interaction directe entre les partenaires. Ces études ont été réalisées au BGPI (Biologie and Génétique des interactions Plante-Parasite, Campus International de Baillarguet, Montpellier). Les données concernant les caractéristiques de l'interaction et de ces NB-LRR seront abordées plus en détail dans le chapitre 2.

L'étude détaillée de la structure de ces effecteurs est un premier pas vers la compréhension de l'interaction entre l'Avr et son récepteur NLR. Dans cette optique, différents groupes se sont lancés dans l'étude de la structure d'effecteurs de *M. oryzae* (Cf tableau 1.2). Ainsi la première structure d'un effecteur de *M. oryzae* a été résolue pour AvrPiz-t et est parue en 2013 (Zhang et al., 2011). Les structures d'Avr-Pik et d'Avr-Pia publiée en 2015 respectivement par (Maqbool et al., 2015) et (Ose et al., 2015) seront abordées dans le chapitre Discussion et Perspectives de cette thèse.

Effecteurs d'Avirulence connus chez <i>Magnaporthe oryzae</i>			
<i>gène Avr</i>	protéine	<i>PDB (date)</i>	<i>Référence</i>
<i>Avr-Pita</i>	224 AA protéine sécrétée	-	-
<i>AvrPiz-t</i>	108 AA protéine sécrétée	2LW6 (2013)	(Zhang et al., 2013)
ACE1	Polyketide synthase	-	-
<i>AvrPii</i>	70 AA protéine sécrétée	-	-
<i>Avr-Pik</i>	113 AA protéine sécrétée, cinq allèles (A-E)	5A6W* (2015)	(Maqbool et al., 2015)
<i>Avr-Pikm</i>	113 AA protéine sécrétée, cinq allèles (A-E)	-	-
<i>Avr-Pikp</i>	113 AA protéine sécrétée, cinq allèles (A-E)	-	-
<i>Avr-Pia</i>	85 AA protéine sécrétée	2MYW (2015) 2N37 (2015)	(de Guillen et al., 2015) (Ose et al., 2015)
<i>Avr1-CO39</i>	89 AA protéine sécrétée	2MYV (2015)	(de Guillen et al., 2015)
PWL1	137 AA protéine sécrétée	-	-
PWL2	145 AA protéine sécrétée	-	-
PWL3	147 AA protéine sécrétée	-	-

Tableau 1.2 Tableau des effecteurs d'avirulence connus chez *M. oryzae* :

Abréviations: AA, acide aminé

*structure d'Avr-PikD

Adapté de (Liu et al., 2014; Rouxel and Balesdent, 2010)

1.2 Connaissances structurales récentes (AvrPiz-t)

Le gène *AvrPiz-t* code une petite protéine de 108 acides aminés, prédite comme étant sécrétée et ne présentant aucune homologie avec des protéines connues (Li et al., 2009). Récemment, il a été démontré que l'effecteur AvrPiz-t est impliqué dans la suppression de la réponse immunitaire médiée par les PAMP (PTI) (Park et al., 2012). L'expression d'AvrPiz-t dans des lignées de riz transgéniques induit la diminution de l'expression d'espèces oxydatives (ROS) déclenchée par la voie de reconnaissance des PAMP. La cible de virulence d'AvrPiz-t a été identifiée comme étant la protéine APIP6 (AvrPiz-t Interacting Protein 6). Cette protéine possède une activité ubiquitine ligase qui est inhibée en présence d'AvrPiz-t (Park et al., 2012). Ainsi, cet effecteur cible le système ubiquitine-protéasome pour supprimer la réponse déclenchée par les PAMP (PTI), augmentant ainsi la susceptibilité de la plante face au champignon.

La structure d'AvrPiz-t a été résolue par RMN à partir d'un échantillon de protéine recombinante uniformément marquée ^{15}N / ^{13}C (Zhang et al., 2013). Cette structure est caractérisée par deux

feuillet β antiparallèles disposés face à face et connectés par la boucle L2 (résidus 40-44) et la boucle L5 (résidus 77-81) (Fig 1.1).

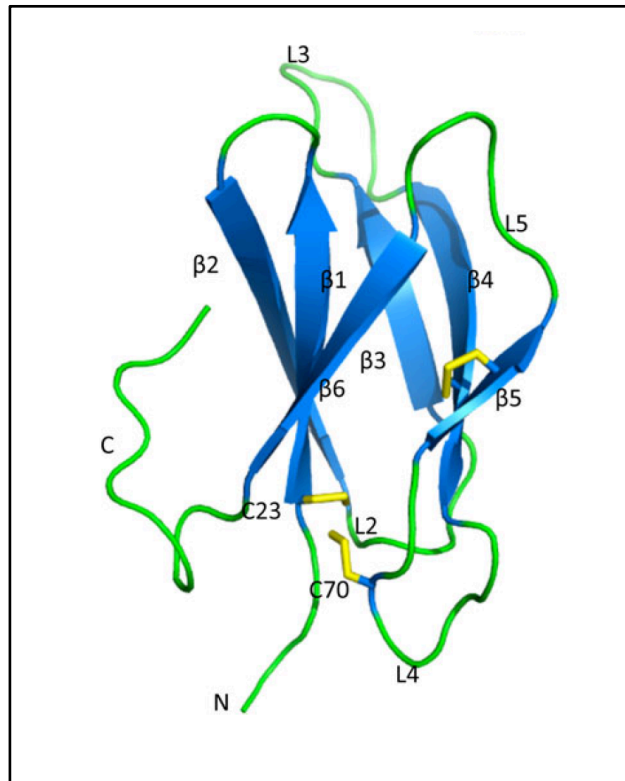


Fig 1.1 Structure d'AvrPiz-t (2LW6) :

Représentation de la structure RMN d'AvrPiz-t dans laquelle le N- et C-terminaux, ainsi que les éléments de structure secondaire sont indiqués; les brins β sont colorés en cyan, les boucles sont dessinées en vert. Les chaînes latérales des quatre résidus cystéines sont représentées en sticks jaunes. Adapté de (Zhang et al., 2013)

Chacun des deux feuillets est formé par trois brins β , le feuillet I est constitué des brins $\beta 1$ (résidus 23-29), $\beta 2$ (résidus 32-39) et $\beta 6$ (résidus 82-88); le feuillet II contient les brins $\beta 3$ (résidus 45-50), $\beta 4$ (résidus 57-64) et $\beta 5$ (résidus 73-76). Dans le feuillet II, un pont disulfure entre les résidus Cys62-Cys75 relie les brins $\beta 4$ et $\beta 5$. Deux autres cystéines Cys23 et Cys70 sont proches dans la structure mais ne forment pas de pont disulfure (Zhang et al., 2013).

La protéine ToxB (Nyarko et al., 2014), une toxine produite par *Pyrenophora tritici-repentis*, un champignon phytopathogène du blé possède une structure homologue à AvrPiz-t (Nyarko et al., 2014). ToxB est une toxine qui possède un homologue *toxb* ayant perdu l'activité toxique. L'analyse des structures de ces deux protéines (ToxB et *toxb*) suggère que l'absence de l'activité serait attribuée à une flexibilité plus élevée de *toxb*, identifiée par une étude de dynamique par RMN, ainsi que de nombreuses différences de résidus exposés au niveau des boucles (Nyarko et al., 2014).

AvrPiz-t et ToxB sont tous les deux des effecteurs de champignons contribuant à leur virulence lorsqu'ils sont produits (Nyarko et al., 2014; Park et al., 2012). D'après Nyarko et al., l'homologie des structures d'AvrPiz-t et de ToxB suggère qu'ils pourraient partager un mécanisme commun conduisant à la mort cellulaire. Ces auteurs proposent que certaines propriétés du récepteur immunitaire Piz-t, conduisant à la HR, seraient partagées avec la cible de virulence de ToxB.

2. Méthodes

La résonance magnétique nucléaire (RMN) liquide est une méthode de choix pour déterminer la structure tridimensionnelle de petites protéines à une résolution atomique. Les protéines possèdent des repliements caractéristiques en hélices α et feuillets β . L'avantage de cette technique est de pouvoir étudier les protéines en solution et d'avoir accès à la flexibilité de la molécule. La RMN est une technique basée sur l'analyse du retour à l'équilibre des noyaux de spins $\frac{1}{2}$ (^1H , ^{13}C et ^{15}N principalement) après perturbation par un champ de radiofréquence. Le retour à l'équilibre des noyaux dépend de leur environnement, ainsi il n'existe qu'une signature spectrale pour une molécule donnée. La fréquence de résonance des atomes est donc modulée de quelques parties par millions (ppm) par rapport à une fréquence de référence, cette variation est appelée déplacement chimique (Cf Fig 1.2). Une caractéristique importante des spectres RMN de protéine est leur extrême complexité, due au grand nombre de noyaux observables dans l'échantillon. Les différentes étapes de la détermination de structure de protéine par RMN sont : la préparation de l'échantillon, l'acquisition et le traitement du signal RMN produit par les noyaux observés, l'analyse et l'attribution des spectres, le calcul des coordonnées des atomes à partir des paramètres mesurés sur les spectres RMN.

Apolaire	RESIDU	Noyaux																				Carbone																		
		N	H	HA	HA2	HA3	HB	HB2	HB3	HG	HG1	HG12	HG13	HG2	HG3	HD1	HD2	ND2	HD21	HD22	HD3	NE	HE	NE1	HE1	NE2	HE2	HE21	HE22	HE3	HZ	HZ2	HZ3	HH	HH2	C	CA	CB	CD	CG
	GLY	108	8.39		3.99	3.88																														173	45			
	ALA	122	8.24	4.2			1.35																													178	53	18		
	VAL	118	8.45	4.06			2.03				0.91		0.8																						175	64	33			
	LEU	120	8.44	4.256				1.76	1.55	1.4						0.9	0.8																		177	56	41			
	ILE	123	8.19	4.12			1.85				1.43	1.09	0.9		0.7																				176	62.5	38			
	CYS*	118	8.31	4.7				3.8	2.8	HG																								175	57	28				
	MET	119	8.43	4.34				1.9	2.1				2.5	2.6									2											176	57	32.5				
	PHE*	119	8.23	4.55				3.1	2.9						7.2	7.2										7.15		7.15						175	59	39				
	TYR*	118	8.18	4.5				3	2.8						7.05	7.05										6.81		6.81						174.5	59	40				
	TRP*	120	8.09	4.65				3.2	3.1						7.25											129	10.16							176	58	29				
	PRO		4.4					2.24	1.9					2	1.95		3.6																	176	63	32				
	SER*	115	8.38	4.5				3.9	3.8	HG																								173	60	63				
	THR	114	8.23	4.42			4.12				3.15		1.15																					174	61	70				
	ASN*	117	8.75	4.71				2.85	2.75									112	7.54	6.89														174	54	28		177		
	GLN	118	8.4	4.17				2.08	1.97				2.24	2.24																				175	56	28	180			
	ASP*	119	8.42	4.5				2.68	2.6																									176	56	40				
	GLU	119	8.37	4.13				2.1	1.9				2.34	2.24																				177	56	29				
	HIS*	117	8.61	4.51				3.21	3.05						7.2	7											7.8	175		HE2				173	55	29				
	LYS	119	8.41	4.1				1.83	1.7				1.5	1.4		1.69																		176	57	32				
	ARG	119	8.27	4.23				1.9	1.7				1.61	1.57		3.2																		177	57	27				

Fig1.2 Déplacement chimique de référence pour les 20 acides aminés :

Ce tableau répertorie les différents déplacements chimiques de référence (CCPNMR) des atomes. Le nom des atomes suit la nomenclature IUPAC. Les acides aminés sont classés en fonction de leur polarité.

L'analyse de la structure d'une protéine par RMN nécessite un échantillon pur, concentré et uniformément marqué à l'azote 15 et/ou au carbone 13. Dans cette optique, les protéines

recombinantes AVR1-CO39 et AVR-Pia ont été produites dans *Escherichia coli*. Pour ce faire, les séquences codantes des protéines sont insérées dans un plasmide de type pET15b auquel un signal de sécrétion au périplasme a été ajouté en N-terminal. En effet, les protéines AVR1-CO39 et AVR-Pia étant sécrétées par le champignon, il nous a semblé judicieux d'insérer ce peptide signal, ce qui permet un meilleur rendement par rapport à une construction « classique ». Les techniques de production et de purification sont décrites dans la partie A de la section Matériels et Méthodes.

Les expériences RMN permettent l'observation des couplages scalaires (par le squelette de la molécule) et dipolaires (par l'espace) entre spins. L'observation du transfert d'aimantation, par couplage dipolaire, produit des effets Overhauser nucléaires (nOe) entre noyaux situés à moins de 5 ångström (Å) l'un de l'autre, et permet l'estimation des distances entre les noyaux observés. L'étude des protéines se fait par RMN hétéronucléaire pour résoudre les superpositions de déplacements chimiques au niveau du proton. La stratégie d'attribution débute par l'acquisition de spectres hétéronucléaires à 2 dimensions comme le spectre HSQC (« heteronuclear single quantum correlation»). L'HSQC permet d'observer les pics de corrélation dus au couplage scalaire entre l'hydrogène amide et l'azote de chaque résidu, elle représente l'empreinte de la protéine. L'enregistrement de spectre à 3 dimensions à partir d'échantillons doublement marqués (^{15}N , ^{13}C) permet l'attribution par l'enregistrement d'expériences HNCO, HN(CA)CO, HNCACB, HNCA ; HN(CO)CACB) (Cf Fig 1.3). Dans cette nomenclature, l'aimantation passe par le carbone entre parenthèse mais il n'est pas édité par une dimension spectrale.

L'HNCO permet de déterminer la position spectrale du carbonyle « CO » autour de 170 ppm porté par le résidu précédent. L' HN(CA)CO donne accès au CO du résidu (i) et du précédent (i-1). En superposant les deux expériences l'attribution des CO des résidus i et i-1 peut être faite.

L'HNCA permet de déterminer principalement la position spectrale du carbone alpha du résidu (i) et il reste une corrélation plus faible pour le résidu précédent (i-1), autour de 55 ppm. Dans l'HN(CO)CACB, les taches de corrélation observées correspondent au carbone alpha et au carbone bêta du résidus précédent. En superposant ces deux dernières expériences avec l'HNCACB les positions spectrales des carbones α et β des résidus i et i-1 sont déterminées. L'attribution des carbones α et β peut alors être faite.

Ainsi basée sur la séquence primaire de la protéine, les fréquences de résonance des noyaux du squelette carbonée peuvent être attribuées, à la main, de façon automatique (par un logiciel comme CCPNMR) ou en alliant les deux méthodes

L'utilisation d'expériences tridimensionnelles comme NOESY-HSQC et TOCSY-HSQC sont également très utiles pour compléter l'attribution. Sur les 20 acides aminés, 12 ont une signature particulière, les 8 autres relevant d'un système AMX (couplage à deux noyaux non équivalents) et donnent le même type de taches de corrélation (Cf Fig 1.4). Ainsi ces expériences permettent d'attribuer les fréquences de résonance des systèmes de spins et de remonter sur les chaînes latérales.

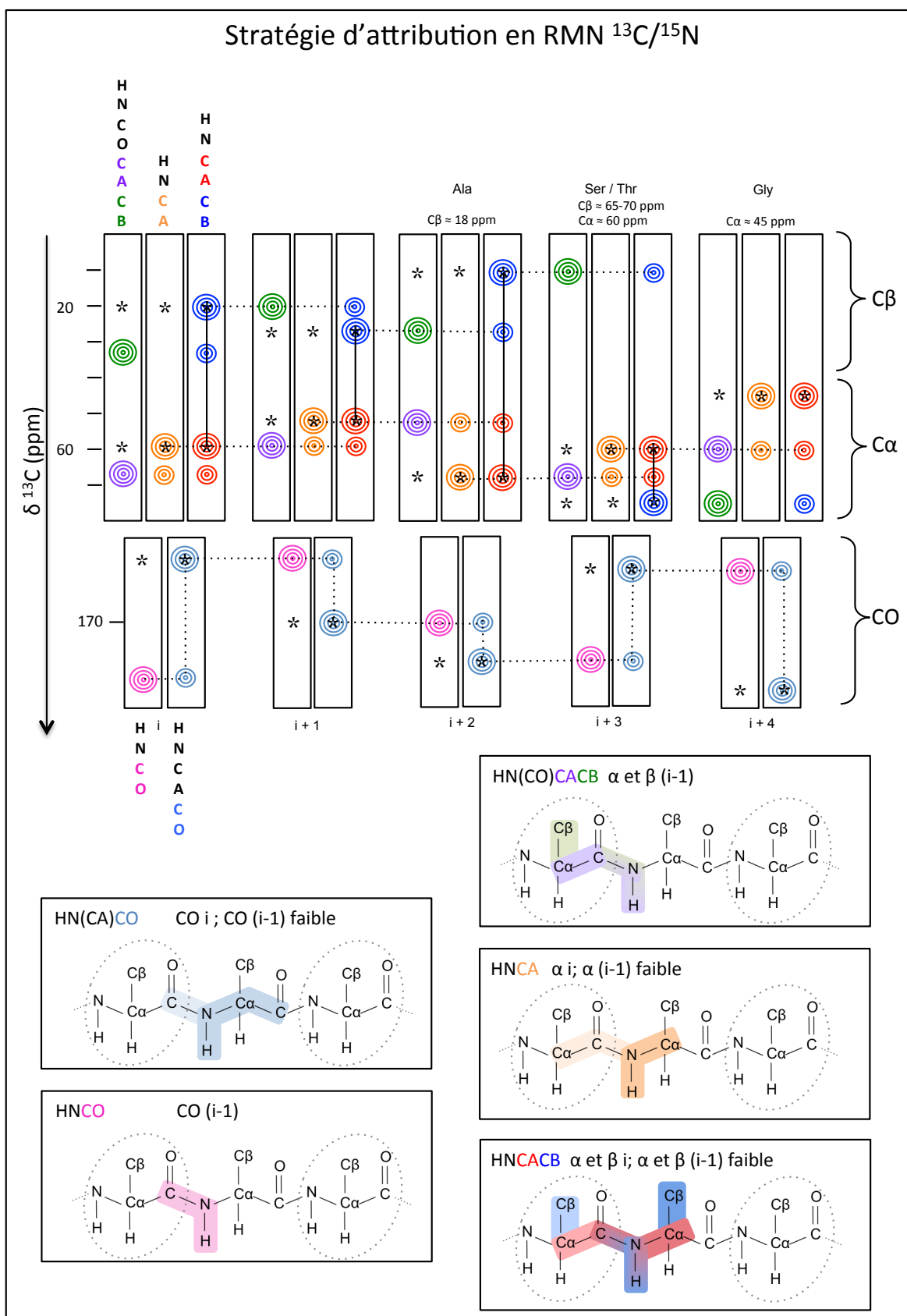


Fig 1.3 Stratégie d'attribution en RMN $^{13}\text{C}/^{15}\text{N}$:

Dans la première partie de la figure sont représentées les corrélations extraites sous forme de bandelettes (ou strip) à partir du pic HSQC en fonction des différentes expériences 3D. Le code couleur se rapporte à l'atome observé. Les étoiles correspondent aux $\text{C}\alpha$, $\text{C}\beta$ et CO du résidu. Dans la deuxième partie de la figure est représenté le transfert d'aimantation en fonction de l'expérience.

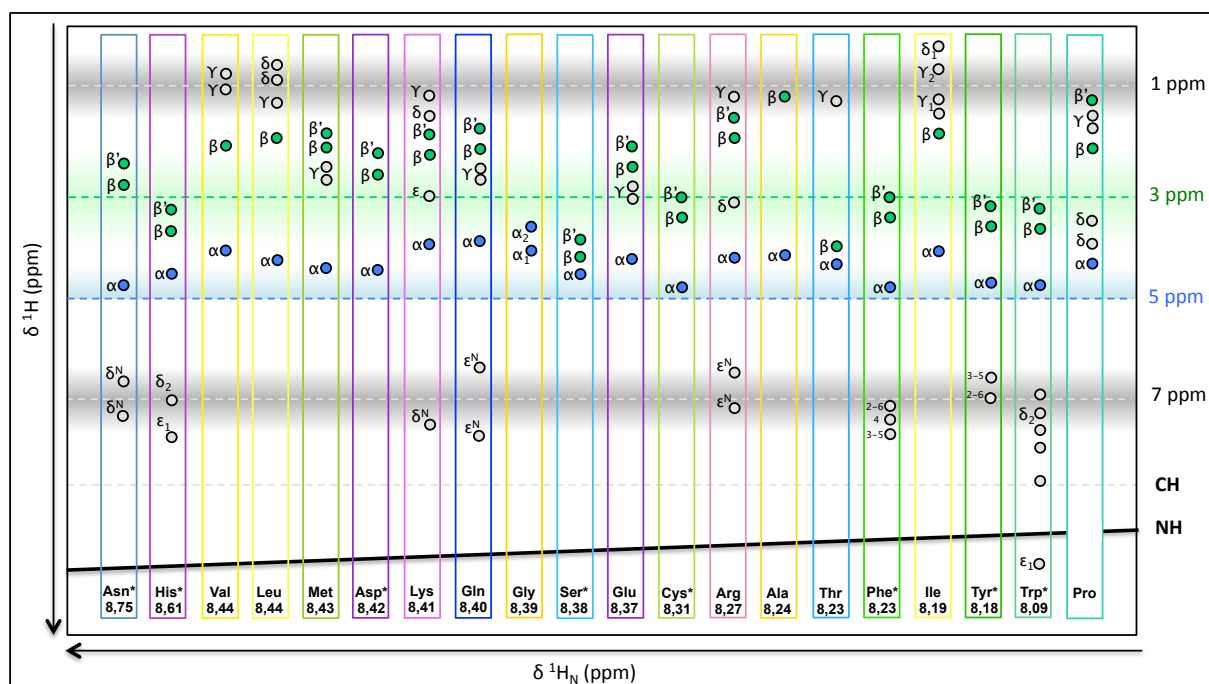


Fig 1.4 Système de SPIN :
Système de spins des 20 acides aminés, * Désigne les AMX

Une liste de déplacements chimiques, la plus complète possible, comprenant tous les noyaux observables de la protéine (^1H , ^{15}N et ^{13}C) est créée à partir de l'analyse des différentes expériences 3D.

L'utilisation du programme TALOS (Torsion Angle Likelihood Obtained from Shift end sequence similarity) permet de transformer la liste de déplacements chimiques attribués ($^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, ^{13}CO , ^{15}N et $^1\text{H}_\text{N}$) en angles de torsion prédits pour la chaîne principale (ϕ et ψ) de la protéine.

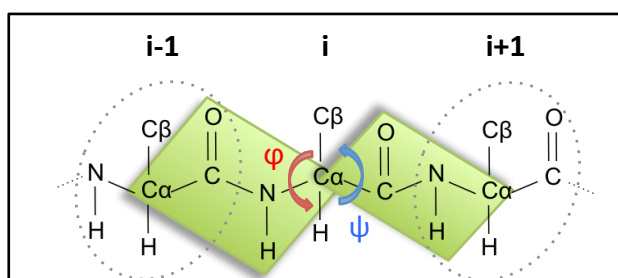


Fig 1.5 Représentation des angles de torsions ϕ et ψ sur une chaîne polypeptidique

L'analyse des expériences nOe, 2D-NOESY, et 3D-NOESY-HSQC qui éditent les corrélations nOe résultant de la proximité spatiale des différents protons en fonction du noyau ^{15}N , permet de créer deux listes de pic nOes avec les intensités.

L'attribution des corrélations présentes sur un spectre nOe est réalisée de façon automatique par le logiciel CYANA. Ce programme est basé sur une procédure itérative qui lève progressivement les

ambiguïtés en générant des modèles moléculaires intermédiaires. La prise de risque de l'attribution (contrainte de distance erronée dont l'impact sur la structure finale pourra être très variable) est gérée en réduisant progressivement des paramètres qui vont jouer sur la précision de l'attribution, et qui démarre par une étape de contraintes non-ambigües. La qualité et la convergence des structures retenues sont des critères importants pour estimer la qualité du processus de génération des structures. Le résultat final est donné par un tableau qui récapitule pour un ensemble de structures sélectionnées, la valeur de la fonction "cible" (qui est une mesure des violations de contraintes pour une structure), et une liste de violations de contraintes de distance et d'angles dièdres globale, qui permet d'apprécier si une contrainte particulière est systématiquement ou épisodiquement violée. Un tri dans les contraintes attribuées automatiquement doit être réalisé pour corriger les erreurs et doit permettre d'obtenir une meilleure convergence des structures (Root Mean Square Déviation faible) et ayant moins de contraintes violées. Ce processus un peu fastidieux consiste à appliquer le protocole de calcul de dynamique par les angles de torsion ancêtre de CYANA (DYANA). A chaque étape, les violations sont analysées, un contrôle spectral des violations de distance est réalisé pour déterminer les superpositions/ambiguïtés éventuelles, l'intensité des pics nOes et l'existence de diffusion de spin (notamment quand le nombre de contraintes dans un réseau de ^1H est important). Pour ce faire la superposition de NOESY théoriques (approximation $1/d^6$) et des NOESY expérimentaux est utilisée. La violation de contrainte d'un angle dièdre nécessite l'inspection des distances dans la structure et l'éventuel ajustement/suppression de certaines contraintes. Le contrôle de la pertinence des structures secondaires observées, de leur compatibilité avec le repliement global de la structure et de toutes les contraintes nOe est basé sur une double carte de contact, l'une calculée par les contraintes nOe et l'autre obtenue à partir d'une des structures obtenues. A la fin de la procédure, une famille de structures similaires est obtenue, respectant les contraintes topologiques propres au système (longueur de liaison, stéréochimie, angles dièdres) et compatibles avec les données expérimentales. L'optimisation de la structure d'une protéine sous les contraintes géométriques déduites de la RMN est basée sur l'utilisation d'un protocole de calcul de recuit simulé. A partir d'un ensemble de structures, un champ de force va être appliqué permettant l'exploration de différentes conformations par une simulation de dynamique moléculaire. Le but étant de trouver les coordonnées minimisant au maximum l'énergie potentielle du système. L'algorithme de recuit simulé consiste à faire une dynamique moléculaire dans un premier temps à haute température (phase exploratoire) puis dans un deuxième temps le système est refroidi. Dans notre cas, nous avons utilisé le logiciel CNS pour effectuer cette étape.

La validation des structures obtenues est réalisée par le logiciel PROCHECK qui donne une vision statistique du jeu de structures par rapport à ce qui est attendu de façon standard dans des structures présentes dans la PDB. Les premiers critères d'acceptation d'une structure sont que les contraintes de distances nOe ne doivent pas être violées à plus de 0,4 Å et que les contraintes d'angles dièdres ne doivent pas être violées à plus de 5° . Les angles dièdres (φ , ψ) du squelette peptidique doivent se trouver dans les régions du diagramme de Ramachandran. Le logiciel PROCHECK donne également accès à des analyses statistiques sur les angles chi1 et chi2 (angles de torsion de la chaîne latérale) ainsi que sur les structures secondaires.

Les structures ont été déposées dans la PDB (Protein Data Bank) sous le nom 2MYV pour AVR1-CO39 et 2MYW pour AVR-Pia.

Les protéines AVR1-CO39 et AVR-Pia ne présentent que peu d'homologie de séquence. Les alignements de séquence des deux protéines matures par les logiciels Clustal2 et T-Coffee (<http://www.ebi.ac.uk/Tools/msa/>) révèlent 24% et 18% d'identité de séquence, respectivement, avec l'introduction de nombreuses insertions (4 gaps de 3 à 5 résidus). Quelque soit le programme utilisé les 2 cystéines centrales (exclusion faite du résidu cystéine C-terminal de l'AVR1-CO39) ne sont jamais alignées. La détermination des structures de ces deux effecteurs m'a permis de réaliser un alignement des séquences par l'utilisation du logiciel DALI (Holm and Rosenstrom, 2010) qui a été basé sur la superposition des deux structures. De manière plus générale le serveur DALI (http://ekhidna.biocenter.helsinki.fi/dali_server/) permet de faire des recherches sur toutes les structures présentes dans la PDB à partir d'une structure-cible. Les résultats consistent en (i) une liste de structures homologues, ordonnées en fonction du Z-score, et (ii) les données de l'alignement qui sont l'alignement des séquences primaires, le taux d'identité de séquence, le nombre de résidus superposés dans les structures et le RMSD associé.





La recherche dans DALI a permis de confirmer que les protéines AVR1-CO39 et AVR-Pia partagent le même repliement et d'identifier deux autres protéines Avr-Piz-t et TOXB qui possèdent également la même structure. La superposition par DALI des quatre structures a permis de guider l'alignement de séquences fortement divergentes, ainsi que la position conservée de deux cystéines. L'outil de bioinformatique PSI-BLAST (PSI pour Position-Specific Iterated et BLAST pour Basic Local Alignment Search Tool) est l'un des programmes les plus populaires pour la recherche d'homologie (Jones and Swindells, 2002). Une clé du succès de PSI-BLAST est sa capacité à proposer des substitutions à chaque position de la séquence-cible, et en pratique ceci est réalisé en générant un profil (matrice de score position-spécifique) à partir des résultats de l'étape précédente de recherche d'homologie. L'étape initiale du processus itératif PSI-BLAST est une recherche BLAST simple. Les séquences homologues retenues sont réalignées par rapport à l'alignement structural d'AVR-Pia, AVR1-CO39, AvrPiz-T et ToxB. A partir de cet alignement, un profil HMM (Hidden Markov Model) a été construit et utilisé pour rechercher de façon itérative par le programme HMMERsearch (Finn et al., 2011) sur une base de donnée rassemblant les génomes de champignons ascomycètes.

Les résultats de ces différentes analyses sont présentés dans l'article qui suit.


3. Article 1 : Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi

RESEARCH ARTICLE

Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi

Karine de Guillen^{1,2}, Diana Ortiz-Vallejo^{3,4}, Jérôme Gracy^{1,2}, Elisabeth Fournier^{3,4}, Thomas Kroj^{3,4}^{†*}, André Padilla^{1,2}^{†*}

1 INSERM U1054, Centre de Biochimie Structurale, Montpellier, France, **2** CNRS UMR5048, Montpellier University, Montpellier, France, **3** INRA, BGPI, Biology and Genetics of Plant-Pathogen Interactions, Campus International de Baillarguet, Montpellier, France, **4** CIRAD, BGPI, Biology and Genetics of Plant-Pathogen Interactions, Campus International de Baillarguet, Montpellier, France

 These authors contributed equally to this work.

[†] TK and AP also contributed equally to this work.

* thomas.kroj@supagro.inra.fr (TK); andre.padilla@cbs.cnrs.fr (AP)



CrossMark
click for updates

 OPEN ACCESS

Citation: de Guillen K, Ortiz-Vallejo D, Gracy J, Fournier E, Kroj T, Padilla A (2015) Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. *PLoS Pathog* 11(10): e1005228. doi:10.1371/journal.ppat.1005228

Editor: Jin-Rong Xu, Purdue University, UNITED STATES

Received: June 18, 2015

Accepted: September 24, 2015

Published: October 27, 2015

Copyright: © 2015 de Guillen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data except sequences from resequenced *Magnaporthe* isolates are within the paper and its Supporting Information files. Sequences from resequenced *Magnaporthe* isolates are available at <http://genome.jouy.inra.fr/gemo>

Funding: This work was supported by a PhD grant for DO from the Colombian National Agency for the Science and Technology COLCIENCIAS, COLFUTURO (call 528) and by grants from the "Agence nationale de la recherche" of France ANR-10-INSB-05-0 (French Infrastructure for Integrated

Abstract

Phytopathogenic ascomycete fungi possess huge effector repertoires that are dominated by hundreds of sequence-unrelated small secreted proteins. The molecular function of these effectors and the evolutionary mechanisms that generate this tremendous number of singleton genes are largely unknown. To get a deeper understanding of fungal effectors, we determined by NMR spectroscopy the 3-dimensional structures of the *Magnaporthe oryzae* effectors AVR1-CO39 and AVR-Pia. Despite a lack of sequence similarity, both proteins have very similar 6 β -sandwich structures that are stabilized in both cases by a disulfide bridge between 2 conserved cysteins located in similar positions of the proteins. Structural similarity searches revealed that AvrPiz-t, another effector from *M. oryzae*, and ToxB, an effector of the wheat tan spot pathogen *Pyrenophora tritici-repentis* have the same structures suggesting the existence of a family of sequence-unrelated but structurally conserved fungal effectors that we named MAX-effectors (*Magnaporthe* **A**vrs and **T**ox**B** like). Structure-informed pattern searches strengthened this hypothesis by identifying MAX-effector candidates in a broad range of ascomycete phytopathogens. Strong expansion of the MAX-effector family was detected in *M. oryzae* and *M. grisea* where they seem to be particularly important since they account for 5–10% of the effector repertoire and 50% of the cloned avirulence effectors. Expression analysis indicated that the majority of *M. oryzae* MAX-effectors are expressed specifically during early infection suggesting important functions during biotrophic host colonization. We hypothesize that the scenario observed for MAX-effectors can serve as a paradigm for ascomycete effector diversity and that the enormous number of sequence-unrelated ascomycete effectors may in fact belong to a restricted set of structurally conserved effector families.

Structural Biology—FRISBI) to AP, KG and JG, ANR-09-GENM-029 (Analysis and comparison of genomes of the fungal pathogen *Magnaporthe oryzae*—GEMO) to EF and TK and ANR-07-GPLA-0007 (Molecular bases of disease and resistance in the Interaction of Rice and *Magnaporthe grisea*—IRMA) to TK. This work benefited from interactions promoted by COST Action FA 1208. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist

Author Summary

Fungal plant pathogens are of outstanding economic and ecological importance and cause destructive diseases on many cultivated and wild plants. Effector proteins that are secreted during infection to manipulate the host and to promote disease are a key element in fungal virulence. Phytopathogenic fungi possess huge effector repertoires that are dominated by hundreds of sequence-unrelated small secreted proteins. The molecular functions of this most important class of fungal effectors and the evolutionary mechanisms that generate this tremendous numbers of apparently unrelated proteins are largely unknown. By investigating the 3-dimensional structures of effectors from the rice blast fungus *M. oryzae*, we discovered an effector family comprising structurally conserved but sequence-unrelated effectors from *M. oryzae* and the phylogenetically distant wheat pathogen *Pyrenophora tritici-repentis* that we named MAX-effectors (*M. oryzae* Avrs and ToxB). Structure-informed searches of whole genome sequence databases suggest that MAX-effectors are present at low frequencies and with a patchy phylogenetic distribution in many ascomycete phytopathogens. They underwent strong lineage-specific expansion in fungi of the *Pyriculariae* family that contains *M. oryzae* where they seem particularly important during biotrophic plant colonization and account for 50% of the cloned Avr effectors and 5–10% of the effector repertoire. Based on our results on the MAX-effectors and the widely accepted concept that fungal effectors evolve according to a birth-and-death model we propose the hypothesis that the majority of the immense numbers of different ascomycete effectors could in fact belong to a limited set of structurally defined families whose members are phylogenetically related.

Introduction

Pathogenic microorganisms have to cope with the immune system of their host and therefore deploy measures to hide their presence, disturb host immunity or inactivate defense responses. In all these strategies, proteins secreted by the pathogen during infection and acting on host proteins and cellular processes play a key role [1–3]. These proteinaceous virulence factors named effectors act either extra-cellularly or inside host cells and can possess, depending on the microorganism, very different molecular features.

In fungal pathogens, the main class of effectors are small secreted proteins of less than 200 amino acids expressed specifically during infection and often rich in cysteins [4–6]. Genome sequencing and expression analysis identified hundreds of such effector candidates in individual plant pathogenic fungal species. Few of them, mainly those acting extra-cellularly, are widely distributed among phytopathogenic fungi and contain known motifs or domains, such as NLPs (necrosis and ethylene-inducing peptide 1 (Nep1)-like proteins), LysM domain-containing proteins or protease inhibitors [5,6]. The vast majority of the fungal effectors do not share sequence similarities with other proteins and do not contain conserved motifs. This is very different from the situation in other phytopathogens and in particular oomycetes, an important class of plant pathogens that have similar lifestyles and infection strategies and whose virulence relies also on large effector repertoires. In oomycete pathogens, large families of cytoplasmic effectors with hundreds of members in individual species are defined by the presence of the RXLR or the LFLAK host cell translocation motifs [7–9]. The effector domains of these RXLR and Crinkler (CRN) effectors that mediate virulence functions are highly diversified but contain, in the majority of cases, conserved motifs or domains that are shared between effectors from the same or other species allowing their classification in distinct

families. On the contrary, most fungal effectors are species-specific while few are lineage specific and occur in closely related species. In most phytopathogenic fungi, no large effector gene families were identified [5,6]. The majority of their effectors are singletons and a small proportion belongs to small paralogous groups of rarely more than 3 members. Effector repertoires dominated by gene families of large size counting more than 5 members were only detected in particular cases such as powdery mildew and rust fungi lineages [10–13]. Due to their high diversity and the lack of similarity with other proteins, the mode of action and the role in infection of fungal effectors have to be elucidated case by case and remain still largely unknown [5,6]. In addition, this tremendous diversity raises the question of the evolutionary trajectories of fungal effectors that do not show traces of common origins.

Rice blast disease caused by the ascomycete fungus *M. oryzae* is present in all rice growing areas and causes important harvest losses. Since rice is the main source of calories for half of the human population and since disease control strategies are frequently overcome by the pathogen due to its high genetic plasticity, blast is considered one of the most dangerous plant diseases threatening global food security and hampering attempts to increase rice yield in many parts of the world [14–16]. Due to its economic importance, the status of the host plant rice as a model plant and the ease of cultivation and genetic manipulation of *M. oryzae*, blast disease has become a model for the molecular and genetic investigation of fungal plant diseases [14]. In particular, molecular mechanisms of fungal disease development were studied intensively in *M. oryzae* uncovering important features of fungal virulence [17,18]. Key steps in infection by *M. oryzae* are (i) penetration into epidermal cells by the breakage of the leaf cuticle and epidermal cell walls by an appressorium, a specialized unicellular structure, (ii) biotrophic growth inside the first invaded host cells, followed by (iii) necrotrophic growth associated with active killing of host tissue and the development of disease symptoms and finally, (iv) clonal reproduction and sporulation.

Effectors and in particular cytoplasmic effectors are key elements in *M. oryzae* virulence and particularly important during the biotrophic phase of infection [6,19,20]. However, the function of individual effectors in the infection process has only been established for the LysM effector SLP1 that sequesters chitin fragments and thereby interferes with their recognition by the rice chitin receptor CEBiP, and AvrPiz-t that interferes with host immunity by inhibiting the E3 ubiquitin ligase APIP6 [21,22]. Mutant analysis aiming to demonstrate that individual effectors are important for virulence have often been unsuccessful, probably due to functional redundancy among effectors [23,24]. Approximately 700 of the 1300–1500 secreted proteins encoded in the *M. oryzae* genome are considered effector candidates according to their size of less than 200 amino acids and their lack of homology to proteins of known function [25,26]. Hundreds of them were found to be expressed during appressoria formation or infection [23,26–28].

Some effectors are recognized in certain plant accessions by immune receptors localized either at the plasma membrane or in the cytosol leading to the induction of strong defense responses and resistance to pathogen isolates possessing this effector [29]. The recognized effector is, in these cases, named an avirulence (Avr) protein. In *M. oryzae*, 8 different effectors acting as Avr proteins named PWL2, AVR-Pia, AVR1-CO39, AVR-Pii, AVR-Pik, AvrPiz-t, AVR-Pita and Avr-Pi9 have been cloned molecularly [26,30–35]. They are all translocated into host cells and do not show similarities to proteins of known function with the exception of AVR-Pita that shows homology to neutral zinc proteases [6]. For 7 of them, the matching rice immune receptors that are in all cases cytoplasmic nucleotide-binding and leucine-rich repeat domain proteins (NLRs) have been identified [36–41].

In the present study, the 3-dimensional structures of the *M. oryzae* effectors AVR-Pia and AVR1-CO39 were investigated to deepen our understanding of fungal effector function and diversity. NMR analysis revealed that the structures of both proteins consist of two anti-parallel β -sheets, each having three strands, and linked by one disulfide bond. Structural similarity

searches revealed that the *M. oryzae* effector AvrPiz-t and the effector ToxB from the wheat pathogen *Pyrenophora tritici-repens* have similar 6 β -sandwich structures with the same topology [42,43]. Comparisons of the structures of the four effectors that we named MAX-effectors revealed that they share a common architecture but no sequence consensus. Structure-informed and pattern-based searches identified large numbers of weakly homologous MAX-effector candidates in *M. oryzae* and *M. grisea*, and limited numbers or no homologs in other phytopathogenic ascomycete fungi. Expression profiling indicated that the majority of the *M. oryzae* MAX-effector candidates are expressed during early infection. MAX-effectors therefore seem to have undergone a lineage-specific expansion in the *Pyricularia* genus that may be driven by duplications and rapid adaptation to new functions involving important changes of surface properties but conservation of protein architecture. This evolutionary process has the potential to generate large families of structurally related proteins without sequence similarity and may serve as a paradigm for effector evolution and diversification in phytopathogenic ascomycete fungi.

Results

Protein expression

AVR1-CO39 and AVR-Pia proteins, deleted for their endogenous secretion signal, were expressed in *E. coli* with an N-terminally fused signal peptide for secretion in the bacterial periplasm that is cleaved upon secretion, an N-terminal His₆-tag for purification and a TEV1 cleavage site. Recombinant proteins were soluble and were purified to homogeneity from periplasmic protein extracts by Ni-agarose affinity and gel exclusion chromatography (S1 Fig). Both recombinant Avr proteins eluted as monomers from gel exclusion chromatography.

NMR analysis

Recombinant, ¹⁵N and ¹³C-labelled AVR1-CO39 and AVR-Pia proteins produced in ¹⁵N and ¹³C-labelled minimal medium were used for structure determination by two- and three-dimensional NMR experiments. Three-dimensional (3D) HNCO, HNCA, HN(CO)CACB, HN(CA)CO, HNCACB, 2D ¹³C-detected CON, CACO and 2D-COSY-DQF(D₂O) and TOCSY(D₂O) experiments were used for the backbone and aliphatic side chain resonance assignments. 3D ¹⁵N-edited NOESY-HSQC and 2D-NOESY(D₂O) spectra were collected to confirm the chemical shift assignments and generate distance restraints for structure calculations. (Fig 1 and S1 Table). The assigned ¹H,¹⁵N-HSQC spectra were well dispersed. Residues from the N-terminal tags are still resolved. All amino acids of AVR-Pia and almost all of AVR1-CO39 have {¹H-¹⁵N} NOE values above 0.8 indicating highly defined structures with low flexibility (S2 Fig). Only N-terminal tags, below residue number 22–23, and C-terminal sequences of AVR1-CO39 (amino acids 80–89) show increased flexibility. The strong $d_{\alpha\text{N}}(i, i+1)$ NOEs and weak $d_{\text{NN}}(i, i+1)$ NOEs are indicative of a β -structure and consistent with the six β -strands observed in AVR-Pia and AVR1-CO39 (S3 Fig). NHs in slow exchange were consistent with hydrogen bonding networks and were used to derive constraints for the structure calculations.

The ratios of R_2 to R_1 relaxation rates of AVR-Pia and AVR1-CO39 were consistent with a monomeric molecular size (AVR-Pia $\tau_c = 6.2 \pm 0.3$ ns and AVR1-CO39 $\tau_c = 5.7 \pm 0.4$ ns) and thus confirm that both Avrs form monomers in solution (S1 Table) [44].

AVR-Pia and AVR1-CO39 have similar β -sandwich structures

The solution structures of AVR-Pia and AVR1-CO39 were determined based on 1541 and 1286 NOE-derived distance restraints, 90 and 72 dihedral angle restraints and 20 and 15

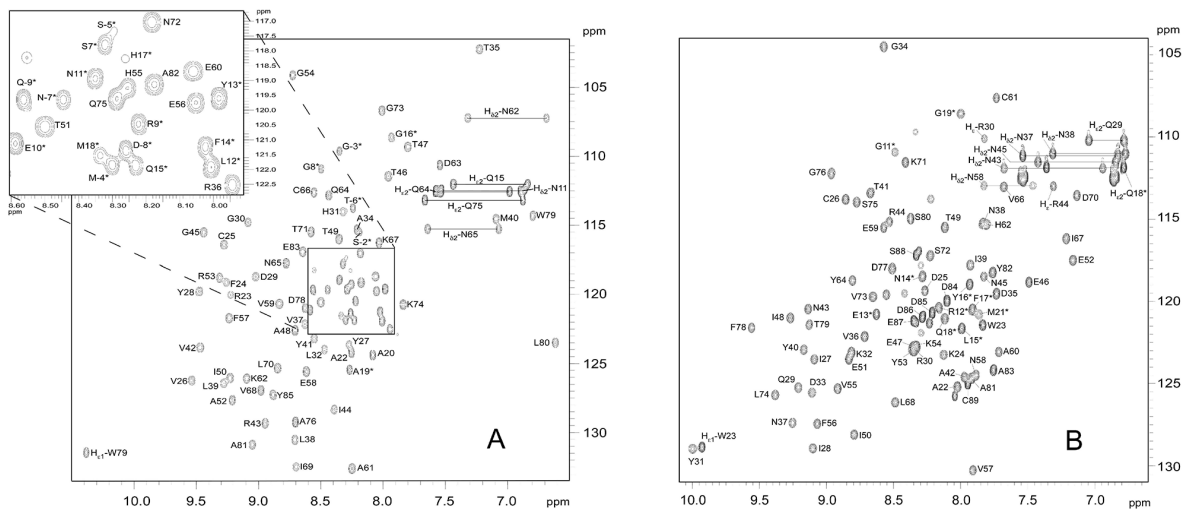


Fig 1. ¹⁵N-HSQC spectra of (A) AVR-Pia and (B) AVR1-CO39. In the ¹⁵N-HSQC spectra of (A) AVR-Pia and (B) AVR1-CO39, each peak comes from N-H chemical connectivity and has ¹⁵N and ¹H chemical shift coordinates. There is one backbone N-H group per residue, leading to one HSQC peak per residue. A side-chain NH₂ group gives two HSQC peaks with one common N coordinate. Other side-chains NH groups may also be observed, as Trp Nε1-H and Arg Nε-H. (*) indicates residues in the N-terminal tail. The mature proteins start at residue Ala20 and Trp23 for AVR-Pia and AVR1-CO39, respectively [26,32]. The NH₂ side chains resonances were assigned. The resonances of the tryptophan indole groups are specifically labelled Hε1.

doi:10.1371/journal.ppat.1005228.g001

hydrogen bond restraints, respectively (Fig 2 and Table 1 and S4 Fig). A disulfide bridge between Cys25-Cys66 for AVR-Pia and between Cys26-Cys61 for AVR1-CO39 was added based on cysteine ¹³Cβ chemical shifts and DTNB quantification of free thiols. The Pro65 in AVR1-CO39 has been determined to be in a cis-conformation according to the ¹³Cβ chemical shift at 34.4 ppm and strong sequential Hα-Hα NOE. The best conformers with the lowest energies, which exhibited no obvious NOE violations and no dihedral violations > 2° were selected for final analysis.

Surprisingly, both AVR-Pia and AVR1-CO39 proved to possess the same secondary structure elements arranged with the same topology in similar three-dimensional structures (Fig 2). Both proteins are composed of 6 β-strands that form two antiparallel β-sheets packed face-to-face and connected by loops (Fig 2). The first sheet is formed by the three β-strands β1, β2 and β6 while the second sheet contains β3, β4 and β5. In both cases, the two β-sheets pack together by an internal core of hydrophobic residues and one disulfide bridge and the structures belong to the β-sandwich classification. In both Avrs, the β-strands overlay and are similarly oriented (vide infra) but loops differ in length and structure.

AVR-Pia and AVR1-CO39 possess a hydrophobic surface patch

The surface properties of AVR-Pia and AVR1-CO39 are different with the exception of a hydrophobic patch located in both proteins on the side of the β-sandwich that is formed by the first β-sheet (β1-β2-β6) (Fig 2C and 2D). In AVR-Pia, this solvent exposed hydrophobic surface is constituted by the residues F24, V26 and Y28 in β1, V37, L38 and Y41 in β2, and Y85 in β6, and has an area of 372 Å². In AVR1-CO39, the solvent exposed hydrophobic surface of the first β-sheet is formed by the residues I27 and Y31 in β1, V36 and I39 in β2 and V73 in β6, as well as W23 from the N-terminus and Y82 from the C-terminus, and has a surface area of 280 Å².

ToxB and AvrPiz-t are structural homologs of AVR1-CO39 and AVR-Pia

To identify structural homologs of AVR-Pia and AVR1-CO39, structural similarity searches were performed using the Dali server and the Protein Data Bank [45]. Both queries, with AVR1-CO39 and AVR-Pia, identified the secreted effector protein ToxB from the wheat tan spot pathogen *Pyrenophora tritici-repentis* as well as its natural allele ToxB as the closest structural homologs with the highest Z-scores (S2 Table and Fig 3) [43]. Like, AVR-Pia and AVR1-CO39, ToxB is secreted during infection and is an important determinant of virulence for the tan spot fungus [46]. In addition, the search with AVR1-CO39 identified AvrPiz-t, another avirulence effector of *M. oryzae* that is sequence-unrelated to AVR-Pia and

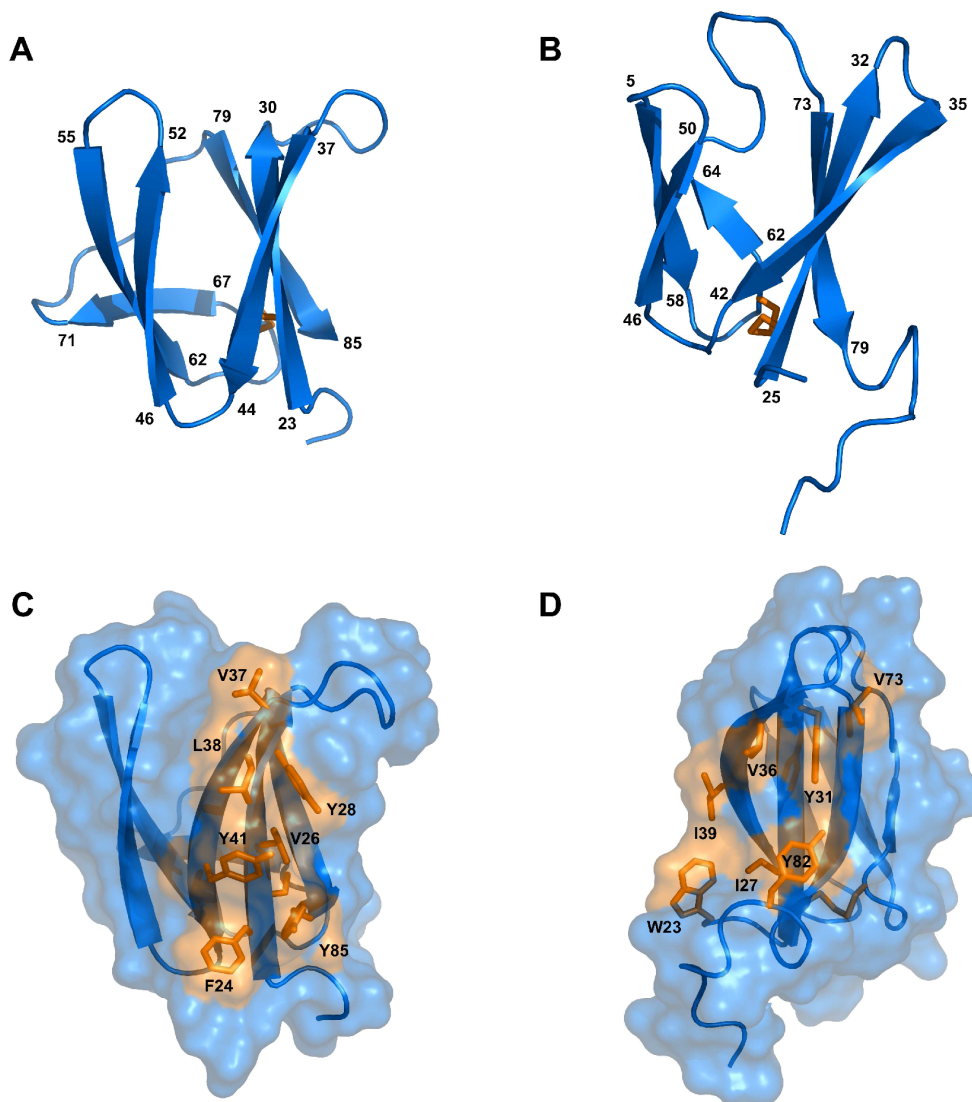


Fig 2. Solution structures of mature AVR-Pia and AVR1-CO39. Cartoon representations of AVR-Pia (A) and (B) AVR1-CO39 highlight the similar β -sandwich structure of both proteins. Yellow sticks represent disulfide bonds. Numbers indicate the residues at β -strands borders. A surface view reveals extended hydrophobic patches on one of the surfaces of AVR-Pia (C) and AVR1-CO39 (D) that are composed of exposed hydrophobic residues labelled in yellow. The Figs were generated using PyMOL (<http://www.pymol.org>).

doi:10.1371/journal.ppat.1005228.g002

AVR1-CO39 but structurally similar [42]. A pairwise similarity matrix using root-mean-square deviation (rmsd, measured in Å) and DALI Z-scores [45] was established revealing that all proteins are structurally related and that ToxB is closer to all other three structures than the others among them (S2 Table). ToxB and AvrPiz-t are like AVR-Pia and AVR1-CO39, composed of two three-stranded antiparallel β -sheets, β 1- β 2- β 6 and β 3- β 4- β 5, forming a six β -sandwich (Fig 3A–3D). Structure-based sequence alignments provided by DALI revealed, at a first glance, no obvious conservation, but also no clear consensus except buried hydrophobic residues alternating with exposed polar amino acids in the β -strands (Fig 3E). The β -strands β 1 and β 2 are very similar in length and position in all four proteins, while β 3, β 4 and β 6 display more variation. β 5 is the shortest and the most irregular strand. As expected for β -strands, buried and exposed residues alternate, with the exception of β 1 where residues have a tendency to be more buried. This is due to the packing of β 1 in between the β 2 and β 6 strands. The loops connecting the β -strands have variable length, and are the sites where most of the residue insertions occur. The disulfide bond between C2 and C43 (ToxB numbering) is well conserved but shifted “in phase” by two residues in AVR-Pia (Fig 3E).

Psi-Blast searches identify in *M. oryzae* and *M. grisea* multiple effector candidates with similarities to *Magnaporthe* AvrS and ToxB

The unexpected finding, that all three *M. oryzae* effectors that have been characterized for their structure so far and one effector from an only very distantly related fungal group are

Table 1. Statistics for 20 NMR structures of AVR-Pia and AVR1-CO39.

	AVR-Pia	AVR1-CO39
NOE restraints	1541	1286
Short range ($ i-j \leq 1$)	1022	745
Medium range ($1 < i-j < 5$)	128	157
Long range ($ i-j \geq 5$)	391	384
Dihedral restraints (a)	90	72
Number of NOE violations		
> 0.1 Å	9.30 ± 1.29	22.45 ± 2.76
> 0.2 Å	0.35 ± 0.47	0.60 ± 0.74
> 0.3 Å	0.05 ± 0.01	0
> 0.4 Å	0	0
Dihedral violations		
> 2°	0.15 ± 0.26	0.05 ± 0.10
> 4°	0	0
Ramachandran plot statistics (b)		
most favorable regions (%)	86.2	83.2
additionally allowed regions (%)	13.1	15.4
generously allowed regions (%)	0	1.3
disallowed regions (%)	0.7	0.3
Pairwise RMSD (Å) (c)		
Backbone	0.72 ± 0.14	0.49 ± 0.11
Heavy atoms	1.34 ± 0.18	1.02 ± 0.14

Structures were calculated using CYANA, refined using CNS, and analyzed using PROCHECK.

(a) Residues in regular secondary structures were derived from the chemical shifts using TALOS+ software.

(b) PROCHECK was used over the residues 24–85 for AVR-Pia and over the residues 23–83 for AVR1-CO39.

(c) Main chain atoms (N, C α , C) over the residues 24–85 for AVR-Pia and over the residues 23–83 for AVR1-CO39.

doi:10.1371/journal.ppat.1005228.t001

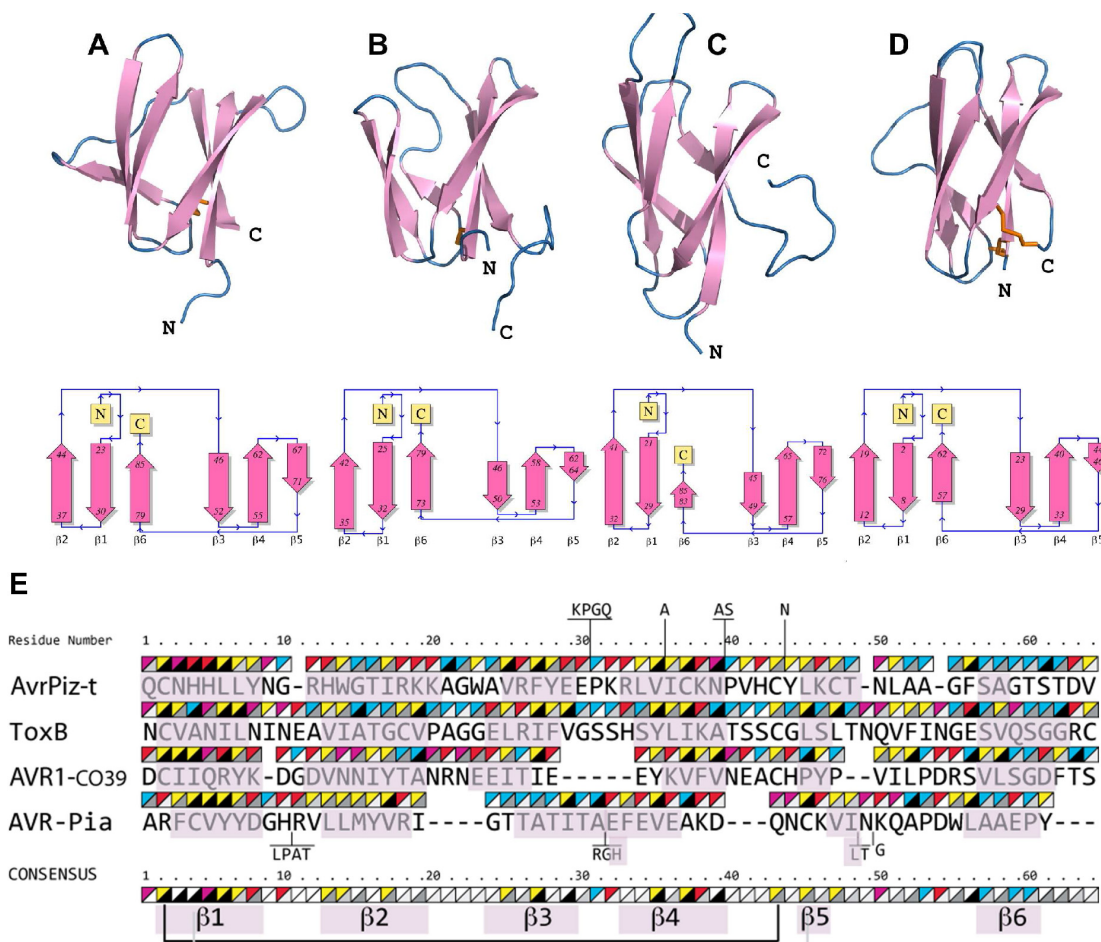


Fig 3. AVR-Pia, AVR1-CO39, AvrPiz-t and ToxB have similar 6 β -sandwich structures. Topology diagrams (lower row) show that AVR-Pia (A), AVR1-CO39 (B), AvrPiz-t (C) and ToxB (D) possess the same fold. Ribbon diagrams (upper row, generated with PyMOL (<http://www.pymol.org>)) highlight similarities of their structures. Disulfide bonds are shown in the ribbon diagrams by orange sticks. All four structures were superimposed and a structural alignment was derived using DALI with the ToxB sequence as the reference for numbering (E). Residues not aligned to ToxB are connected by vertical lines and correspond to insertions in loops of AvrPiz-t and AVR-Pia. Triangles over the residues indicate chemical properties (upper-left triangle: yellow for hydrophobic, red for charged, pink for Asn and Gln and blue for other residues) and solvent accessibility (lower-right triangle: from black for buried to white for solvent-exposed). The consensus is defined by at least three similar residues per position. Residues forming β -strands are pink. Disulfide bridges in AVR1-CO39 and ToxB are shown below the consensus by a black line and for AVR-Pia by a grey line. For AvrPiz-t, no disulfide bridge was reported despite presence of the two conserved cysteins [42].

doi:10.1371/journal.ppat.1005228.g003

structurally related raised the possibility that these four effectors are members of a widely distributed and abundant fungal effector family characterized by a common β -sandwich structure and high sequence divergence. Simple Blast searches are not suited to identify such distantly related proteins and when performed with the protein sequence of effectors from ascomycete fungi, generally identify no or only very few conserved homologs in the same species. Therefore, more sensitive Psi-Blast searches that use position-specific scoring matrices were performed with AVR-Pia, AVR1-CO39, AvrPiz-t and ToxB. The searches were performed on a protein sequence database combining the protein sequences of the *M. oryzae* reference isolate 70-15, of 5 other rice-infecting *M. oryzae* isolates (TH16, TH12, PH14, FR13 and Guy11), three *M. oryzae* isolates with other host specificities (BR32, US71 and CD156 specific for

wheat, *Setaria italica* and *Eleusine coracana*) and one isolate of the sister species *M. grisea* (BR29). These additional *M. oryzae* and *M. grisea* protein sequences were obtained by whole genome re-sequencing and de novo annotation of proteins and are accessible at <http://genome.jouy.inra.fr/gemo> [47]. After 4 Psi-Blast iterations and filtering of the results for sequences having an alignment length of at least 40 residues, an overall protein size of less than 180 amino acids and the presence of a predicted signal peptide, 3, 8 and 4 homologs of AVR-Pia, 16, 25 and 16 homologs of AVR1-CO39 and 5, 9 and 6 homologs of ToxB were detected in respectively 70–15, TH16 and BR29 (S3 Table, orthologous sequences present in 70–15 and TH16 were only counted for 70–15). For the other *M. oryzae* isolates similar numbers of homologs as in TH16 were found. The elevated number of homologs present in these isolates but not in 70–15 are due to the fact that the pipeline used for protein annotation in the re-sequenced genomes identified many additional small secreted proteins that are not annotated in 70–15 although the corresponding coding sequences are present in its genome [47]. The similarities were weak (frequently less than 25% identity) but they were consistent with the structural alignment (Fig 3) and included the two cysteine residues. For AvrPiz-t, no homologs that were not already identified by standard Blast were identified in the Psi-Blast search. When 25 additional fungal genomes, including the closely related fungi *M. poae* and *Gaeumannomyces graminis* were added to the database for the Psi-Blast searches, only very limited numbers of homologs (0, 1 or 2) with frequently low e-value scores were identified in other fungi. This suggested that effectors with similarity to *Magnaporthe* Avr and ToxB named in the following MAX-effectors that potentially also have 6 β -sandwich structure are present with low frequency in other fungal pathogens but were strongly amplified and diversified in *M. oryzae* and *M. grisea* that both belong to the genus *Pyricularia* in the *Pyriculariaceae* family [48].

HMM searches identify a huge MAX-effector family in *M. oryzae* and *M. grisea*

To exclude that the Psi-Blast search missed MAX-effectors in the additional fungal genomes due to biases in the search matrix or too low sensitivity and to deepen the search for this class of effectors in *M. oryzae* and *M. grisea* genomes, a hidden Markov model (HMM)-based profile search was performed. This type of profile search is among the most powerful procedure for detecting with high accuracy remote homologies between proteins.

As a first step, a high stringency Blast search with the three *M. oryzae* effectors and a Psi-Blast search with ToxB was performed and the resulting set of closely related sequences was aligned in a multiple sequence alignment constrained by the structural alignment of AVR-Pia, AVR1-CO39, AvrPiz-t and ToxB (S5A Fig). For the *M. oryzae* effectors, the Blast search identified orthologs of the effectors with few polymorphisms in different *M. oryzae* isolates. In addition, for each *M. oryzae* effector, one paralog was identified in *M. oryzae* and one or two paralogs were identified in the *M. grisea* isolate BR29 (S5B Fig). For the *M. oryzae* paralogs, generally several different alleles were identified. For ToxB, in addition to highly homologous sequences from *P. tritici-repentis* and *P. bromi*, 1 homolog was identified in *M. oryzae*, *Bipolaris oryzae* and *Colletotrichum higginsianum*, 2 in *C. fiorinae*, 3 in *C. orbiculare* and 4 in *C. gloeosporioides*. (S5B Fig).

As a second step, an HMM profile was built, starting from the structure-guided multiple sequence alignment from step1 (S5A Fig) and by iteratively searching for homologs in a database containing the small secreted proteins (<170 amino acids) of 25 pathogenic and non-pathogenic ascomycete fungi and of the 9 re-sequenced *M. oryzae* and *M. grisea* isolates from which completely redundant sequences had been removed. At each iteration, the recovered sequences were filtered for alignment of the two cysteins with a spacing of 34 to 49 amino acids

and used to generate a new profile used in the next iteration. The interval of 34 to 49 amino acids was fixed, based on the frequencies of cysteine spacings in HMM searches run without this constraint.

This search recovered 161 new, more distantly related sequences of which 154 were from *M. oryzae* or *M. grisea*, 5 from 3 different *Colletotrichum* species, 1 from *Leptosphaeria maculans* and 1 from *Mycosphaerella graminicola* (recently renamed *Zymoseptoria tritici*) (S6A Fig). This suggests that MAX-effectors have been massively and specifically expanded in *M. oryzae* and *M. grisea*. However, it also indicates their presence in other fungal species, i. e. in *Colletotrichum* spp. where they seem to occur at elevated frequencies. The alignment and clustering of the set of 200 sequences combining the 39 sequences used for the initial profile and the 161 new sequences revealed clusters of orthologous sequences originating from the different *M. oryzae* isolates with weak sequence polymorphism between orthologs (S6A and S6B Fig). Frequently, orthologs of *M. oryzae* can be identified in *M. grisea* but never in other fungi. Sequences from different orthologous clusters have high sequence diversity. Only in 3 cases, statistically significant clusters, supported by bootstrap values bigger than 50% can be identified that contain 2 distantly related MAX-effectors or MAX-effector clusters of *M. oryzae*.

A sequence logo derived from the multiple alignment shows the invariant cysteine residues (position 2 and 43 in mature ToxB) that constitute the alignment framework, as well as additional positions that are specifically enriched (Fig 4A). There is a propensity for hydrophobic residues in positions 4 and 6, corresponding to hydrophobic positions in strand β 1, in position 27, corresponding to a hydrophobic residue in β 3 and in positions 35, 37 and 39 corresponding to β 4. Positions 10, 23, 40 and 49 are in loop regions between the pairs of strands β 1- β 2, β 2- β 3, β 4- β 5 and β 5- β 6 respectively, and are enriched in glycine, polar or charged residues.

The resulting HMM profile was used to search with a relaxed cut-off two different databases: (i) the UniRef90 database that contains non-redundant sequences from a wide range of different organisms and that was used to determine in which type of organisms proteins with the MAX-effector motif occur and to evaluate by this the specificity of the motif and (ii) the previously described fungal genomes and *M. oryzae* and *M. grisea* database to get a precise view of the occurrence of MAX-effectors in a broad range of ascomycete fungi.

The search of the UniRef90 database recovered 70 sequences. All but 3 were from phytopathogenic ascomycete fungi (S7A Fig). The exceptions were from a bacteria, *Pseudomonas* sp. *StFLB209*, living in association with plants, from tomato (*Solanum lycopersicum*) and from a nematode-parasitic fungus (*Arthrobotrys oligospora*) and had low e-values. Among the fungal sequences, 49 were from *M. oryzae* and included AVR1-CO39 and AVR-Pia. The remaining 18 corresponded to previously identified effectors from *Colletotrichum* species (5 *C. orbiculare*, 2 *C. higginsianum*, 3 *C. gloeosporioides*, 2 *C. fioriniae*) that belong as *M. oryzae* to the class of Sordariomycetes and *Z. tritici*, *L. maculans* and *B. oryzae* as well as ToxB from *P. tritici-repentis* and *P. bromi* that are all from the class of Dothideomycete fungi. Clustering of the sequences revealed high sequence diversity and, apart from the Tox-B cluster, no or extremely limited relatedness could be identified (S7A and S7B Fig). Interestingly, with slightly different settings, this search also recovered the well characterized AVR-Pik effector from *M. oryzae* [26]. AVR-Pik clearly fits the MAX-effector pattern but was discarded in the other searches since its secretion signal is not recognized by the SignalP4.1 program used for filtering of the results.

The search of the previously described *Magnaporthe* and other fungal genomes database not filtered for redundancy recovered only limited numbers of MAX-effectors in non-*Magnaporthe* fungal genomes that had, with the exception of one effector from *Fusarium fujicuroi*, already been retrieved in the other searches (Fig 4B and S8A Fig). In *M. oryzae*, between 67 and 38 MAX-effectors per isolate were identified while in *M. grisea*, 37 MAX-effectors were identified (Fig 4B). 46 of the 55 MAX-effectors identified by Psi-Blast in *M. oryzae* 70–15 and

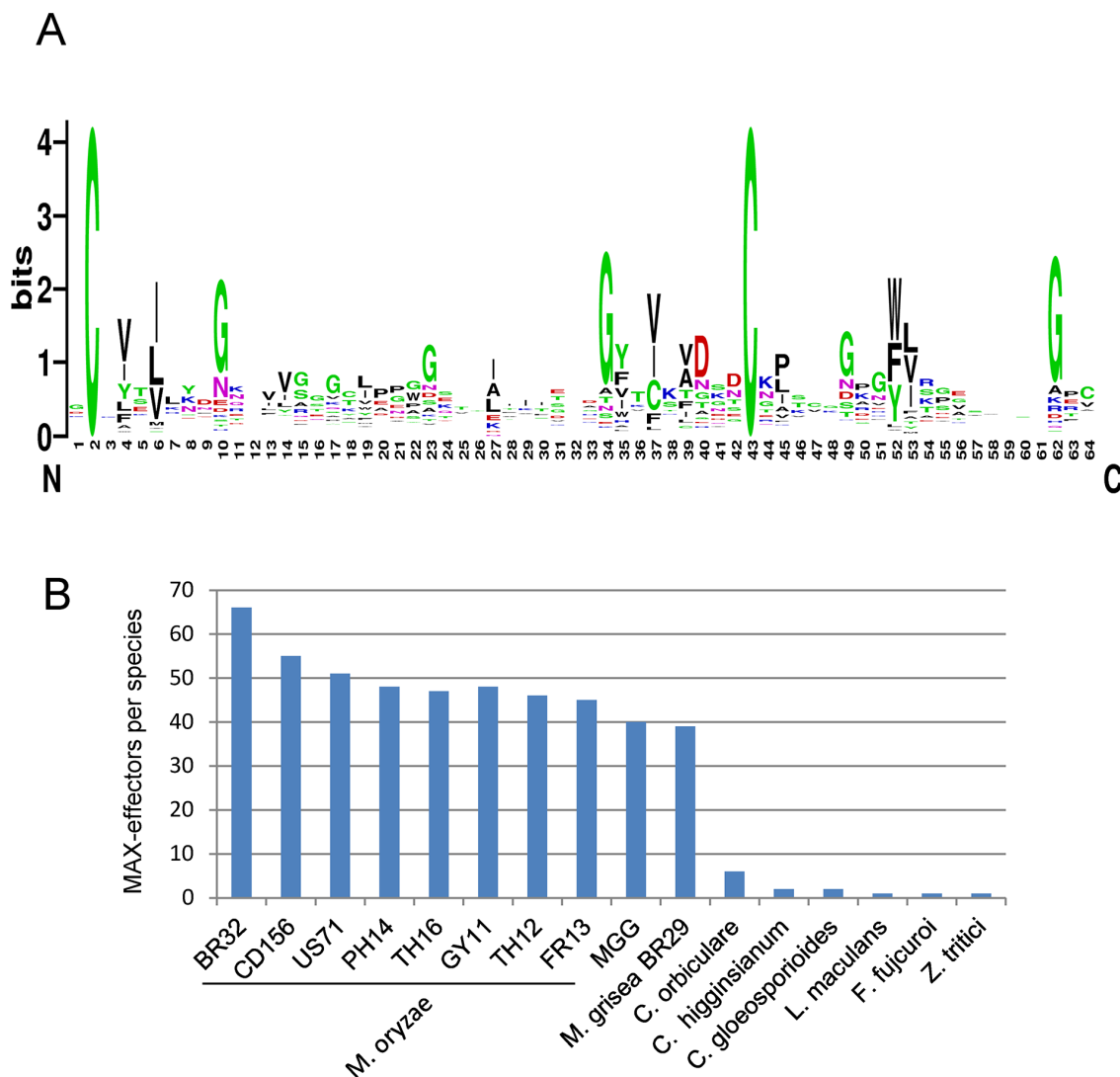


Fig 4. Large numbers of MAX-effectors sharing a characteristic sequence pattern are present in *M. oryzae* and *M. grisea*. A) Sequence pattern of MAX-effectors. The sequence logo was generated using the alignment of MAX-effector candidates identified by a high stringency HMM search (S6 Fig). (B) Numbers of MAX-effector candidates detected by a low stringency HMM sequence pattern search. A database combining 25 pathogenic and non-pathogenic ascomycete fungi and 9 *M. oryzae* and *M. grisea* isolates was searched with an HMM pattern based on a structural alignment of AVR-Pia, AVR1-CO39, AVR-Pia and AvrPiz-t.

doi:10.1371/journal.ppat.1005228.g004

TH16 and in *M. grisea* BR29 (S3 Table) were also found by this HMM search. Alignment and clustering shows that the *M. oryzae* MAX-effectors are generally present in the majority of *M. oryzae* isolates and are grouped in clusters of orthologs (S8A and S8B Fig). Many of these orthologous clusters also contain an ortholog from the *M. grisea* isolate BR29 that shows however higher sequence divergence. Only six statistically significant clusters (bootstrap > 50%) that contain more distantly related *M. oryzae* effectors from different orthologous groups are identified. Otherwise, the sequence diversity between proteins from different *M. oryzae* ortholog clusters is so strong that classical tree building methods do not detect statistically significant sequence relatedness. The non-*Magnaporthe* MAX-effectors do not cluster significantly with

Magnaporthe MAX-effectors and 8 of the 10 *Colletotrichum* effectors are comprised in three different *Colletotrichum*-specific clusters.

Taken together, the different HMM searches reveal that the MAX-effector motif is specific for effectors from phytopathogenic ascomycete fungi. MAX-effectors are identified with low frequencies in phytopathogenic ascomycete fungi from the class of Dothideomycetes and seem to have expanded moderately in different *Colletotrichum* species (i.e. *Colletotrichum orbiculare*). Only in *M. oryzae* and *M. grisea*, MAX-effectors expanded and diversified massively to become a dominating family of virulence effectors in these pathogens.

Expression profiling shows that a majority of MAX-effectors is expressed specifically during biotrophic infection

To test if the *M. oryzae* MAX-effectors identified by the HMM profile search could be involved in plant infection, the expression of 50 different candidate MAX-effector-coding genes was analyzed by qRT-PCR in infected rice leaves and in *in vitro* grown mycelium (S4 Table). 30 genes showed early infection-specific expression with a majority of profiles (25) that strongly resemble the biotrophy effector marker gene *BAS3* (Fig 5 and S9A and S9D Fig) [23]. The expression pattern of all these genes and of 3 genes coding for MAX-effectors identified only by Psi-Blast searches was clearly different from the markers of very early or late infection (Orf3 and MGG01147, respectively). For 18 genes, no significant expression was detected and only 2 genes were expressed constitutively with significant expression in the mycelium (Fig 5 and S9C and S9D Fig). Therefore, the majority of the MAX-effector candidates seems specifically expressed during biotrophic infection and can therefore be considered as potential virulence effectors.

Discussion

In this study, we have determined by NMR spectroscopy the 3-dimensional structures of two different effectors of *M. oryzae*, AVR1-CO39 and AVR-Pia. Although the two proteins have no evident sequence similarity they possess similar 6 β -sandwich structures formed in both cases by two β -sheets each formed of three β -strands oriented in an antiparallel manner. Interestingly, similar β -sandwich structures have previously been found for AvrPiz-t, the only other structurally characterized *M. oryzae* effector and for ToxB, an effector from an only very distantly related plant pathogenic ascomycete fungus, *P. tritici-repentis*. Overlay of the structures and structural alignments revealed that the nature and number of secondary structural elements and the topology of their fold are the same in all four effectors. In addition, all four proteins are stabilized by buried hydrophobic residues coming for their majority from the β -strands and by a disulfide bond between conserved cysteins located in the beginning of β 1 and in the beginning or just before β 5. However, the orientation and the length of certain β -strands, i.e. β -5, vary considerably and the sequences and the length of loops are highly variable resulting in proteins with very different shapes and surface properties. Due to the high sequence diversity, similarity among the MAX-effectors is therefore only detected when their structure is taken into consideration.

Hydrophobic surface patches in AVR-Pia and AVR1-CO39 are potential sites of protein-protein interaction

The only similarity of the surfaces of AVR1-CO39 and AVR-Pia is an extended hydrophobic area on the surface formed by β 1, β 2 and β 6. Such extended and exposed hydrophobic areas are uncommon since protein surfaces are generally in contact with solvent water molecules

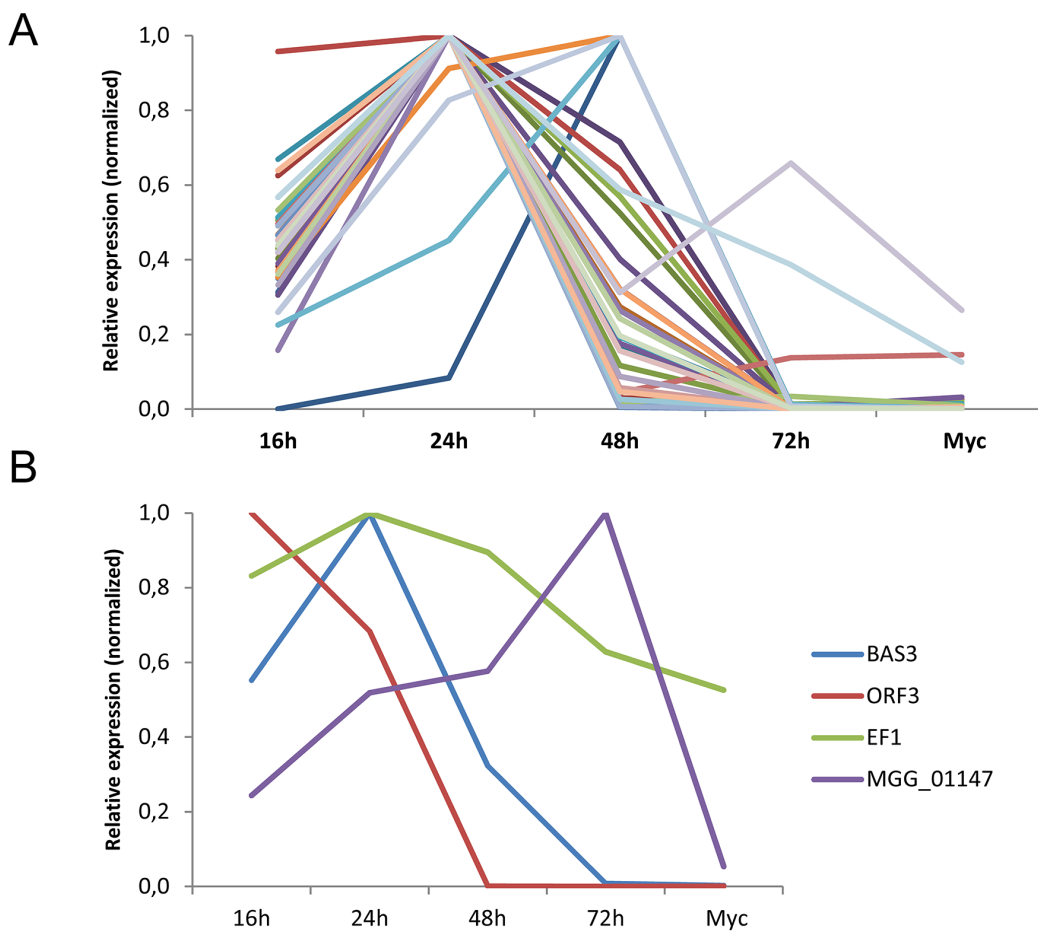


Fig 5. The majority of *M. oryzae* MAX-effectors is expressed specifically during biotrophic infection. mRNA levels of *M. oryzae* genes coding for 32 different MAX-effectors (A) and marker genes (B) for appressorium formation and very early infection (*ORF3* of the *ACE1* cluster, *MGG_08381*), biotrophic infection (*BAS3*, *MGG_11610*), late infection (*MGG_01147*) and constitutive expression (*EF1 α* , *MGG_03641*) were determined by q-RT-PCR in rice leaf samples harvested 16, 24, 48 and 72 h after inoculation and mycelium grown *in vitro*. Relative expression levels were calculated by using expression of a constitutively expressed *Actin* gene (*MGG_03982*) as a reference and normalized with respect to the highest expression value. Values are means calculated from the relative expression values of three independent biological samples. Individual expression profiles are in [S9 Fig](#).

doi:10.1371/journal.ppat.1005228.g005

and they are frequently involved in protein-protein interactions. Previous studies on the recognition of AVR-Pia by the rice NLR immune receptor RGA5 support that the hydrophobic surface of AVR-Pia could indeed be involved in protein binding [37]. AVR-Pia binds physically to a C-terminal domain of RGA5 homologous to heavy metal-associated (HMA) domain proteins related to the copper chaperone ATX1 from *Saccharomyces cerevisiae* (RATX1 domain). This binding is required to derepress a second NLR RGA4 that activates resistance signaling [49]. A natural allele of AVR-Pia (AVR-Pi-H3) where the surface exposed phenylalanine 24 and threonine 46 situated respectively in and at the border of the hydrophobic patch are replaced by serine and asparagine loses binding to RGA5_{RATX1} and does not trigger resistance [37]. Structural information will now guide further functional studies to elucidate if other amino acids situated in or at the border of the hydrophobic patch are also involved in RGA5_{RATX1}-binding and to validate by this the role of the hydrophobic patch as a protein-protein interaction surface.

MAX-effectors have different molecular properties and activities

Common features of the *M. oryzae* MAX-effectors are that they act intracellular in host cells [21,24,32] and are recognized by NLR immune receptors in resistant rice genotypes: AVR1-CO39 and AVR-Pia by the same NLR pair RGA4/RGA5 and AvrPiz-t by the NLR immune receptor Piz-t [37,39,41]. While the molecular bases of the recognition of AVR1-CO39 and AVR-Pia by RGA4/RGA5 are beginning to be elucidated, details of AvrPiz-t recognition are not known. Also, whether the three *M. oryzae* MAX-effectors target similar host processes and host proteins is not known. AvrPiz-t was described to target the host ubiquitin proteasome system by binding and inactivating the RING E3 ubiquitin ligase APIP6 [21] but virulence targets of AVR-Pia and AVR1-CO39 have not been described. However, it has been hypothesized that both proteins target RATX1 proteins homologous to the RGA5_{RATX1} domain that was suggested to act as a mimic for AVR-Pia and AVR1-CO39 targets [50]. Therefore, we assume that AvrPiz-t on the one hand and AVR-Pia and AVR1-CO39 on the other have different molecular activities and target different host proteins. This would be in accordance with the high divergence of their shapes and their surface properties. That AVR-Pia and AVR1-CO39 interact with the same immune receptor by binding to the same sensor domain and potentially interact with the same host targets is striking because apart from the extended hydrophobic patch on the $\beta 1\beta 2\beta 6$ surface they share no apparent similarities with respect to their shapes and surfaces. It will therefore be important to elucidate in the future which amino acids of AVR-Pia and AVR1-CO39 bind to RGA5_{RATX1} and which surfaces of RGA5_{RATX1} are involved in binding to each of the two effectors to better understand specificity in effector recognition. In addition, identification of AVR1-CO39 and AVR-Pia targets as well as ToxB targets for which molecular details of activity are also lacking will be important to understand how MAX-effectors promote virulence and to understand the link between MAX-effector structure and function.

MAX-effectors are a highly diversified effector family specific to phytopathogenic ascomycete fungi and underwent expansion in *M. oryzae* and *M. grisea*

Structure-informed pattern searches identified huge numbers of MAX-effector candidates that possess as the structurally characterized MAX-effectors very high sequence diversity and probably also possess a 6 β -sandwich structure stabilized by buried hydrophobic residues from β -strands and a disulfide bond between conserved cysteins connecting $\beta 1$ and $\beta 5$. Systematic prediction of the secondary structure of the MAX effector candidates using SS-PRO 5 software identified with high frequency two β -strands, $\beta 1$ located after the first cysteine and $\beta 4$ located before the second cysteine (S10 Fig). The other regions of the sequences had more variable secondary structure predictions which is also reflected by a less defined pattern in these regions (Fig 4A). High sequence diversity among MAX-effector candidates could as in the case of the structurally characterized MAX-effector be the consequence of interchangeability of buried hydrophobic core residues, variation in the lengths of some β -strands (i.e. $\beta 5$), exchange of surface exposed residues and deletion or insertion of residues in exposed loops.

MAX-effectors were specifically detected in phytopathogenic ascomycetes from the classes of Sordariomycetes and Dothideomycetes. One MAX-effector per species was detected in phytopathogenic fungi of the class of Dothideomycetes (*L. maculans*, *P. tritici-repentis*, *Z. tritici* and *B. oryzae*) and higher numbers (2–6) occur in fungi from the genus *Colletotrichum*. Only in *M. oryzae* and *M. grisea* that are both from the genus *Pyricularia* huge numbers of MAX-effector candidates were detected and expression profiling confirmed that most of them are likely *bona fide* effectors expressed specifically during biotrophic early infection. With 40–60

effectors which represents 5–10% of the candidate effectors of individual *M. oryzae* or *M. grisea* isolates, MAX-effectors can be considered a dominant class of effectors in these fungi [24,47]. This is further supported by the finding that 5 of the 51 biotrophy-associated proteins identified by transcriptome analysis are MAX effectors (MG02546, MG08414, MG08482, MG09425 and MG09675) [23]. Also, the *M. oryzae* effector AVR-Pik fits the MAX-effector pattern further highlighting the outstanding importance of this effector family that comprises 4 out of 8 cloned Avr effectors in the blast fungus [6]. It is striking that the only other group of fungi with elevated numbers of MAX effectors are *Colletotrichum* species. *Colletotrichum* fungi are phylogenetically only distantly related to *M. oryzae* and *M. grisea* but employ a similar hemibiotrophic infection strategy characterized by appressorium-mediated penetration into the host and growth inside invaded plant cells during biotrophic infection. It will be interesting to determine in the future whether MAX effectors play similar roles in these early infection processes in both groups of fungi.

In *Gaeumannomyces graminis* and *M. poae* that belong to the closely related *Magnaporthaceae* family no MAX-effectors were detected [48]. The expansion of MAX-effectors therefore occurred probably in a common ancestor of *M. oryzae* and *M. grisea* since clear orthologous relations can be established between many MAX-effectors from *M. oryzae* and *M. grisea* but after the split of the *Magnaporthaceae*. Expansion and diversification of the MAX-effectors is clearly continuing since frequently orthologs in *M. oryzae* or *M. grisea* cannot be identified and duplication, loss and diversification of MAX-effectors in host specific lineages of *M. oryzae* is observed (S8B Fig). Genome sequencing of additional species from *Pyricularia* and other genera in the *Periculariae* will allow to further strengthen the hypothesis of lineage-specific expansion of MAX-effectors.

Lineage specific expansion of effector families has been observed in other fungi such as mildew and rust fungi whose effector repertoires are dominated by effector families that contain frequently numerous members and are for their majority restricted to individual species or precise clades [10,51]. However, in these cases, sequence divergence is not as strong as in MAX-effectors since sequence-based comparisons allow the establishment to these effector families.

On the contrary, the effector repertoires of ascomycete phytopathogens outside the mildew lineage contain hundreds of sequence-unrelated effectors and the evolutionary origin of these huge amounts of species or clade specific genes is an open question. Duplication and diversification eventually driven by localization of the genes in transposon rich regions, genome reshuffling or transfer of accessory chromosomes were convincingly proposed as potential mechanisms to create effector diversity but the apparent lack of relatedness of ascomycete effectors remains unexplained [52–55]. Establishment of a huge effector family in *M. oryzae* and *M. grisea* that is also present at much lower frequency in other ascomycete pathogens sheds new light on the origin and relatedness of ascomycete effectors.

Diversifying rather than convergent evolution leads to highly diversified effector families

Theoretically, convergent evolution as well as diversifying evolution can explain the situation observed for the MAX-effectors characterized by a broad and patchy distribution, high diversification and limited sequence homology as well as a shared sequence pattern and probably the same structure. Convergent evolution would apply if these proteins with similar functions and a similar fold appeared repeatedly in phytopathogenic ascomycetes and eventually evolved independently in different clades. Under diversifying evolution, a protein or protein family present in a common ancestor has been strongly diversified in different lineages of ascomycete fungi and frequently lost during evolution in certain lineages and species. The scenario of

convergent evolution of MAX-effectors cannot be excluded but is clearly less parsimonious. It raises the question why MAX-effectors do not occur in organisms with similar lifestyles outside the Sordariomycete and Dothideomycete pathogens such as phytopathogenic basidiomycetes or oomycetes. In addition, there are no well-documented examples of convergent evolution towards similar folds or sequence patterns for pathogenic effectors or secreted fungal proteins involved in adaptation to the environment while comparative genomics studies in fungi and oomycetes are beginning to identify such widely distributed gene families that are shaped by strong diversifying selection and that can only be properly reconstructed when pattern-based searches and structure information are taken into consideration. The best documented example is certainly the WY-domain family among the RXLR effectors that is specific to the *Pero- nosporales* clade in oomycetes and evolves by diversifying evolution [8,9,56–58]. Careful sequence analysis involving pattern searches identified the W, Y and L sequence motifs in the effector domains of a majority of the *Phytophthora* RXLR effectors that are frequently completely sequence unrelated [9]. Functional analysis confirmed the importance of these motifs for effector function [59] and structure analysis of the effector domain of different RXLR effectors with limited sequence homology revealed that conserved sequence motifs reflected a conserved, highly similar 3-dimensional structure named the WY-domain fold [56,60–62]. PexRD2 and AVR3a11 show e.g. only 14% amino acid identity in a structure-based alignment but overlay of their structures has an RMSD score of 0.73 Å. As in the case of the β -sandwich fold of the MAX-effectors, the WY-domain fold tolerates insertion or deletion of amino acids in the loops, exchange of surface exposed amino acids and is stabilized by hydrophobic core residues that can be exchanged as long as hydrophobicity is maintained [56]. This flexible structure allows to generate effectors with highly variable shapes and surface properties and studied WY-domain effectors showed very diverse molecular activities, target different host proteins and are recognized by different NLR immune receptors [7,56].

An example of rapidly evolving proteins from fungi that are structurally but not sequence-conserved are hydrophobins that are low molecular mass secreted proteins important for the impermeabilization of fungal cell walls, adhesion to hydrophobic surfaces and pathogenicity [63]. Hydrophobins were shown to evolve rapidly according to a birth-and-death mechanism [64], are widely distributed in a broad range of basidio- and ascomycete fungi and are characterized by sequence patterns but no sequence homology [63,65]. Structure analysis demonstrated that distantly related hydrophobins are structurally related supporting a common evolutionary origin [66].

Another example of a fungal gene family that is rapidly evolving according to a birth-and-death model are the Hce2 proteins (homologs of *Cladosporium fulvum* ECP2) that are present in a wide range of basidio and ascomycete fungi and seem to act as effectors in pathogenic fungi and potentially in stress responses in non-pathogenic fungi [67]. Much like MAX-effectors they show patchy distribution, lineage-specific expansions and high sequence diversification.

MAX-effectors may serve as a paradigm for the evolution and diversification of effectors in phytopathogenic ascomycetes

Based on our discovery of the MAX-effector family and the widely accepted concept that fungal effectors evolve according to a birth-and-death model we propose the hypothesis that the majority of the immense number of different ascomycete effectors could in fact belong to a restricted set of structurally defined families whose members are phylogenetically related. These families of structurally conserved effectors are expected to be, as the MAX-effectors widely distributed with frequent losses on the one hand and lineage specific expansions on the

other leading to effector families that are particularly important in certain fungal clades but not in others. The evolution of individual effectors is so rapid and their adaptation to new functions so profound that sequence homology and resulting phylogenetic signals are rapidly lost although the basic protein architecture may frequently be conserved because it represents a good solution to many general constraints effectors have to face such as stability in the fungus-host interface or translocation into host cytosol. Sequence homology can therefore only be detected in orthologs from closely related species but in paralogs from the same species or homologs from more distantly related species no similarity is detected on the sequence level. Only structure-informed and pattern-based searches reveal the hidden relatedness of ascomycete effectors. This hypothesis is also supported by the recent identification of an effector superfamily in the powdery mildew fungus *Blumeria graminis* fsp *hordei* by structural modelling [51]. 72 effectors from different families established by sequence homology or with no homology to other proteins had 3D structure models with similarity to ribonucleases suggesting a common origin and a conserved structure in this superfamily of sequence diverse effectors.

Knowledge on the structures of fungal effector proteins is extremely limited and outside of the MAX-effectors the structures of only three cytoplasmic fungal effectors have been determined. AvrL567 from the rust fungus *Melampsora lini* and ToxA from *P. tritici-repentis* have distantly related β -sandwich structures whose topologies are completely different from the MAX-effectors and AvrM has a helical structure [68–70]. Therefore, the elucidation of the 3-dimensional structures of additional fungal effectors is a priority for a better understanding of their diversity and will teach us to what extent structurally conserved but sequence-diversified effector families dominate the huge and extremely diverse effector repertoires of phytopathogenic fungi.

Methods

Protein expression and purification

The sequence for the mature protein (residues 20–85 for AVR-Pia, and residues 23–89 for AVR1-CO39) was inserted into the pET-SP vector by ligation of PCR using NdeI-BamHI sites. PCR products were generated using the forward and reverse oligos tatcatatggctGCGCCAGCTAGATTTTGCGTCTAT and tatggatccCTAGTAAGGCTCGGCAGCAAG or tatcatatGCTTG GAAAGATTGCATCATCCA and tatggatccGATCAACAAGACTCATCGTCGTCA for respectively AVR-Pia or AVR1-CO39. The pET-SP vector was constructed from pET-15b (Merck-Millipore, Darmstadt Germany) by inserting a periplasmic secretion sequence, a hexahistidine tag and a TEV cleavage site at the N-terminus of the protein adding an extra 31 amino acid sequence at the N-terminus of the recombinant proteins (sequence MKKTAIAIA VALAGFATVAQA_APQDNTSMGSSHHHHHHSSGRENLYFQGHMA). The plasmids pET-SP-AVR-Pia and pET-SP-AVR1-CO39 were used to transform *E. coli* BL21 (DE3).

Transformed cells were grown in an autoinducing minimal media C-750501 [71] at 37°C for 24h. To generate isotopically-labeled samples for NMR spectroscopy, we used $^{15}\text{NH}_4\text{Cl}$, $^{13}\text{C}_3$ -glycerol and $^{13}\text{C}_6$ -glucose as the primary nitrogen and carbon sources. Cells were harvested by centrifugation and the pellet was resuspended in lysis buffer (200 mM TrisHCl pH8, 200mM Sucrose, 0.05mM EDTA, 50 μM lysozyme). After 30 minutes incubation, cell debris were removed by centrifugation at 12 000 g for 15 min at 4°C. The resulting crude protein extracts were loaded on an AKTA basic system into a HisTrap 5ml HP column (GE Healthcare), equilibrated in buffer A (50 mM TrisHCl, pH 8.0, 300 mM NaCl, 1 mM DTT, 0.1 mM Benzamidine). The His-tagged protein was eluted from the affinity column with buffer B (buffer A supplemented with 500 mM imidazole). Fractions containing the protein were

identified by SDS-PAGE and pooled. The protein was further purified by gel filtration using a Superdex S75 26/60 (GE Healthcare) column in buffer A and pure fractions were pooled.

The elution profiles indicated that AVR-Pia and AVR1-CO39 eluted as single monomeric species (Fig 1). Ellman's reagent, 5, 5'-dithio-bis-(2-nitrobenzoic acid), DTNB, was used for quantitating free sulfhydryl groups [72]. Briefly, aliquots of standard (cysteine, Sigma, 12.5 μ M to 75 μ M) or sample (50 μ M) were reacted with 0.1 mM DTNB reagent in 100 mM sodium phosphate pH 8.0, 1mM EDTA buffer. Free sulfhydryl groups were also measured in denaturing conditions using the same buffer supplemented with 6M Guanidinium Chloride. Absorbance was read at 412 nm on a NanoDrop 2000, and the concentration of free thiols was determined from the standard curves.

NMR samples

The NMR samples were prepared with 1mM of purified protein at 10% D₂O and 0.5 mM DSS as a reference. For AVR-Pia the purification buffer was exchanged with phosphate buffer (20 mM potassium-sodium phosphate, pH 5.4 and 150 mM NaCl), by filtrating with Centricon. The purified AVR1-CO39 proteins were dialyzed in 20 mM sodium phosphate, pH 6.8, 150 mM NaCl and 1 mM DTT. For the D₂O experiments, a non-labeled sample was lyophilized and dissolved in D₂O.

Nuclear magnetic resonance spectroscopy

Spectra were acquired on 500 and 700 MHz Avance Bruker spectrometers equipped with triple-resonance (¹H, ¹⁵N, ¹³C) z-gradient cryo-probe at 305 K. Experiments were recorded using the TOPSPIN pulse sequence library (v. 2.1) (S2 Fig). 2D-NOESY experiments with excitation sculpting water suppression were acquired at 305K, with mixing times from 100 to 150 msec. All spectra are referenced to the internal reference DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) for the ¹H dimension and indirectly referenced for the ¹⁵N and ¹³C dimensions [73].

NMR data was processed using Topspin (v. 3.2) and were analyzed using strip-plots with Cindy in house software and CCPN [74] [analysis v 2.3]. Side chain assignments were carried out using 2D-NOESY, 2D-TOCSY and COSY-DQF experiments with D₂O samples, combined with ¹⁵N-NOESY-HSQC and ¹⁵N-TOCSY-HSQC 3D spectra. For AVR-Pia, the two N-terminal residues Ala-Pro and the His-tag, Ser-His₆-Ser were not assigned. For AVR1-CO39, the tag-residues Asp(-7)-Asn(-8) and the stretch Ser₂-His₆-Ser₂ were not assigned. The ¹⁵N and ¹³C assignments were derived from the 3D spectra at 500 MHz.

¹⁵N backbone amide NMR relaxation data

Relaxation data were acquired at 305K on a Bruker Avance 500 MHz spectrometer using R₁, R₂ and ¹⁵N{¹H} heteronuclear NOE pulse sequences (TOPSPIN library, v 2.1). NMR samples of 500 μ L at 0.85 mM and 0.3 mM were used for AVR-Pia and AVR1-CO39, respectively. R₁ experiments were performed with nine relaxation delays (18, 54, 102, 198, 294, 390, 582, 774 and 966 ms). R₂ experiments were carried out employing a Carr-Purcell-Meiboom-Gill (CPMG) pulse train [75,76] with eight relaxation delays (16, 32, 48, 64, 96, 128, 192 and 256 ms). A recycle delay of 2.5 s was employed in R₁ and R₂, experiments, and ¹⁵N decoupling during acquisition was performed using a GARP-4 sequence. In heteronuclear ¹⁵N{¹H}NOEs, proton saturation was achieved during the relaxation time by application of high-power 120° pulse spaced at 20 ms intervals for 3 s prior to the first pulse on ¹⁵N [77]. A relaxation delay equal to 6 s between each scan was used. Relaxation parameters, R₁, R₂ and NOEs were determined from the analysis module of CCPN [74].

Structure calculation

The programs CYANA [78] and CNS [79] were used for automatic NOE assignments and structure calculations. The NH, H α , ^{15}N , $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts were converted into Φ/Ψ dihedral angle constraints using TALOS+ (v. 1.2) [80]. The CANDID procedure of CYANA (v 2.1) was used to assign the 3D-peaks list from the ^{15}N -NOESY-HSQC spectra. NOE assignments were inspected and used in a new CANDID assignment run including peaks from the 2D-NOESY spectra (with 100 and 150 msec mixing times for AVR-Pia and 100 and 200 msec for AVR1-CO39). A disulfide bridge Cys25-Cys66 for AVR-Pia and between Cys26-Cys61 for AVR1-CO39 was added based on cysteine C β chemical shifts and DTNB quantification of free thiols. NOE constraints were inspected and classified from very strong, strong, medium weak and very weak, corresponding to 2.4, 2.8, 3.6, 4.4 and 4.8 Å upper bound constraints, respectively. Final structure calculations were performed with CYANA (v. 2.1) using 1541 and 1286 distance restraints, for AVR-Pia and AVR1-CO39, with 90 and 72 Φ/Ψ dihedral angle constraints, respectively. The 30 conformers with lowest target function starting from 200 initial structures, were refined by CNS (v. 1.2) using the refinement in water of RECOORD [81]. The final 20 conformers were selected with the lowest NOE and dihedral angle violations. These are the structures discussed herein and deposited (PDBs, 2MYV and 2MYW). The final 20 structures contained no NOE violations greater than 0.3 Å and no dihedral angle constraint violations greater than 2°. Structures were validated using PROCHECK [82].

Sequence analysis

Two sequence databases were used, the UniRef90 release 2015_03 [83] and a database build from the genomes of the ascomycete fungi *Magnaporthe oryzae* (reference isolate 70–15), *Colletotrichum graminicola*, *Colletotrichum higginsianum*, *Fusarium graminearum*, *Fusarium oxysporum*, *Gaeumannomyces graminis*, *Magnaporthe poae*, *Neurospora crassa*, *Pyrenophora tritici-repentis*, *Verticillium dahliae*, *Aspergillus fumigatus*, *Aspergillus nidulans*, *Blumeria graminis*, *Botrytis cinerea*, *Colletotrichum gloeosporioides*, *Colletotrichum orbiculare*, *Dothistroma septosporum*, *Fusarium fujikuroi*, *Fusarium pseudograminearum*, *Fusarium verticillioides*, *Lepidosphaeria maculans*, *Phaeosphaeria nodorum*, *Pyrenophora teres*, *Trichoderma virens*, *Tuber melanosporum* and *Zymoseptoria tritici* (all from the Ensembl Fungi database <http://fungi.ensembl.org>) as well as the genomes of eight *M. oryzae* isolates specific for *Eleusine coracana* (CD156), *Triticum aestivum* (BR32), *Setaria italica* (US71) and *Oryza sativa* (TH16, GY11, FR13, TH12, PH14) and one *M. grisea* isolate (BR29) pathogenic to *Digitaria* ssp (genome sequences at <http://genome.jouy.inra.fr/gemo>) [47]. Sequences without signal peptide (according to SIGNALP 4.1 [84]) bigger than 170 amino acids or with less than 2 cysteine residues were removed. For the initial HMM search, identical sequences were reduced to only one occurrence in the databases.

The start of the search was a structural alignment with TM-align [85] and the structures of AVR-Pia, AVR1-CO39, AvrPiz-t and ToxB complemented with sequence homologues found by single queries using BLAST (v 2.2.27+) with a stringent cut-off E-value = 1e-6. For the ToxB query, two iterations of NCBI PSI-BLAST were used on the NR database with a cut-off E-value = 1e-4 (S5A Fig).

This initial alignment was used as input to look for homologs in the filtered and non-redundant fungi database using HMMERsearch program from the HMMER package v 3.0 [86] with a 1e-6 E-value cut-off. For each run, only sequences where the two cysteine residues were aligned were kept, and the output alignment was used as input query for a new HMMERsearch run. This run was repeated until reaching convergence. New iterations were then done with

increased E-value cut-off at $1e-5$ and $1e-4$. From the last alignment, a histogram indicated that the two aligned cysteine residues were separated by at least 34 and at most 49 amino acids.

The full homolog search was re-started, as described above, but this time using also the aligned cysteine separations as an additional constraint for filtering homologs after each HMMERsearch run. The HMMERsearch runs were repeated until convergence for raised threshold E-values $1e-6$, $1e-5$, $1e-4$ and finally $1e-2$. The homolog ensembles obtained for the three E-values cut-off, $1e-6$, $1e-4$ and $1e-2$ were aligned with Muscle v3.8 [87] (S6A Fig for E-value $1e-4$). The derived logo was built from the HMMER search with E-value of $1e-4$ using Weblogo3 [88]. The multiple sequence alignment (MSA) derived from the HMMER search with E-value $1e-4$ was used as input to look for homologs in the redundant fungi database and the UniRef90 database, using HMMERsearch with an E-value threshold of $1e-1$. Diversity trees were built from alignments generated with Muscle v3.8 using the Neighbor-Joining method with the MEGA6 package [89].

Fungal growth and infection assays

For analysis of gene expression *in vitro* grown mycelium, *M. oryzae* isolate Guy11 was grown in liquid medium (glucose 10g/L, KNO_3 3g/L, KH_2PO_4 2g/L and yeast extract 2g/L) at 120 rpm on a rotary shaker at 25°C for five days. Mycelium was harvested over a piece of cheese-cloth (Merck-Millipore, Darmstadt Germany).

For production of spores for infection assays, *M. oryzae* isolate Guy11 was grown on rice flour agar for spore production [90]. A suspension of fungal conidiospores was prepared at a density of 2×10^5 spores/ml and spotted on detached leaves of the japonica rice variety Saraceltik grown for 3 weeks as described [91]. Infected leaf samples were harvested 16, 24, 48 and 72 hours post inoculation (hpi).

RNA extraction and qRT-PCR analysis

RNA extraction and reverse transcription was performed as described [92]. Quantitative PCR were performed with a LightCycler 480 instrument (Roche, Basel, Switzerland) using LC 480 SYBR Green I Master Mix (Roche) and the primers listed in the S4 Table. Amplification was performed as follows: 95°C for 10 min; 40 cycles of 95°C for 15 s, 60°C for 20s and 72°C for 30 s; then 95°C for 5 min and 40°C for 30 s. Data were analyzed using the delta-delta Ct method and applying the formula $2^{-\Delta\Delta CT}$, where $\Delta\Delta CT$ is the difference in threshold cycle (CT) between the gene of interest and the housekeeping gene *Actin* (MGG_03982) used as a constitutively expressed reference gene. For each condition, three biological replicates were analyzed.

Supporting Information

S1 Table. NMR experiments acquired for structure calculations and chemical shift assignments.

(PDF)

S2 Table. DALI statistics for structural alignment of AVR-Pia, AVR1-C039, AvrPiz-t and ToxB.

(PDF)

S3 Table. MAX-effector candidates identified by psi-Blast in the genomes of the *M. oryzae* isolates 70–15 and TH16 and the *M. grisea* isolate BR29.

(PDF)

S4 Table. Primers used for qRT-PCR.

(PDF)

S1 Fig. Gel filtration profile and SDS-PAGE analysis of purified AVR-Pia (A) and AVR1-CO39 (B) proteins.

(PDF)

S2 Fig. ¹⁵N Relaxation data at 500 MHz and 305K for AVR-Pia (panels A, B and C) and AVR1-CO39 (panels D, E and F).

(PDF)

S3 Fig. Backbone sequential and medium range NOEs observed for (A) AVR-Pia and (B) AVR1-CO39. The line width is proportional to the NOE intensity. The dots (•) indicate slow exchange NH observed in 2D-NOESY in D₂O. Grey arrows indicate the β-strands determined from the structure analysis (vide infra).

(PDF)

S4 Fig. Solution structures of (A) AVR-Pia and (B) AVR1-CO39. Superposition of the backbone atoms of the 20 lowest energy conformers used to calculate the final structures. Only mature chains are shown, from residues Ala20 and Trp23 for AVR-Pia and AVR1-CO39, respectively.

(PDF)

S5 Fig. Structure-guided alignment and diversity of MAX-effector homologs identified by Blast. A) Homologs of AVR1-CO39, AvrPiz-t and AVR-Pia identified by Blast in *M. oryzae* and *M. grisea* genomes and ToxB homologs identified by Psi-Blast in the GeneBank database were aligned to the structural alignment of mature ToxB, AVR1-CO39, AvrPiz-t and AVR-Pia. (B) A diversity tree was constructed by the neighbor-joining method using the alignment in (A). It highlights the high diversity of MAX-effector homologs. Branch supports are based on 1000 bootstraps and horizontal branch length reflects sequence divergence. Accession numbers of non-Magnaporthe sequences were completed by a 2 letter identifier for the species: BO for *Bipolaris oryzae*, CF is for *Colletotrichum fioriniae*, CH for *C. higginsianum*, CG for *C. gloeosporioides*, CO for *C. orbiculare*, LM for *Lepthosphaeria maculans*, PT for *Pyrenophora tritici-repentis* and PB for *Pyrenophora bromi*.

(PDF)

S6 Fig. MAX-effector homologs identified by a high stringency HMM search. (A) Histogram showing the numbers of MAX-effectors identified by an HMM pattern search in a non-redundant database comprising the small secreted proteins of 25 ascomycete fungi and of 8 additional *M. oryzae* and one *M. grisea* isolate. (B) MAX-effectors were aligned to the structural alignment of mature ToxB, AVR1-CO39, AvrPiz-t and AVR-Pia and gaps were removed. (C) A diversity tree was constructed by the neighbor-joining method using the alignment in (B). Branch supports are based on 1000 bootstraps and horizontal branch length reflects sequence divergence. Accession numbers of non-Magnaporthe sequences were completed by a 2 letter identifier for the species: BO for *Bipolaris oryzae*, CF for *Colletotrichum fioriniae*, CH for *C. higginsianum*, CG for *C. gloeosporioides*, CO for *C. orbiculare*, LM for *Lepthosphaeria maculans*, PT for *Pyrenophora tritici-repentis* PB for *Pyrenophora bromi* and ZT for *Zymoseptoria tritici*.

(PDF)

S7 Fig. MAX-effector homologs identified in the UniRef90 database by a low stringency HMM search. (A) Histogram showing the numbers of MAX-effectors identified by an HMM

pattern search in a non-redundant UniRef90 database. (B) MAX-effectors identified by HMM pattern search were aligned to the structural alignment of mature ToxB, AVR1-CO39, AvrPiz-t and AVR-Pia. (C) A diversity tree was constructed by the neighbor-joining method using the alignment in (B). This highlights the high diversity of MAX-effector homologs. Branch supports are based on 1000 bootstraps and horizontal branch length reflects sequence divergence. Accession numbers contain the following information on the species: MAGGR, MAGO7, MAGOP and MAGOR are from *M. oryzae*, COLGC and COLGN from *C. gloeosporioides*, COLHI from *Colletotrichum higginsianum*, 9PEZI from *C. fioriniae* and COLOR from *Colletotrichum orbiculare*, 9PLEO from *P. tritici-repentis* or *P. bromi*, ARTOA from *Arthrobotrys oligospora*, COCMI from *Bipolaris oryzae*, LEPMJ from *Leptosphaeria maculans*, MYCGM from *Zymoseptoria tritici*, 9PSED from *Pseudomonas sp. StFLB209* and SOLLC from *Solanum lycopersicum*.
(PDF)

S8 Fig. MAX-effector homologs identified by a low stringency HMM search in the fungal genomes database. (A) MAX-effectors identified by an HMM pattern search in a redundant database comprising the small secreted proteins of 25 ascomycete fungi, 8 additional *M. oryzae* and one *M. gisea* isolate were aligned to the structural alignment of mature ToxB, AVR1-CO39, AvrPiz-t and AVR-Pia and gaps were removed. (B) A diversity tree was constructed by the neighbor-joining method using the alignment in (A). Branch supports are based on 1000 bootstraps and horizontal branch length reflects sequence divergence. Accession numbers of non-Magnaporthe sequences were completed by a 2 letter identifier for the species: BO for *Bipolaris oryzae*, CF for *Colletotrichum fioriniae*, CH for *C. higginsianum*, CG for *C. gloeosporioides*, CO for *C. orbiculare*, GF for *Fusarium fujicuroi*, LM for *Leptosphaeria maculans*, PT for *Pyrenophora tritici-repentis*, PB for *Pyrenophora bromi* and ZT for *Zymoseptoria tritici*.
(PDF)

S9 Fig. Expression of *M. oryzae* MAX-effector candidates and marker genes during rice infection and in *in vitro* grown mycelium. mRNA levels of *M. oryzae* genes coding for MAX-effectors (A, B and C) and marker genes (D) was determined by q-RT-PCR in rice leaf samples harvested 16, 24, 48 or 72 h after inoculation and mycelium grown liquid medium for 72 hours. (A) Infection specific MAX-effectors identified in the HMM search, (B) infection specific MAX-effectors identified in the Psi Blast search but nor in the HMM search, (C) constitutively expressed MAX-effectors identified in the HMM search and (D) marker genes for appressorium and very early infection (*ORF3* of the ACE1 cluster, *MGG_08381*), biotrophic infection (*BAS3*, *MGG_11610*), late infection (*MGG_01147*), constitutive expression (*EF1 α* , *MGG_03641*). Relative expression levels were calculated by using expression of a constitutively expressed *Actin* (*MGG_03982*) as a reference. Mean values and standard deviation were calculated from three independent biological samples. The analyzed genes, were in most cases not or extremely weakly expressed in the mycelium. For genes with significant expression in the mycelium (ratio gene versus actine > 0,01) a T-test was performed to determine if in planta expression was significantly different from expression in the mycelium. In these cases (*MGG_11967*, *MGG_14793*, *MGG_15207*, *MGG_17266*, *MGG_18019*, *M_TH16_00000541*, *M_TH16_00040131*, *M_TH16_00079081*, *M_TH16_00104561*, *M_TH16_00120731*, *M_TH16_00124981*), a star or two stars (* or **) mark conditions where the expression was different from expression in the mycelium at respectively $p < 0,05$ or $p < 0,005$.
(PDF)

S10 Fig. Prediction of the secondary structure of *M. oryzae* MAX effectors. The secondary structures of the MAX-effectors from the 70–15, TH16 and BR29 genomes was predicted with SS-PRO5 [93]. The predictions are shown at the bottom of the figure and are aligned onto the corresponding primary sequence alignment shown at the top of the figure. Sequence identifiers for the secondary structure predictions are suffixed with ".2d.SS-PRO5". Blue "H", red "E" and yellow "C" correspond respectively to helix, extended sheet and coil predictions. The sequences of the 4 MAX effectors with experimentally determined structures are displayed at the top of the multiple sequence alignment and, for clarity, the alignment positions corresponding to shared gaps in the known structures were removed.

(TIF)

Acknowledgments

Véronique Chalvon is acknowledged for technical assistance. Martin Cohen-Gonsaud, Christian Roumestand, Jean-Benoit Morel and Stella Césari are acknowledged for fruitful discussions.

Author Contributions

Conceived and designed the experiments: TK AP JG. Performed the experiments: DOV KdG JG AP. Analyzed the data: JG DOV KdG EF TK AP. Contributed reagents/materials/analysis tools: EF. Wrote the paper: TK AP JG.

References

1. Hogenhout S a, Van der Hoorn R a L, Terauchi R, Kamoun S. Emerging concepts in effector biology of plant-associated organisms. *Mol Plant Microbe Interact.* 2009; 22: 115–122. doi: [10.1094/MPMI-22-2-0115](https://doi.org/10.1094/MPMI-22-2-0115) PMID: [19132864](https://pubmed.ncbi.nlm.nih.gov/19132864/)
2. Jones JDG, Dangl JL. The plant immune system. *Nature.* 2006; 444: 323–9. PMID: [17108957](https://pubmed.ncbi.nlm.nih.gov/17108957/)
3. Doehlemann G, Requena N, Schaefer P, Brunner F, Connell RO, Parker JE. Reprogramming of plant cells by filamentous plant-colonizing microbes. 2014; 803–814.
4. Stergiopoulos I, de Wit PJGM. Fungal effector proteins. *Annu Rev Phytopathol.* 2009; 47: 233–63. doi: [10.1146/annurev.phyto.112408.132637](https://doi.org/10.1146/annurev.phyto.112408.132637) PMID: [19400631](https://pubmed.ncbi.nlm.nih.gov/19400631/)
5. Presti L Lo, 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. doi: [10.1146/annurev-arplant-043014-114623](https://doi.org/10.1146/annurev-arplant-043014-114623) PMID: [25923844](https://pubmed.ncbi.nlm.nih.gov/25923844/)
6. Giraldo MC, Valent B. Filamentous plant pathogen effectors in action. *Nat Rev Microbiol.* Nature Publishing Group; 2013; 11: 800–14. doi: [10.1038/nrmicro3119](https://doi.org/10.1038/nrmicro3119) PMID: [24129511](https://pubmed.ncbi.nlm.nih.gov/24129511/)
7. Bozkurt TO, Schomack S, Banfield MJ, Kamoun S. Oomycetes, effectors, and all that jazz. *Curr Opin Plant Biol.* Elsevier Ltd; 2012; 1–10.
8. Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature.* 2009; 461: 393–8. doi: [10.1038/nature08358](https://doi.org/10.1038/nature08358) PMID: [19741609](https://pubmed.ncbi.nlm.nih.gov/19741609/)
9. Jiang RHY, Tripathy S, Govers F, Tyler BM. RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proc Natl Acad Sci U S A.* 2008; 105: 4874–4879. doi: [10.1073/pnas.0709303105](https://doi.org/10.1073/pnas.0709303105) PMID: [18344324](https://pubmed.ncbi.nlm.nih.gov/18344324/)
10. Duplessis S, Cuomo C a, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci U S A.* 2011; 108: 9166–71. doi: [10.1073/pnas.1019315108](https://doi.org/10.1073/pnas.1019315108) PMID: [21536894](https://pubmed.ncbi.nlm.nih.gov/21536894/)
11. Pedersen C, Themaat EVL van, McGuffin LJ, 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](https://doi.org/10.1186/1471-2164-13-694) PMID: [23231440](https://pubmed.ncbi.nlm.nih.gov/23231440/)
12. Spanu PD, Abbott JC, Amselem J, Burgis T a, Soanes DM, Stüber K, et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science.* 2010; 330: 1543–6. doi: [10.1126/science.1194573](https://doi.org/10.1126/science.1194573) PMID: [21148392](https://pubmed.ncbi.nlm.nih.gov/21148392/)

13. Hacquard S, Joly DL, Lin Y-C, Tisserant E, Feau N, Delaruelle C, et al. A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Mol Plant Microbe Interact*. 2012; 25: 279–93. doi: [10.1094/MPMI-09-11-0238](https://doi.org/10.1094/MPMI-09-11-0238) PMID: [22046958](https://pubmed.ncbi.nlm.nih.gov/22046958/)
14. Dean R, Kan JANALVAN, Pretorius ZA, Hammond-kosack KIME, Pietro ADI, Spanu PD, et al. The Top 10 fungal pathogens in molecular plant pathology. 2012;13: 414–430.
15. Gurr S, Samalova M, Fisher M. The rise and rise of emerging infectious fungi challenges food security and ecosystem health. *Fungal Biol Rev*. 2011; 25: 181–188.
16. Skamnioti P, Gurr SJ. Against the grain: safeguarding rice from rice blast disease. *Trends Biotechnol*. 2009; 27: 141–50. doi: [10.1016/j.tibtech.2008.12.002](https://doi.org/10.1016/j.tibtech.2008.12.002) PMID: [19187990](https://pubmed.ncbi.nlm.nih.gov/19187990/)
17. Galhano R, Talbot NJ. The biology of blast: Understanding how *Magnaporthe oryzae* invades rice plants. *Fungal Biol Rev*. Elsevier Ltd; 2011; 25: 61–67.
18. Liu J, Wang X, Mitchell T, Hu Y, Liu X, Dai L, et al. Recent progress and understanding of the molecular mechanisms of the rice-*Magnaporthe oryzae* interaction. *Mol Plant Pathol*. 2010; 11: 419–427. doi: [10.1111/j.1364-3703.2009.00607.x](https://doi.org/10.1111/j.1364-3703.2009.00607.x) PMID: [20447289](https://pubmed.ncbi.nlm.nih.gov/20447289/)
19. Mentlak TA, Talbot NJ, Kroj T. Effector Translocation and Delivery by the Rice Blast Fungus *Magnaporthe oryzae*. In: Francisrtin, Kamoun S, editors. *Effectors in Plant–Microbe Interactions*. Wiley-Blackwell; 2011. pp. 219–241.
20. Valent B, Khang CH. Recent advances in rice blast effector research. *Curr Opin Plant Biol*. Elsevier Ltd; 2010; 13: 434–41. doi: [10.1016/j.pbi.2010.04.012](https://doi.org/10.1016/j.pbi.2010.04.012) PMID: [20627803](https://pubmed.ncbi.nlm.nih.gov/20627803/)
21. Park C-H, Chen S, Shirsekar G, Zhou B, Khang CH, Songkumarn P, et al. The *Magnaporthe oryzae* Effector AvrPiz-t Targets the RING E3 Ubiquitin Ligase APIP6 to Suppress Pathogen-Associated Molecular Pattern-Triggered Immunity in Rice. *Plant Cell*. 2012.
22. Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, et al. Effector-mediated suppression of chitin-triggered immunity by *magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell*. 2012; 24: 322–35. doi: [10.1105/tpc.111.092957](https://doi.org/10.1105/tpc.111.092957) PMID: [22267486](https://pubmed.ncbi.nlm.nih.gov/22267486/)
23. Mosquera G, Giraldo MC, Khang CH, Coughlan S, Valent B. Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as Biotrophy-associated secreted proteins in rice blast disease. *Plant Cell*. 2009; 21: 1273–90. doi: [10.1105/tpc.107.055228](https://doi.org/10.1105/tpc.107.055228) PMID: [19357089](https://pubmed.ncbi.nlm.nih.gov/19357089/)
24. Saitoh H, Fujisawa S, Mitsuoka C, Ito A, Hirabuchi A, Ikeda K, et al. Large-scale gene disruption in *Magnaporthe oryzae* identifies MC69, a secreted protein required for infection by monocot and dicot fungal pathogens. *PLoS Pathog*. 2012; 8: e1002711. doi: [10.1371/journal.ppat.1002711](https://doi.org/10.1371/journal.ppat.1002711) PMID: [22589729](https://pubmed.ncbi.nlm.nih.gov/22589729/)
25. Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, et al. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. *PLoS One*. 2008; 3: e2300. doi: [10.1371/journal.pone.0002300](https://doi.org/10.1371/journal.pone.0002300) PMID: [18523684](https://pubmed.ncbi.nlm.nih.gov/18523684/)
26. Yoshida KK, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Tosa Y, et al. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell*. 2009; 21: 1573–91. doi: [10.1105/tpc.109.066324](https://doi.org/10.1105/tpc.109.066324) PMID: [19454732](https://pubmed.ncbi.nlm.nih.gov/19454732/)
27. Chen X, Coram T, Huang X, Wang M, Dolezal A. Understanding Molecular Mechanisms of Durable and Non-durable Resistance to Stripe Rust in Wheat Using a Transcriptomics Approach. *Curr Genomics*. 2013; 14: 111–126. doi: [10.2174/1389202911314020004](https://doi.org/10.2174/1389202911314020004) PMID: [24082821](https://pubmed.ncbi.nlm.nih.gov/24082821/)
28. Kim S, Hu J, Oh Y, Park J, Choi J, Lee Y-H, et al. Combining ChIP-chip and expression profiling to model the MoCRZ1 mediated circuit for Ca/calcieneurin signaling in the rice blast fungus. *PLoS Pathog*. 2010; 6: e1000909. doi: [10.1371/journal.ppat.1000909](https://doi.org/10.1371/journal.ppat.1000909) PMID: [20502632](https://pubmed.ncbi.nlm.nih.gov/20502632/)
29. Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet*. Nature Publishing Group; 2010; 11: 539–48. doi: [10.1038/nrg2812](https://doi.org/10.1038/nrg2812) PMID: [20585331](https://pubmed.ncbi.nlm.nih.gov/20585331/)
30. Miki H, Matsui K, Kito H, Otsuka K, Ashizawa T, Yasuda N, et al. Molecular cloning and characterization of the AVR-Pia locus from a Japanese field isolate of *Magnaporthe oryzae*. 2009;10: 361–374.
31. Orbach MJ, Farrall L, Sweigard J a, Chumley FG, Valent B. A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. *Plant Cell*. 2000; 12: 2019–32. PMID: [11090206](https://pubmed.ncbi.nlm.nih.gov/11090206/)
32. Ribot C, Césari S, Abidi I, Chalvon V, Bournaud C, Vallet J, et al. The *Magnaporthe oryzae* effector AVR1-CO39 is translocated into rice cells independently of a fungal-derived machinery. *Plant J*. 2012; 74: 1–12.
33. Sweigard J a, Carroll a M, Kang S, Farrall L, Chumley FG, Valent B. Identification, cloning, and characterization of PWL2, a gene for host species specificity in the rice blast fungus. *Plant Cell*. 1995; 7: 1221–33. PMID: [7549480](https://pubmed.ncbi.nlm.nih.gov/7549480/)
34. Wu J, Kou Y, Bao J, Li Y, Tang M, Zhu X, et al. Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice. 2015;

35. Li W, Wang B, Wu J, Lu G, Hu Y, Zhang X, et al. The Magnaporthe oryzae avirulence gene AvrPiz-t encodes a predicted secreted protein that triggers the immunity in rice mediated by the blast resistance gene Piz-t. *Mol Plant Microbe Interact.* 2009; 22: 411–20. doi: [10.1094/MPMI-22-4-0411](https://doi.org/10.1094/MPMI-22-4-0411) PMID: [19271956](https://pubmed.ncbi.nlm.nih.gov/19271956/)
36. Ashikawa I, Hayashi N, Yamane H, Kanamori H, Wu J, Matsumoto T, et al. Two adjacent nucleotide-binding site-leucine-rich repeat class genes are required to confer Pikm-specific rice blast resistance. *Genetics.* 2008; 180: 2267–76. doi: [10.1534/genetics.108.095034](https://doi.org/10.1534/genetics.108.095034) PMID: [18940787](https://pubmed.ncbi.nlm.nih.gov/18940787/)
37. Cesari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, et al. The rice resistance protein pair RGA4/RGA5 recognizes the Magnaporthe oryzae effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell.* 2013; 25: 1463–81. doi: [10.1105/tpc.112.107201](https://doi.org/10.1105/tpc.112.107201) PMID: [23548743](https://pubmed.ncbi.nlm.nih.gov/23548743/)
38. Qu S, Liu G, Zhou B, Bellizzi M, Zeng L, Dai L, et al. The broad-spectrum blast resistance gene Pi9 encodes a nucleotide-binding site-leucine-rich repeat protein and is a member of a multigene family in rice. *Genetics.* 2006; 172: 1901–14. PMID: [16387888](https://pubmed.ncbi.nlm.nih.gov/16387888/)
39. Okuyama Y, Kanzaki H, Abe A, Yoshida K, Tamiru M, Saitoh H, et al. A multifaceted genomics approach allows the isolation of the rice Pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes. *Plant J.* 2011; 66: 467–79. doi: [10.1111/j.1365-3113.2011.04502.x](https://doi.org/10.1111/j.1365-3113.2011.04502.x) PMID: [21251109](https://pubmed.ncbi.nlm.nih.gov/21251109/)
40. Bryan GT, Wu KS, Farrall L, Jia Y, Hershey HP, McAdams S a, et al. tA single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene Pi-ta. *Plant Cell.* 2000; 12: 2033–46. PMID: [11090207](https://pubmed.ncbi.nlm.nih.gov/11090207/)
41. Zhou B, Qu S, Liu G, Dolan M, Sakai H, Lu G, et al. The eight amino-acid differences within three leucine-rich repeats between Pi2 and Piz-t resistance proteins determine the resistance specificity to Magnaporthe grisea. *Mol Plant Microbe Interact.* 2006; 19: 1216–28. PMID: [17073304](https://pubmed.ncbi.nlm.nih.gov/17073304/)
42. Zhang Z-M, Zhang X, Zhou Z-R, Hu H-Y, Liu M, Zhou B, et al. Solution structure of the Magnaporthe oryzae avirulence protein AvrPiz-t. *J Biomol NMR.* 2013; 55: 219–23. doi: [10.1007/s10858-012-9695-5](https://doi.org/10.1007/s10858-012-9695-5) PMID: [23334361](https://pubmed.ncbi.nlm.nih.gov/23334361/)
43. Nyarko A, Singarapu KK, Figueroa M, Manning V a., Pandelova I, Wolpert TJ, et al. Solution NMR Structures of Pyrenophora tritici-repentis ToxB and Its Inactive Homolog Reveal Potential Determinants of Toxin Activity. *J Biol Chem.* 2014; 289: 25946–25956. doi: [10.1074/jbc.M114.569103](https://doi.org/10.1074/jbc.M114.569103) PMID: [25063993](https://pubmed.ncbi.nlm.nih.gov/25063993/)
44. Barthe P, Ropars V, Roumestand C. DYNAMOF: a program for the dynamics analysis of relaxation data obtained at multiple magnetic fields. *Comptes Rendus Chim.* 2006; 9: 503–513.
45. Holm L, Rosenström P. Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* 2010; 38: 545–549.
46. Ciuffetti LM, Manning V a., Pandelova I, Betts MF, Martinez JP. Host-selective toxins, Ptr ToxA and Ptr ToxB, as necrotrophic effectors in the Pyrenophora tritici-repentis-wheat interaction. *New Phytol.* 2010; 187: 911–919. doi: [10.1111/j.1469-8137.2010.03362.x](https://doi.org/10.1111/j.1469-8137.2010.03362.x) PMID: [20646221](https://pubmed.ncbi.nlm.nih.gov/20646221/)
47. Chiapello H, Mallet L, Guérin C, Aguilera G, Amselem J, Kroy T, et al. Deciphering genome content and evolutionary relationships of isolates from the fungus Magnaporthe oryzae attacking different hosts. *Genome Biol Evol.* 2015; in press.
48. Klaubauf S, Tharreau D, Fournier E, Groenewald JZ, Crous PW, de Vries RP, et al. Resolving the phylogenetic nature of Pyricularia (Pyriculariaceae). *Stud Mycol. ELSEVIER B.V.* 2014; 79: 85–120. doi: [10.1016/j.simyco.2014.09.004](https://doi.org/10.1016/j.simyco.2014.09.004) PMID: [25492987](https://pubmed.ncbi.nlm.nih.gov/25492987/)
49. Césari S, Kanzaki H, Fujiwara T, Bernoux M, Chalvon V, Kawano Y, et al. The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *EMBO J.* 2014; 33: 1941–1959. doi: [10.15252/embj.201487923](https://doi.org/10.15252/embj.201487923) PMID: [25024433](https://pubmed.ncbi.nlm.nih.gov/25024433/)
50. Césari S, Bernoux M, Moncuquet P, Kroy T, Dodds PN. A novel conserved mechanism for plant NLR protein pairs: the “integrated decoy” hypothesis. *Front Plant Sci.* 2014; 5: 606. doi: [10.3389/fpls.2014.00606](https://doi.org/10.3389/fpls.2014.00606) PMID: [25506347](https://pubmed.ncbi.nlm.nih.gov/25506347/)
51. Pedersen C, Themaat V, Ver E, McGuffin L, Abbott JC, Burgis TA, et al. Structure and evolution of barley powdery mildew effector candidates. *BMC Genomics.* 2012
52. Chuma I, Isobe C, Hotta Y, Ibaragi K, Futamata N, Kusaba M, et al. Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus Magnaporthe oryzae and related species. *PLoS Pathog.* 2011; 7: e1002147. doi: [10.1371/journal.ppat.1002147](https://doi.org/10.1371/journal.ppat.1002147) PMID: [21829350](https://pubmed.ncbi.nlm.nih.gov/21829350/)
53. 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. doi: [10.1038/ncomms1189](https://doi.org/10.1038/ncomms1189) PMID: [21326234](https://pubmed.ncbi.nlm.nih.gov/21326234/)
54. Ma L-J, van der Does HC, Borkovich K a, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature.* 2010; 464: 367–73. doi: [10.1038/nature08850](https://doi.org/10.1038/nature08850) PMID: [20237561](https://pubmed.ncbi.nlm.nih.gov/20237561/)

55. Jonge R De, Bolton MD, Kombrink A, De Jonge R, Bolton MD, Kombrink A, et al. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res.* 2013; 23: 1271–1282. doi: [10.1101/gr.152660.112](https://doi.org/10.1101/gr.152660.112) PMID: [23685541](https://pubmed.ncbi.nlm.nih.gov/23685541/)
56. Win J, Krasileva K V., Kamoun S, Shirasu K, Staskawicz BJ, Banfield MJ. Sequence Divergent RXLR Effectors Share a Structural Fold Conserved across Plant Pathogenic Oomycete Species. Heitman J, editor. *PLoS Pathog.* 2012; 8: e1002400. doi: [10.1371/journal.ppat.1002400](https://doi.org/10.1371/journal.ppat.1002400) PMID: [22253591](https://pubmed.ncbi.nlm.nih.gov/22253591/)
57. Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, et al. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science.* 2010; 330: 1549–1551. doi: [10.1126/science.1195203](https://doi.org/10.1126/science.1195203) PMID: [21148394](https://pubmed.ncbi.nlm.nih.gov/21148394/)
58. Raffaele S, Farrer R a, Cano LM, Studholme DJ, MacLean D, Thines M, et al. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science.* 2010; 330: 1540–3. doi: [10.1126/science.1193070](https://doi.org/10.1126/science.1193070) PMID: [21148391](https://pubmed.ncbi.nlm.nih.gov/21148391/)
59. Dou D, Kale SD, Wang X, Chen Y, Wang Q, Wang X, et al. Conserved C-terminal motifs required for avirulence and suppression of cell death by *Phytophthora sojae* effector Avr1b. *Plant Cell.* 2008; 20: 1118–33. doi: [10.1105/tpc.107.057067](https://doi.org/10.1105/tpc.107.057067) PMID: [18390593](https://pubmed.ncbi.nlm.nih.gov/18390593/)
60. Boutemy LS, King SRF, Win J, Hughes RK, Clarke T a, Blumenschein TM a, et al. Structures of *Phytophthora* RXLR effector proteins: a conserved but adaptable fold underpins functional diversity. *J Biol Chem.* 2011; 286: 35834–42. doi: [10.1074/jbc.M111.262303](https://doi.org/10.1074/jbc.M111.262303) PMID: [21813644](https://pubmed.ncbi.nlm.nih.gov/21813644/)
61. Chou S, Krasileva K V, Holton JM, Steinbrenner AD, Alber T, Staskawicz BJ. *Hyaloperonospora arabidopsidis* ATR1 effector is a repeat protein with distributed recognition surfaces. *Proc Natl Acad Sci U S A.* 2011; 108: 13323–8. doi: [10.1073/pnas.1109791108](https://doi.org/10.1073/pnas.1109791108) PMID: [21788488](https://pubmed.ncbi.nlm.nih.gov/21788488/)
62. Yaeno T, Li H, Chaparro-Garcia A, Schornack S, Koshiba S, Watanabe S, et al. Phosphatidylinositol monophosphate-binding interface in the oomycete RXLR effector AVR3a is required for its stability in host cells to modulate plant immunity. *Proc Natl Acad Sci U S A.* 2011; 108: 14682–7. doi: [10.1073/pnas.1106002108](https://doi.org/10.1073/pnas.1106002108) PMID: [21821794](https://pubmed.ncbi.nlm.nih.gov/21821794/)
63. Bayry J, Aïmanianda V, Guijarro JI, Sunde M, Latgé JP. Hydrophobins-unique fungal proteins. *PLoS Pathog.* 2012; 8: 6–9.
64. Kubicek CP, Baker S, Gamauf C, Kenerley CM, Druzhinina IS. Purifying selection and birth-and-death evolution in the class II hydrophobin gene families of the ascomycete *Trichoderma/Hypocrea*. *BMC Evol Biol.* 2008; 8: 4. doi: [10.1186/1471-2148-8-4](https://doi.org/10.1186/1471-2148-8-4) PMID: [18186925](https://pubmed.ncbi.nlm.nih.gov/18186925/)
65. Wosten H a. HYDROPHOBINS: Multipurpose Proteins. *Annu Rev Microbiol.* 2001; 55: 625–46. PMID: [11544369](https://pubmed.ncbi.nlm.nih.gov/11544369/)
66. Kwan a HY, Winefield RD, Sunde M, Matthews JM, Haverkamp RG, Templeton MD, et al. Structural basis for rodlet assembly in fungal hydrophobins. *Proc Natl Acad Sci U S A.* 2006; 103: 3621–3626. PMID: [16537446](https://pubmed.ncbi.nlm.nih.gov/16537446/)
67. Stergiopoulos I, Kourmpetis Y a I, Slot JC, Bakker FT, De Wit PJGM, Rokas A. In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. *Mol Biol Evol.* 2012; 29: 3371–84. doi: [10.1093/molbev/mss143](https://doi.org/10.1093/molbev/mss143) PMID: [22628532](https://pubmed.ncbi.nlm.nih.gov/22628532/)
68. Ve T, Williams SJ, Catanzariti A-M, Rafiqi M, Rahman M, Ellis JG, et al. Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. *Proc Natl Acad Sci U S A.* 2013
69. Wang C-IAC-I a, Guncar G, Forwood JK, Teh T, Catanzariti A-MA-M, Lawrence GJ, et al. Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell.* 2007; 19: 2898–912. PMID: [17873095](https://pubmed.ncbi.nlm.nih.gov/17873095/)
70. Sarma GN, Manning VA, Ciuffetti LM, Karplus PA. Structure of Ptr ToxA: An RGD-Containing Host-Selective Toxin from *Pyrenophora tritici-repentis*. 2005;17: 3190–3202.
71. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif.* 2005; 41: 207–234. PMID: [15915565](https://pubmed.ncbi.nlm.nih.gov/15915565/)
72. Habeeb AFSA. [37] Reaction of protein sulfhydryl groups with Ellman's reagent. In: Hirs C. H. W. *SNT*, editor. *Methods in Enzymology.* Academic Press; 1972. pp. 457–464. doi: [10.1016/S0076-6879\(72\)25041-8](https://doi.org/10.1016/S0076-6879(72)25041-8)
73. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, et al. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR.* 1995; 6: 135–140. PMID: [8589602](https://pubmed.ncbi.nlm.nih.gov/8589602/)
74. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct Funct Bioinforma.* 2005; 59: 687–696.
75. Carr HY, Purcell EM. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys Rev.* 1954; 94: 630–632.

76. Meiboom S, Gill D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev Sci Instrum.* 1958; 29: 688.
77. Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry.* 1989; 28: 8972–8979. PMID: [2690953](#)
78. Güntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol.* 2004; 278: 353–378. PMID: [15318003](#)
79. Brünger AT. Version 1.2 of the Crystallography and NMR system. *Nat Protoc.* 2007; 2: 2728–2733. PMID: [18007608](#)
80. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR.* 2009; 44: 213–223. doi: [10.1007/s10858-009-9333-z](#) PMID: [19548092](#)
81. Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CAEM, Nabuurs SB, et al. RECOORD: A recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagRes-Bank. *Proteins Struct Funct Bioinforma.* 2005; 59: 662–672.
82. Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. *J Mol Biol.* 1993; 231: 1049–1067. PMID: [8515464](#)
83. Sużek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniProt: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics.* 2007; 23: 1282–1288. PMID: [17379688](#)
84. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011; 8: 785–786. doi: [10.1038/nmeth.1701](#) PMID: [21959131](#)
85. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 2005; 33: 2302–2309. PMID: [15849316](#)
86. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011; 39: W29–W37. doi: [10.1093/nar/gkr367](#) PMID: [21593126](#)
87. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32: 1792–1797. PMID: [15034147](#)
88. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14: 1188–1190. PMID: [15173120](#)
89. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013; 30: 2725–2729. doi: [10.1093/molbev/mst197](#) PMID: [24132122](#)
90. Berruyer R, Adreit H, Milazzo J, Gaillard S, Berger a., Dió W, et al. Identification and fine mapping of Pi33, the rice resistance gene corresponding to the Magnaporthe grisea avirulence gene ACE1. *Theor Appl Genet.* 2003; 107: 1139–1147. PMID: [12838393](#)
91. Faivre-Rampant O, Thomas J, Allègre M, Morel J-B, Tharreau D, Nottéghem J-L, et al. Characterization of the model system rice-Magnaporthe for the study of nonhost resistance in cereals. *New Phytol.* 2008; 180: 899–910. doi: [10.1111/j.1469-8137.2008.02621.x](#) PMID: [19138233](#)
92. Delteil A, Blein M, Faivre-rampant O, Guellim A, Estevan J, Hirsch J, et al. Building a mutant resource for the study of disease resistance in rice reveals the pivotal role of several genes involved in defence. *2012;13: 72–82.*
93. Magnan CN, Baldi P. SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics.* 2014; 30: 2592–2597. doi: [10.1093/bioinformatics/btu352](#) PMID: [24860169](#)

Supporting information

Supplemental Table S1. NMR experiments acquired for structure calculations and chemical shift assignments

a) AVR-Pia

Experiments	nuclei	Size			Sweep width (ppm)			Mix(ms)	NS	D1(s)	B ₀ (MHz)
		F3	F2	F1	F3	F2	F1				
¹⁵ N-HSQC	¹ H, ¹⁵ N	1024	128	-	-	-	-	-	16	1	700
HNCO(*)	¹ H, ¹⁵ N, ¹³ C	1024	64	64	12.98	36	17.67	-	8	0.001	500
HNCA(*)	¹ H, ¹⁵ N, ¹³ C	1024	64	80	12.98	36	31.81	-	16	0.1	500
HNCOCACB(*)	¹ H, ¹⁵ N, ¹³ C	1024	64	128	12.98	36	75.02	-	32	0.2	500
HNCACO(*)	¹ H, ¹⁵ N, ¹³ C	1024	64	80	12.98	36	17.67	-	64	0.001	500
HNCACB(*)	¹ H, ¹⁵ N, ¹³ C	1024	64	128	12.98	36	75.02	-	32	0.2	500
CON(**)	¹³ C, ¹⁵ N	1024	200	-	30	36	-	-	4	1	500
CACO(**)	¹³ C, ¹³ C	1024	256	-	30	50.32	-	-	8	1	500
¹⁵ N-NOESY-HSQC	¹ H, ¹⁵ N, ¹ H	1500	76	360	14	35	14	120	8	1	700
¹⁵ N-TOCSY-HSQC	¹ H, ¹⁵ N, ¹ H	1500	70	320	14	35	14	56	8	1	700
NOESY (D ₂ O)	¹ H, ¹ H	2048 4096	512 512	-	15.00 12.01	15.00 12.01	-	100 150	128 128	1 1	700 700
TOCSY (D ₂ O)	¹ H, ¹ H	4096	512	-	12.01	12.01	-	39.6	64	1	700
COSY-dqf (D ₂ O)	¹ H, ¹ H	4096	800	-	12.01	12.01	-	-	64	1	700
¹ H, ¹⁵ N NOE	¹ H, ¹⁵ N	1024	128	-	14	35	-	Sat. 3s	64	6	500
R ₁	¹ H, ¹⁵ N	1024	128	-	14	35	-	-	16	2.5	500
R ₂	¹ H, ¹⁵ N	1024	128	-	14	35	-	-	16	2.5	500

b) AVR1-CO39

Experiments	nuclei	Size			Sweep width (ppm)			Mix(ms)	NS	D1(s)	B ₀ (MHz)
		F3	F2	F1	F3	F2	F1				
¹⁵ N-HSQC	¹ H, ¹⁵ N	1500	160	-	12.02	28	-	-	16	1	700
¹³ C-HSQC	¹ H, ¹³ C	2048	256	-	15.00	80	-	-	512	1.5	700
HNCO(*)	¹ H, ¹⁵ N, ¹³ C	1024	64	64	12.02	28	17.67	-	8	0.001	500
HNCA(*)	¹ H, ¹⁵ N, ¹³ C	1024	64	80	18.02	28	31.81	-	16	0.1	500
HNCOCACB(*)	¹ H, ¹⁵ N, ¹³ C	1024	64	128	18.02	28	75.02	-	32	0.2	500
HNCACO(*)	¹ H, ¹⁵ N, ¹³ C	1024	64	80	18.02	28	17.67	-	64	0.001	500
HNCACB(*)	¹ H, ¹⁵ N, ¹³ C	1024	64	128	18.02	28	75.02	-	32	0.2	500
CON(**)	¹³ C, ¹⁵ N	1024	200	-	30	28	-	-	4	1	500
CACO(**)	¹³ C, ¹³ C	1024	256	-	30	50.32	-	-	2	1	500
¹⁵ N-NOESY-HSQC	¹ H, ¹⁵ N, ¹ H	1500	76	360	12	28	12	120	8	1	700
¹⁵ N-TOCSY-HSQC	¹ H, ¹⁵ N, ¹ H	1024	72	320	14	35	14	56	8	1	700
NOESY (D ₂ O)	¹ H, ¹ H	2048 2048	512 512	-	15.00 15.00	15.00 15.00	-	100 200	64 32	1 1	700 700
TOCSY	¹ H, ¹ H	2048	512	-	15.00	15.00	-	40	32	1	700
COSY-dqf	¹ H, ¹ H	2048	800	-	15.00	15.00	-	-	32	1	700
¹ H, ¹⁵ N NOE	¹ H, ¹⁵ N	1024	128	-	14	35	-	Sat. 3s	64	6	500
R ₁	¹ H, ¹⁵ N	1024	128	-	14	35	-	-	16	2.5	500
R ₂	¹ H, ¹⁵ N	1024	128	-	14	35	-	-	16	2.5	500

Experiments were recorded using the TOPSPIN Library (v. 2.1) at 305 K.

(*) Best pulse sequences

(**)IPAP scheme for virtual decoupling

S2 Table. DALI Statistics for structural alignment of AVR-Pia, AVR1-CO39, AVRPiz-t and ToxB.

RMSD (Å)	AVR-Pia	AVR1-CO39	AVRPiz-t	ToxB
Z-score				
AVR-Pia		2.3	2.8	2.2
AVR1-CO39	2.9		3.0	2.2
AVRPiz-t	3.3	3.1		2.2
ToxB	4.4	4.2	5.4	

The rmsd (upper right) and the Z-score (lower left) are from the pairwise superposition. The low RMSDs and the high Z-scores indicate that the ToxB(b) structure is the closest to the other three structures.

S3 Table MAX-effector candidates identified by psi-Blast in the genomes of the *M. oryzae* isolates 7015 and TH16 and the *M. grisea* isolate BR29.

geneID	localization	AVR-Pia*	AVR1-CO39*	ToxB*
M_BR29_EuGene_00004921	scaffold00001	9,0E-12	-	6,0E-12
M_BR29_EuGene_00041131	scaffold00007	-	9,0E-06	-
M_BR29_EuGene_00043011	scaffold00008	-	9,0E-06	-
M_BR29_EuGene_00060181	scaffold00013	-	4,0E-08	-
M_BR29_EuGene_00081821	scaffold00023	-	3,0E-07	-
M_BR29_EuGene_00082031	scaffold00023	-	1,0E-05	-
M_BR29_EuGene_00085071	scaffold00025	-	-	3,0E-15
M_BR29_EuGene_00087671	scaffold00027	4,0E-13	4,0E-12	8,0E-11
M_BR29_EuGene_00088411	scaffold00027	-	-	5,0E-12
M_BR29_EuGene_00091361	scaffold00030	3,0E-12	-	-
M_BR29_EuGene_00091681	scaffold00031	-	2,0E-05	-
M_BR29_EuGene_00095641	scaffold00035	-	3,0E-13	5,0E-13
M_BR29_EuGene_00106461	scaffold00049	7,0E-23	-	-
M_BR29_EuGene_00107481	scaffold00051	-	2,0E-06	-
M_BR29_EuGene_00112111	scaffold00059	-	6,0E-09	-
M_BR29_EuGene_00113041	scaffold00061	-	7,0E-08	-
M_BR29_EuGene_00118801	scaffold00076	-	2,0E-05	-
M_BR29_EuGene_00119491	scaffold00079	-	3,0E-22	-
M_BR29_EuGene_00119511	scaffold00079	-	9,0E-17	-
M_BR29_EuGene_00121691	scaffold00087	-	1,0E-13	-
M_BR29_EuGene_00125811	scaffold00145	-	7,0E-08	-
M_BR29_EuGene_00126081	scaffold00163	-	-	2,0E-20
MGG_00821	Chromosome_8.5	-	6,0E-14	-
MGG_04384	Chromosome_8.2	-	2,0E-04	-
MGG_08482	Chromosome_8.4	-	1,0E-06	-
MGG_08944	Chromosome_8.2	-	-	5,0E-09
MGG_10120	Chromosome_8.4	-	5,0E-04	-
MGG_14793	Chromosome_8.2	4,0E-19	3,0E-06	9,0E-15
MGG_14834	Chromosome_8.4	-	2,0E-05	-
MGG_15207	Chromosome_8.3	-	5,0E-08	-
MGG_15459	Chromosome_8.1	-	8,0E-08	-
MGG_16058	Chromosome_8.1	-	1,0E-05	-
MGG_16113	Chromosome_8.1	-	3,0E-09	-
MGG_16175	Chromosome_8.1	-	5,0E-08	-
MGG_16619	Chromosome_8.3	-	3,0E-06	-
MGG_17132	Chromosome_8.4	2,0E-06	7,0E-04	9,0E-10
MGG_17255	Chromosome_8.4	-	1,0E-06	-
MGG_18019	Chromosome_8.7	-	7,0E-05	8,0E-11
MGG_18060	Chromosome_8.7	5,0E-14	5,0E-14	4,0E-06
M_TH16_EuGene_00000541	scaffold00001	-	1,0E-14	-
M_TH16_EuGene_00027191	scaffold00004	4,0E-10	-	-
M_TH16_EuGene_00027411	scaffold00004	4,0E-12	-	-
M_TH16_EuGene_00034081	scaffold00004	-	3,0E-08	-
M_TH16_EuGene_00040131	scaffold00005	2,0E-25	-	-
M_TH16_EuGene_00045871	scaffold00007	-	3,0E-12	-
M_TH16_EuGene_00079081	scaffold00016	6,0E-10	-	-
M_TH16_EuGene_00079311	scaffold00016	-	2,0E-08	-
M_TH16_EuGene_00099371	scaffold00026	-	3,0E-08	-
M_TH16_EuGene_00101881	scaffold00028	-	2,0E-05	2,0E-08
M_TH16_EuGene_00106621	scaffold00033	-	-	5,0E-14
M_TH16_EuGene_00120731	scaffold00052	3,0E-06	-	-
M_TH16_EuGene_00124981	scaffold00063	-	7,0E-11	-
M_TH16_EuGene_00127871	scaffold00072	-	2,0E-21	-
M_TH16_EuGene_00134971	scaffold00110	-	9,0E-14	1,0E-14
M_TH16_EuGene_00135161	scaffold00112	-	-	2,0E-19

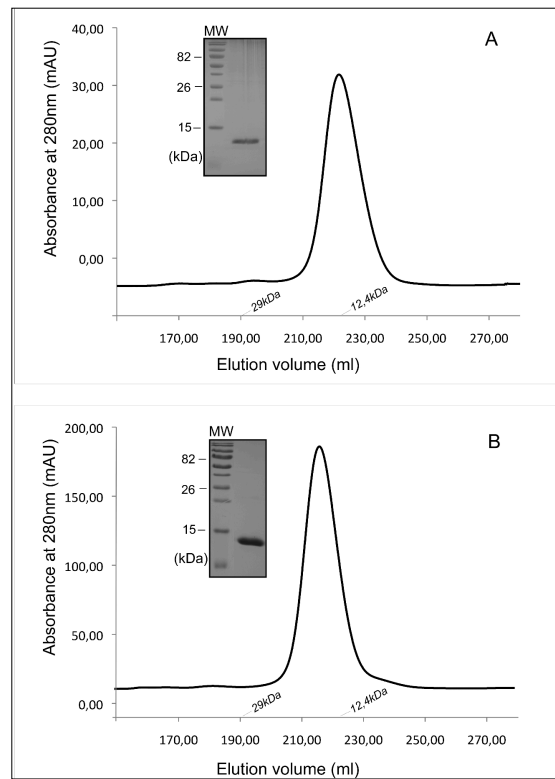
Orthologs of 7015 proteins in TH16 are not listed.

* Values correspond to e-values in psi-blast search with the corresponding protein

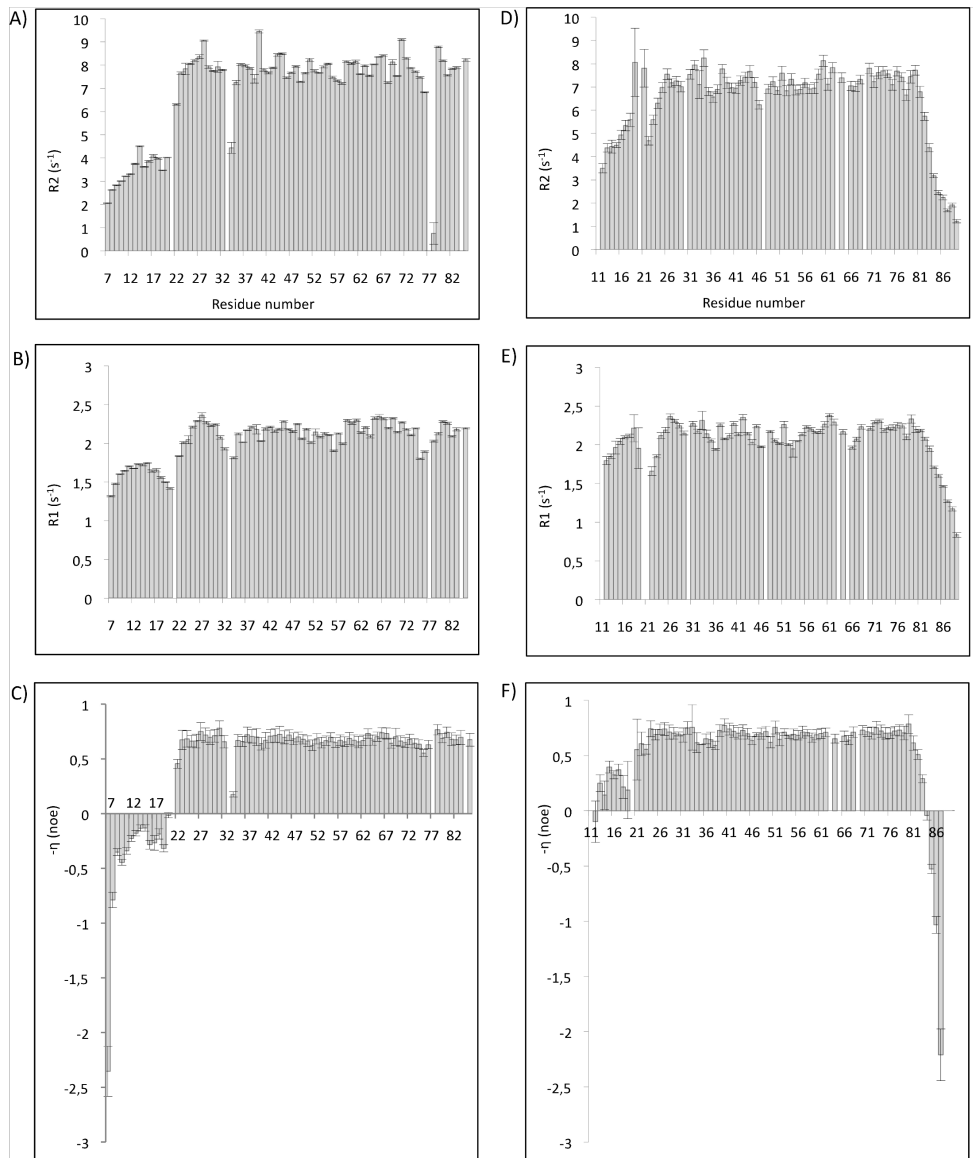
S 4 Table Primers used for qRT-PCR

GENE	Forward	Reverse
Infection specific expression		
MGG_00954	GTACCCAAGCCCTGGAAAGT	CCAGGACAATGCTCAAATCA
MGG_02207	GTGCCAATTGGTGGCTTACT	TCTTTCGCTTAGCGTCGTAAC
MGG_02546	AAGCTACGCCGTTTGTGAAC	CGGGTAATTTCTCGCTGTCA
MGG_04735	GCAAAGTCGTACCCGAGATT	CGTCCATTGTCCCTCAATTA
MGG_05896	GACGTTTGCTGGGTTAAAG	GGCAAGTGTAGCCGTTGATAA
MGG_08414	CGTCGTGGGTGATTGAAC	GTACCAATTCCACGCCAGTC
MGG_08482	ACCATGCAAGGTCACGATTT	TGATATGTGTGGCCGGTAGA
MGG_08944	CCTGACCAATGGCAACAAG	AAATTCGCACCTTGGAGTTCA
MGG_08992	AAGAGACGGGAACGTTGATG	ATCCACGTCGTACCCGTTAT
MGG_09675	CCGTCAAAGAGGAAGGATTC	TTTGGCAGAATGCAATTCAC
MGG_11967	ATGGGCTCAAAGCTTCTGTC	TGGATTCTTGGTCCAGTT
MGG_13133	AGTCGAAAGGTTGGGAGAT	ATCGGATTGCACTCCTTGTC
MGG_15207	CCCACCTCAATCCTGAGAGA	TATCTCGCTCCACGGTTAC
MGG_15625	CAGCAAAGAAGGTGGGAAGA	GCCGATACACATTTGCACTG
MGG_15972	TCGAGAGCGAGACCCTAACT	CCCAGTTTGGACCAACTTTC
MGG_16175	CGATTGCACACTTGGTTGTAA	TGCATCCTCTCCGATCTTT
MGG_16619	GAAGCTAGGCGATCGTTCTG	GCAATCCTGTTCCACATGAA
MGG_16939	ATTGCTGGCCAAGGTTGAG	CGGCAGTTTCTGAACATGAG
MGG_17266	AAACACCAGCTCCCATCAAC	GCTGGAAGTGGGACTCTG
MGG_18019	TGCCTTCCAGCTACAGAGT	CTTTCGGAACGCTCTTCTTG
MGG_18062	CAACAGCAGCTTCGATATCAT	CCAACTTGGCAAGCAATTC
M_TH16_00000541	GGAAATCAGCGTTTCAAGT	CAAACCTCCAGCTCGTGACA
M_TH16_00027411	AGAGTCGAAGCAGGCGAAAT	TTCGCGGAGACTGTACATGA
M_TH16_00034081	AATCAGACGCTGAGCAAGT	GCGATCGCAGGTAATGTAGC
M_TH16_00040131	GCTTGGCTGGGTTAAGATTG	TTTGCAGTTGGTGTCTCTG
M_TH16_00079081	GGTTGGTGCAAGGTCAAAGT	CCGTGAGATTCCAATGAGT
M_TH16_00079311	AAAGCTGCATGGCAAGACT	AACCTCCGATTGTTCCACTG
M_TH16_00104561	CGTCGTCGATTCAGCAATA	GCTGCTTCTTATCCACAA
M_TH16_00119711	CAGCGGTGTAGTGCAATAA	TCCTTCCAGCTTTCACAGT
M_TH16_00120731	ACCGTGGGAATGGGTTACTT	TAAACTCCCGGCTGGATAAG
M_TH16_00124981	GCTTGGGTTGCACCTACAGT	TTCAAACCTCGCAGGTGTCTG
M_TH16_00127871	GTTGCCTGTGCAGCCTATCT	GATTCGGCTGGAATCATCTC
M_TH16_00136331	AAGGAAGAAGTCGAGGGTGAG	AACGATATCCAACCCGTCAG
Constitutive expression and controls		
MGG_14793	ACAGCCGTCTGCGACTTTAT	GTCCGCTCAGGCTAAGTTTG
MGG_07184	AGCGTGTCTGCAAAGCTGTA	AGACCTCCCAACGGTTCTCT
Actin, MGG_03982	TCTTCGAGACCTTCAACGCC	ACCGGAGTCGAGCAGGATAC
EF1 α , MGG_03641	GCC CGG TAT GGT CGT TAC CT	AGC TGC TGG TGG TGC ATC TC
ORF3, MGG_08381	GGTGAGGTTGGAGGTAGTG	TGGAGTGGCCAACATGTC
MGG_01147	CGACGACCTACCTGCTGACT	CAATGCTCCTTCTCTGGAG
BAS3, MGG_11610	CCCGTGTGGAGGAATTGTG	CTTGAGGTTGTCGGTGCTCT
No expression		
MGG_00821	GCGGCTACACGATTGAGAC	ACCTCGTGCTATTCCTGAC
MGG_08469	CCAGACGTACTCCCACTTG	AGGCCCTCAATTTCTTCTC
MGG_08607	ACGCAACAACCGAAGAATC	GGTCCAACCGGAAGTATAA
MGG_09425	TAGCCAGGAAGGCACCTTACA	GCGATCAAATCGAACCTTGA
MGG_10335	ATCCTTGTAGGCGCAAAGTG	GCAGGTTAATGCCCTTGACA
MGG_12426	GGCAGGAGAGTCACCTTCAG	CGGAGATGATGCAACGTTTA
MGG_15106	TTTGGATCACCGGGAATAC	CTCGCCCAATAAACTTTC
MGG_15575	AGGACTTGCCAAAGGGATTT	GCGCAAGAAGTCAATCAATC
MGG_16113	GGGTTATGGCAGGTTGTTA	GCCTGAGGCAGTTTCAGTTC
MGG_16416	GCAACTCATTCCGCCACAGTA	GCAATGCCATAAACAGTCT
MGG_17132	AAATGCAATTCAAATCCCTCTT	ACCATCTCCTCCGATTCTT
MGG_17255	AGTCGTGCACTGGCCTCAG	GCCACCAATGACGTTCTTT
MGG_17799	TTGGCACGAAATGTGCTATT	TCGTAGTCCAACCGGACTGT
MGG_18041	GGGTTGGGCCGTTAGATT	CCAAATTTGGTGCATTTCAGA
MGG_18060	GGGTGCCATATTCCTGAG	TGATCGGCTGAGATTGTAGC
M_TH16_EuGene_00099371	TCCTGGCTTCGTTCAATAAGA	TCGAATGATAGTGGGGTAAC
M_TH16_EuGene_00101881	GTGCAAGGTCGAGATGTTTG	CGGACCAAGTTCTGGGTAAG
M_TH16_EuGene_00106621	TGCAAATTCACGTCCTCAA	GTGAAAGGGCTGACGGTAAC

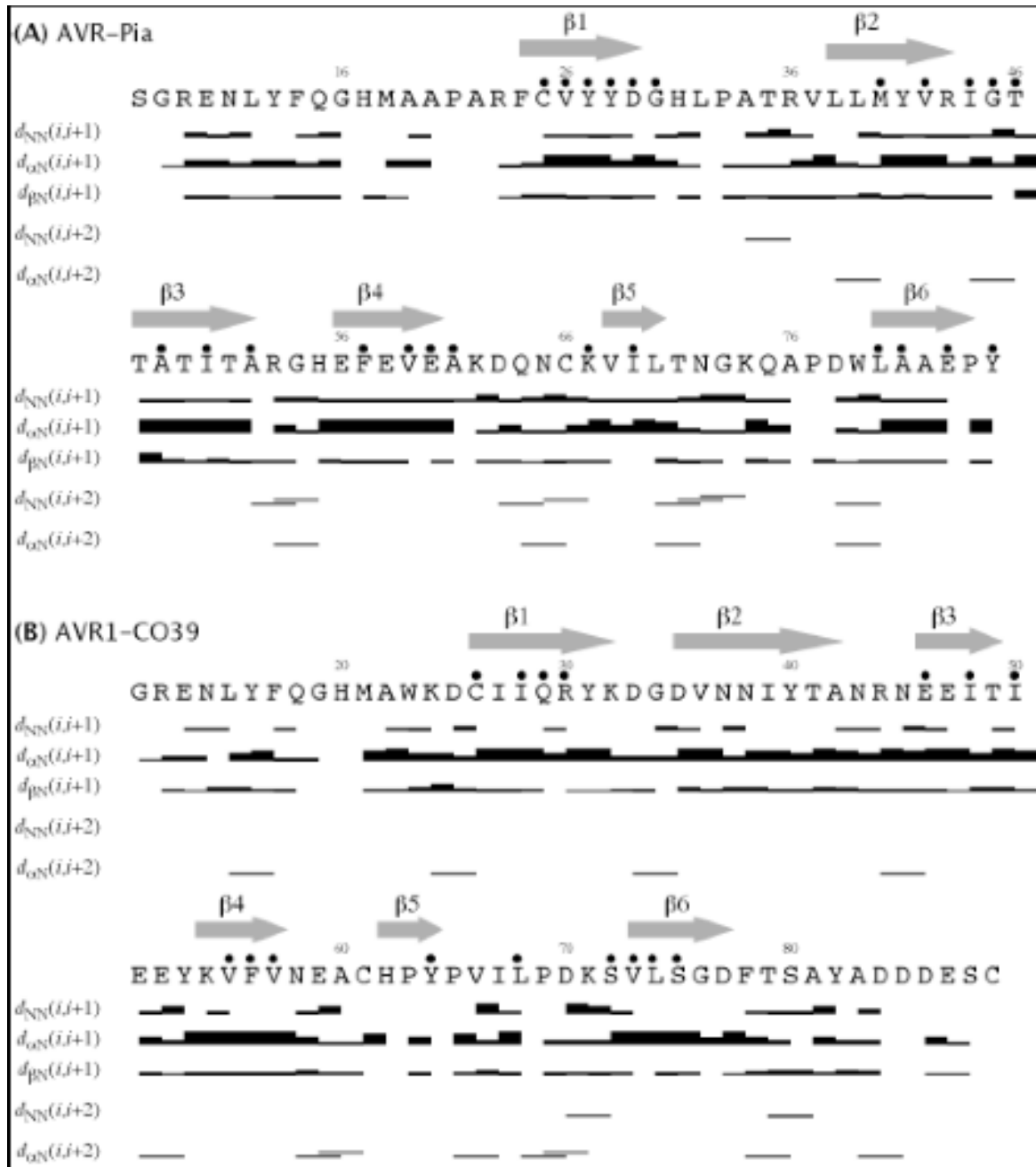
S1 Fig. Gel filtration profile and SDS-PAGE analysis of purified AVR-Pia (A) and AVR1-CO39 (B) proteins.



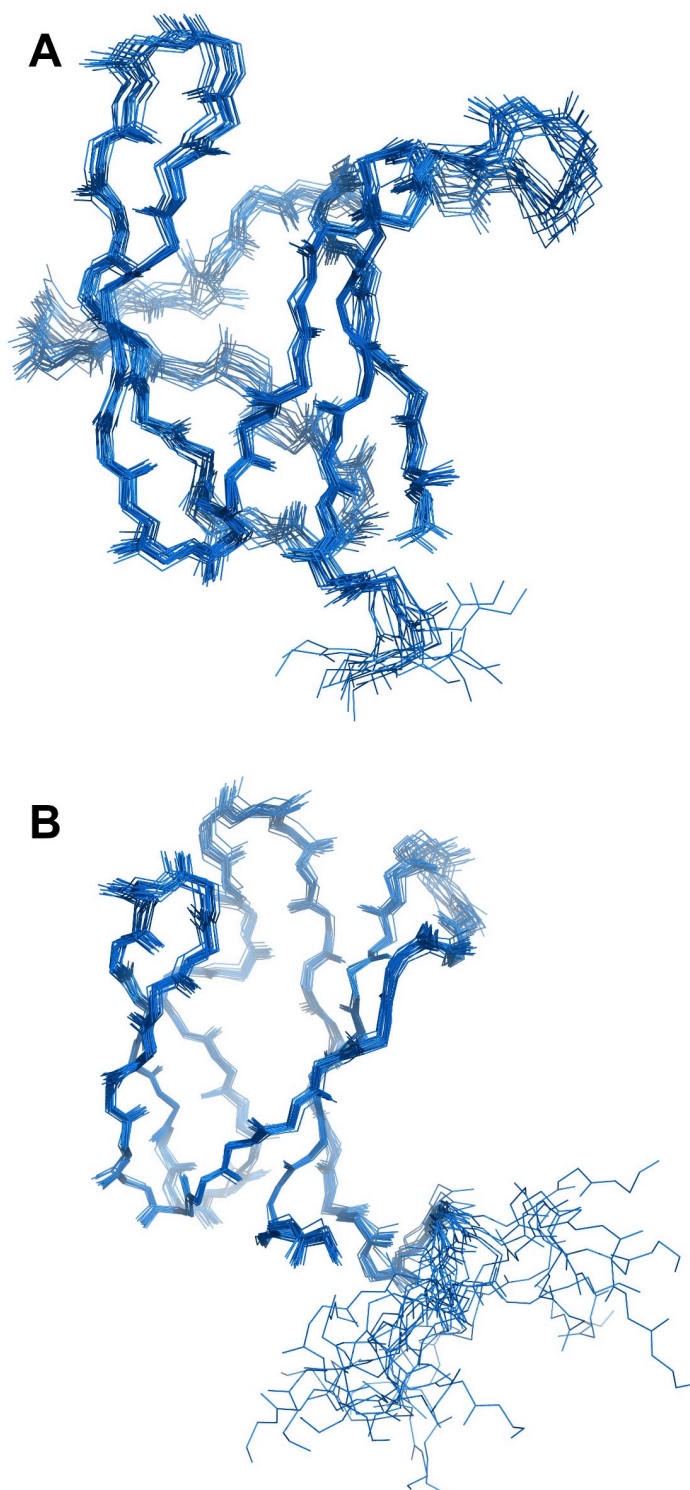
S2 Fig. 15N Relaxation data at 500 MHz and 305K for AVR-Pia (panels A, B and C) and AVR1-CO39 (panels D, E and F).



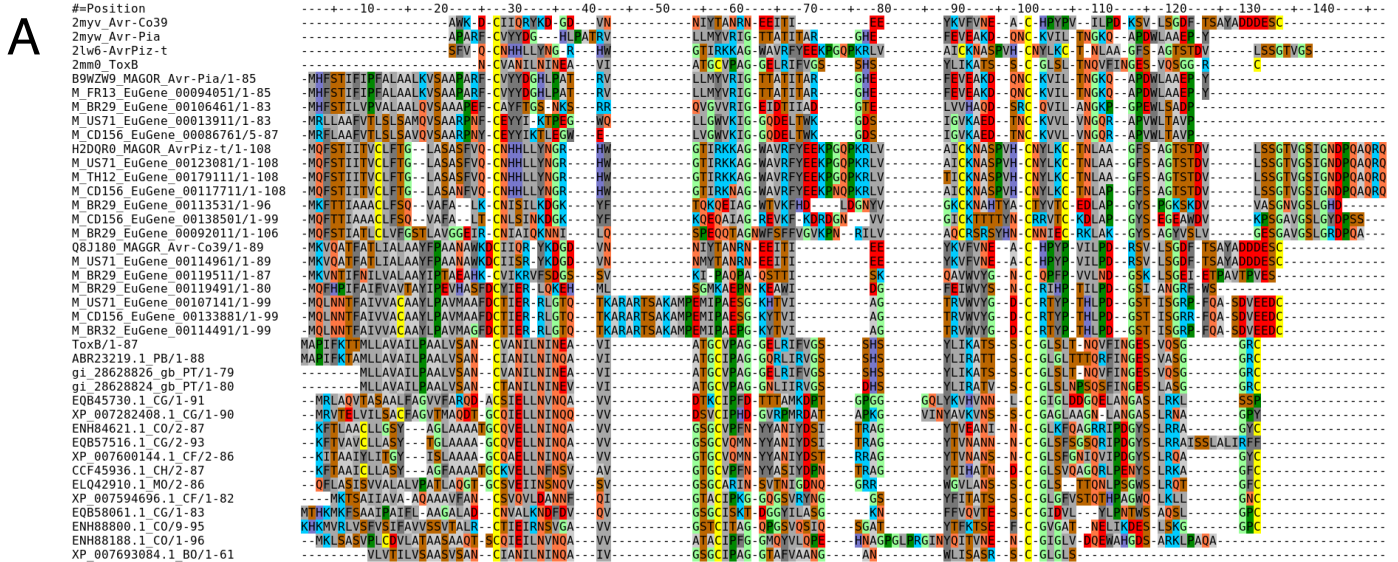
S3 Fig. Backbone sequential and medium range NOEs observed for (A) AVR-Pia and (B) AVR1-CO39. The line width is proportional to the NOE intensity. The dots (•) indicate slow exchange NH observed in 2D-NOESY in D₂O. Grey arrows indicate the β -strands determined from the structure analysis (vide infra).



S4 Fig. Solution structures of (A) AVR-Pia and (B) AVR1-CO39. Superposition of the backbone atoms of the 20 lowest energy conformers used to calculate the final structures. Only mature chains are shown, from residues Ala20 and Trp23 for AVR-Pia and AVR1-CO39, respectively.

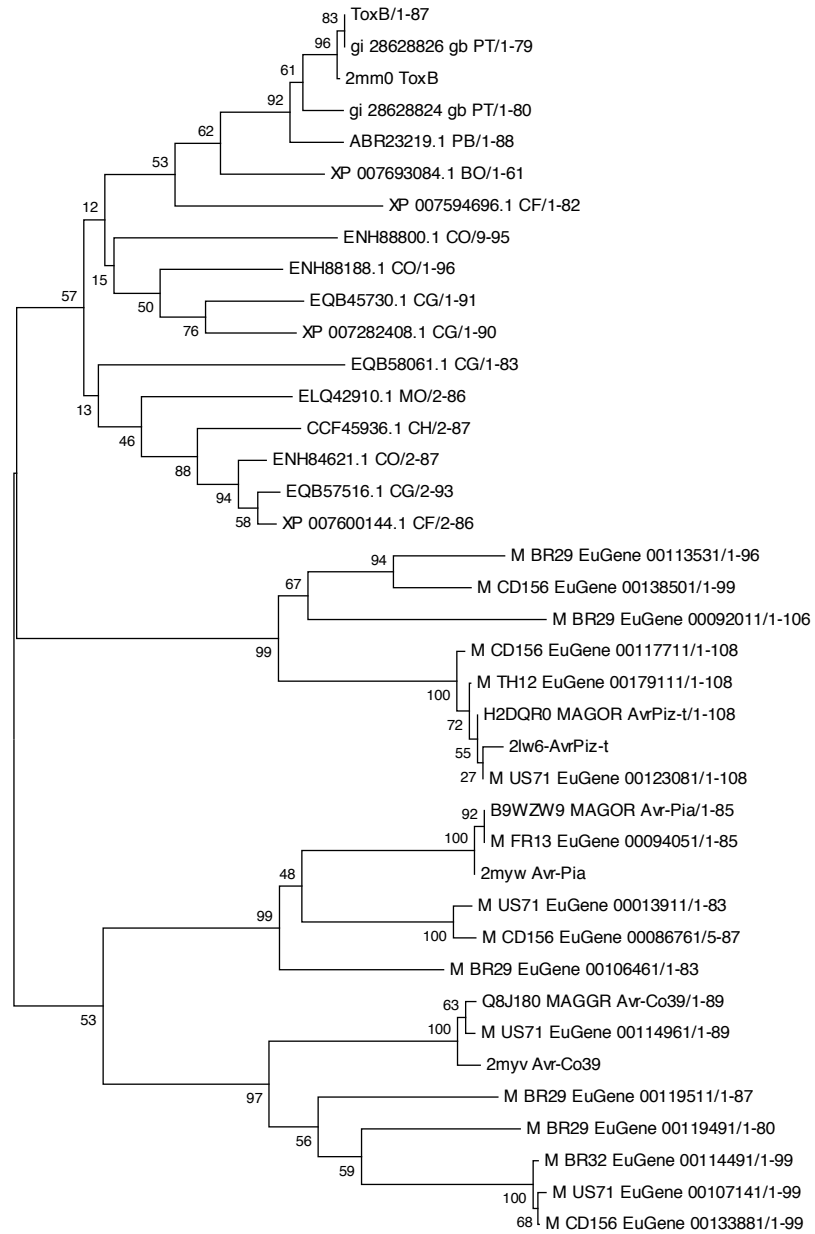


S5 Fig. Structure-guided alignment and diversity of MAX-effector homologs identified by Blast. A) Homologs of AVR1-CO39, AvrPiz-t and AVR-Pia identified by Blast in *M. oryzae* and *M. grisea* genomes and ToxB homologs identified by Psi-Blast in the GeneBank database were aligned to the structural alignment of mature ToxB, AVR1-CO39, AvrPiz-t and AVR-Pia. (B) A diversity tree was constructed by the neighbor-joining method using the alignment in (A). It highlights the high diversity of MAX-effector homologs. Branch supports are based on 1000 bootstraps and horizontal branch length reflects sequence divergence. Accession numbers of non-Magnaporthe sequences were completed by a 2 letter identifier for the species: BO for *Bipolaris oryzae*, CF is for *Colletotrichum fioriniae*, CH for *C. higginsianum*, CG for *C. gloeosporioides*, CO for *C. r biculare*, LM for *Leptosphaeria maculans*, PT for *Pyrenophora tritici-repentis* and PB for *Pyrenophora bromi*.

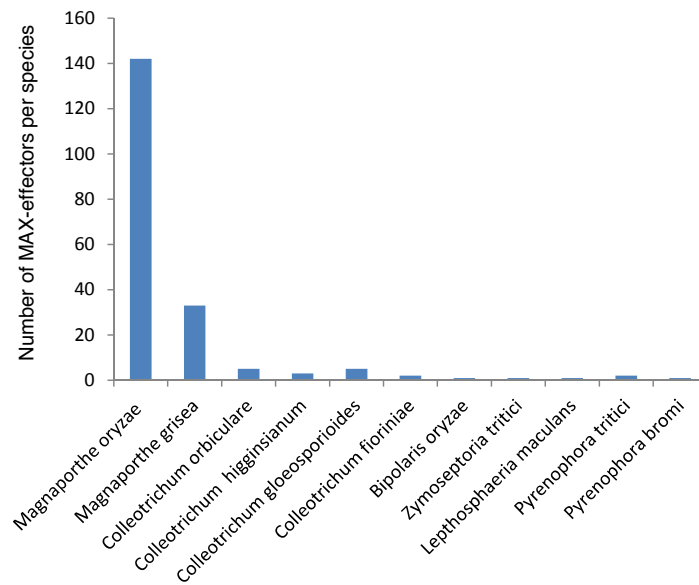


S5 Fig. continued

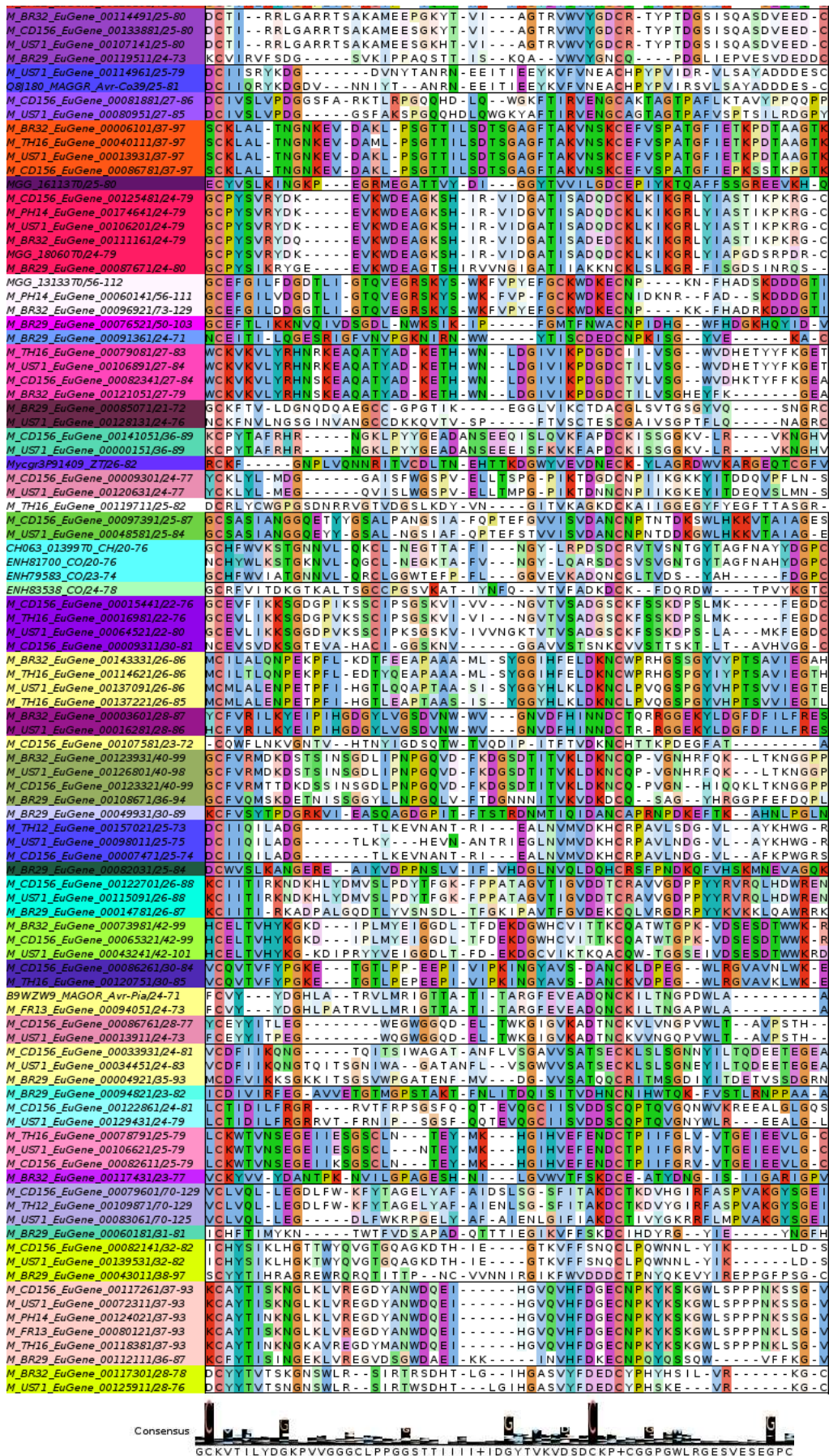
B



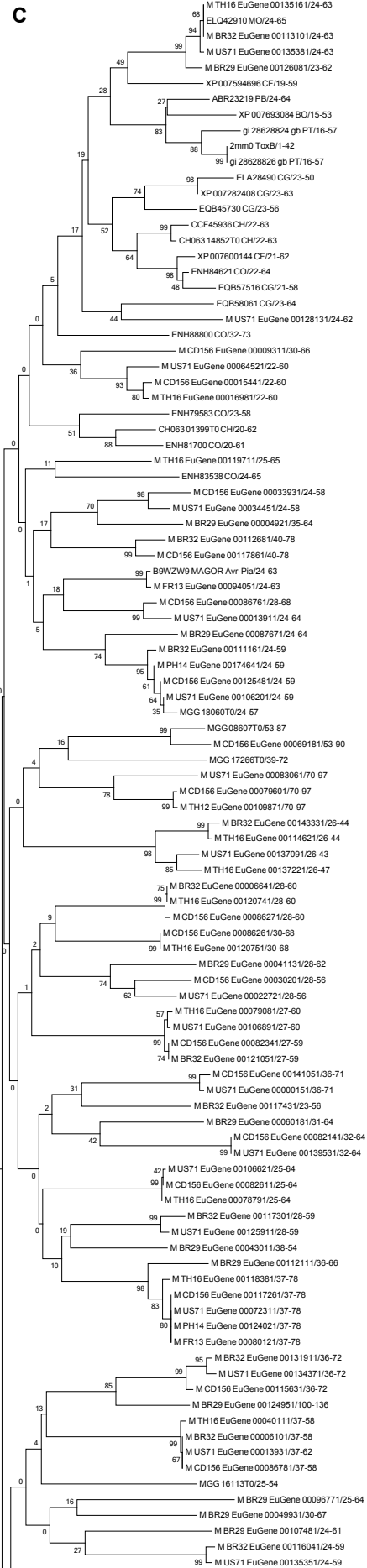
S6 Fig. MAX-effector homologs identified by a high stringency HMM search. (A) Histogram showing the numbers of MAX-effectors identified by an HMM pattern search in a nonredundant database comprising the small secreted proteins of 25 ascomycete fungi and of 8 additional *M. oryzae* and one *M. grisea* isolate. (B) MAX-effectors were aligned to the structural alignment of mature ToxB, AVR1-CO39, AvrPiz-t and AVR-Pia and gaps were removed. (C) A diversity tree was constructed by the neighbor-joining method using the alignment in (B). Branch supports are based on 1000 bootstraps and horizontal branch length reflects sequence divergence. Accession numbers of non-Magnaporthe sequences were completed by a 2 letter identifier for the species: BO for *Bipolaris oryzae*, CF for *Colletotrichum fioriniae*, CH for *C. higginsianum*, CG for *C. gloeosporioides*, CO for *C. orbiculare*, LM for *Lepthosphaeria maculans*, PT for *Pyrenophora tritici-repentis* PB for *Pyrenophora bromi* and ZT for *Zymoseptoria tritici*.

A

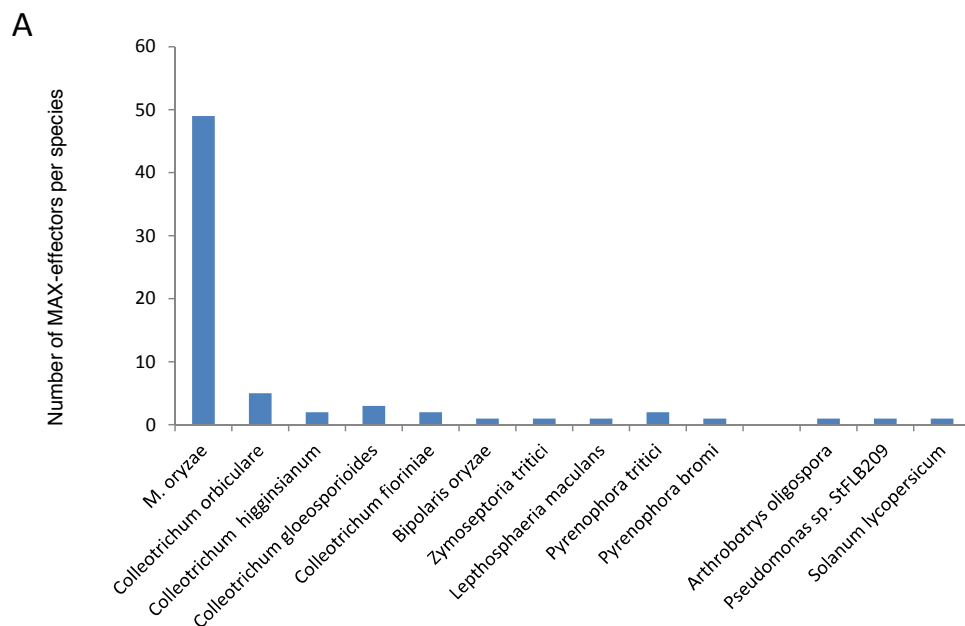
S6 Fig.B continued



S6 Fig. continued

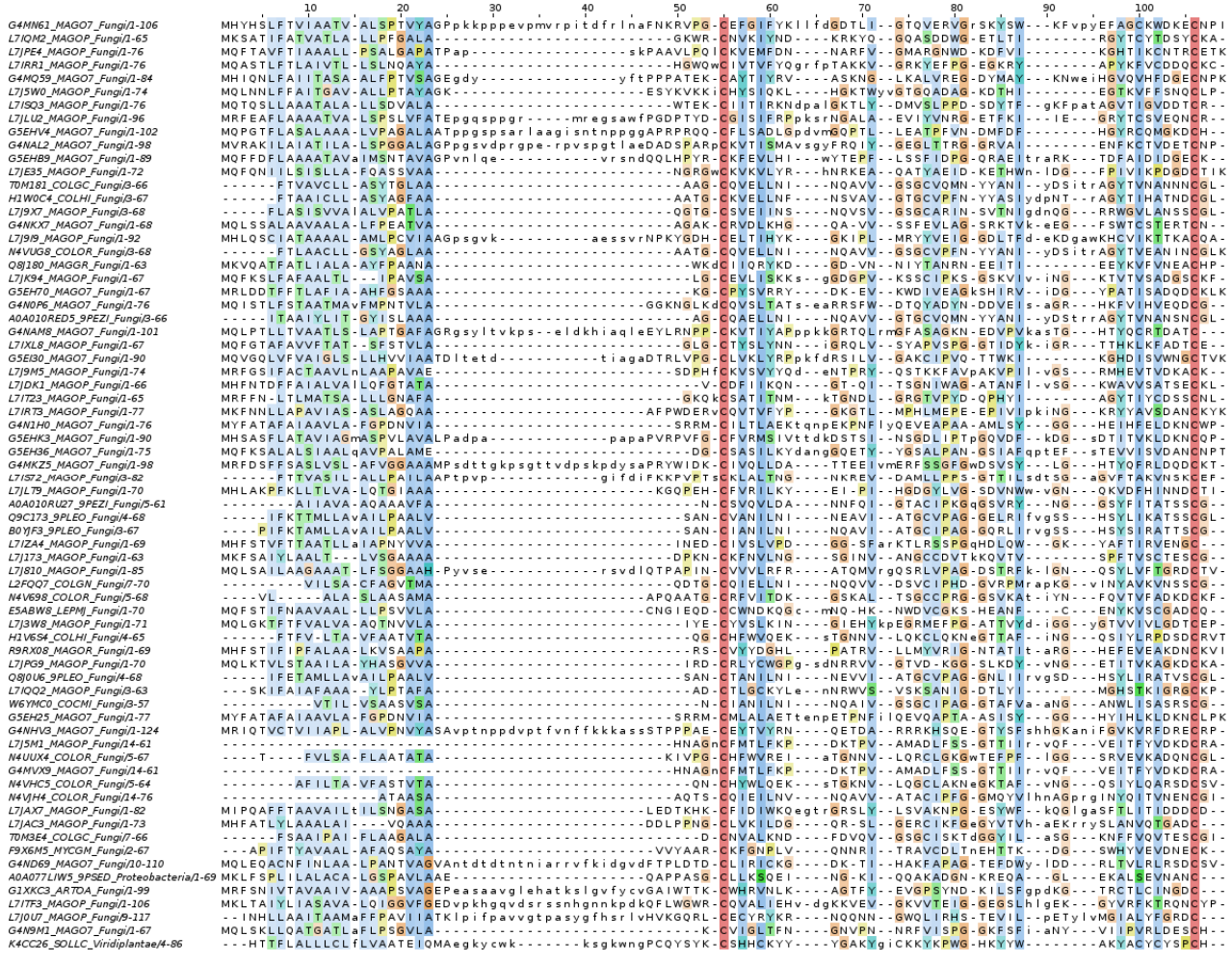


S7 Fig. MAX-effector homologs identified in the UniRef90 database by a low stringencyHMM search. (A) Histogram showing the numbers of MAX-effectors identified by an HMM pattern search in a non-redundant UniRef90 database. (B) MAX-effectors identified by HMMpattern search were aligned to the structural alignment of mature ToxB, AVR1-CO39, AvrPizt and AVR-Pia. (C) A diversity tree was constructed by the neighbor-joining method using the alignment in (B). This highlights the high diversity of MAX-effector homologs. Branch supports are based on 1000 bootstraps and horizontal branch length reflects sequence divergence. Accession numbers contain the following information on the species: MAGGR, MAGO7, MAGOP and MAGOR are from *M. oryzae*, COLGC and COLGN from *C. gloeosporioides*, COLHI from *Colletotrichum higginsianum*, 9PEZI from *C. fioriniae* and COLOR from *Colletotrichum orbiculare*, 9PLEO from *P. tritici-repentis* or *P. bromi*, ARTOA from *Arthrobotrys oligospora*, COCMI from *Bipolaris oryzae*, LEPMJ from *Leptosphaeria maculans*, MYCGM from *Zymoseptoria tritici*, 9PSED from *Pseudomonas sp. StFLB209* and SOLLC from *Solanum lycopersicum*.



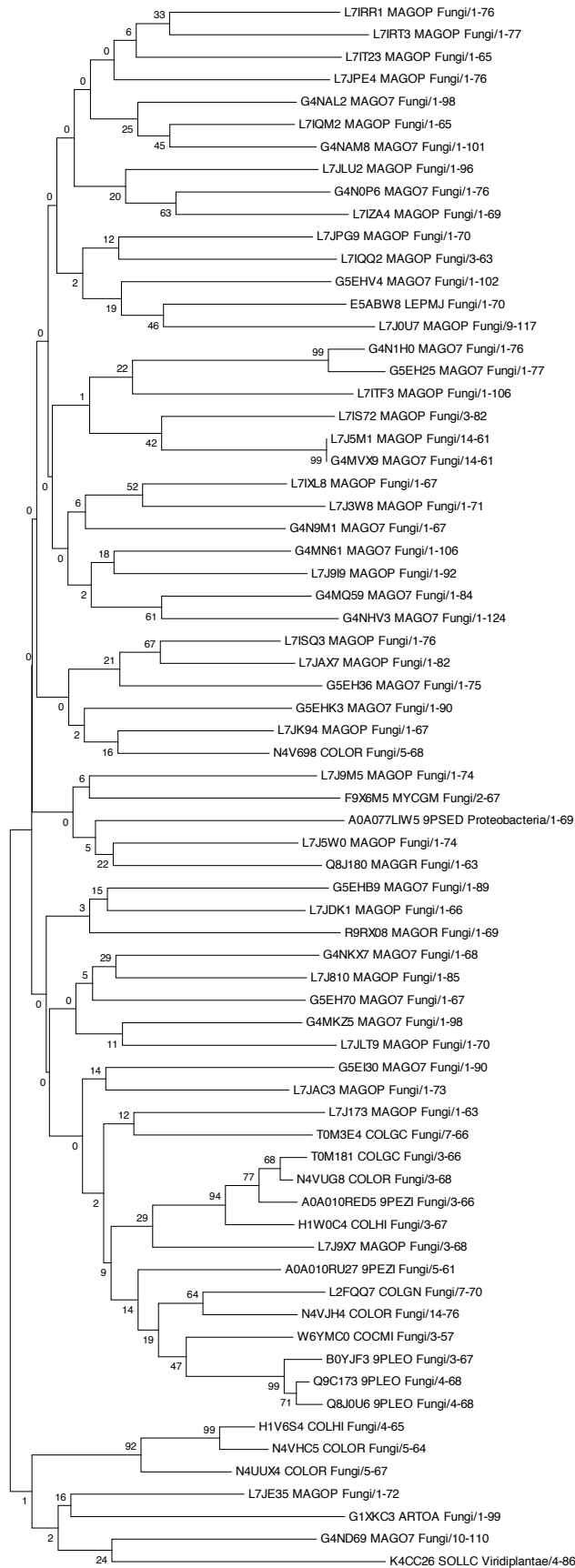
S7 Fig. continued

B



S7 Fig. continued

C



S8 Fig.A continued

M_BR32_EuGene_00143331/1-76	MYFATAFAIAAVLAFGPDNVIA---SRRMCILALAEKKEPNFEEVEAPAAAMLSY-GGHEIHFELDKNCWF
MGG_1693970/1-76	MYFATAFAIAAVLAFGPDNVIA---SRRMCILTLAEKKEPNFEEVEAPAAAMLSY-GGHEIHFELDKNCWF
M_FR13_EuGene_00096711/1-76	MYFATAFAIAAVLAFGPDNVIA---SRRMCILTLAEKKEPNFEEVEAPAAAMLSY-GGHEIHFELDKNCWF
M_TH16_EuGene_00114621/1-76	MYFATAFAIAAVLAFGPDNVIA---SRRMCILTLAEKKEPNFEEVEAPAAAMLSY-GGHEIHFELDKNCWF
M_TH12_EuGene_00151361/1-76	MYFATAFAIAAVLAFGPDNVIA---SRRMCILTLAEKKEPNFEEVEAPAAAMLSY-GGHEIHFELDKNCWF
M_GY11_EuGene_00073911/1-76	MYFATAFAIAAVLAFGPDNVIA---SRRMCILTLAEKKEPNFEEVEAPAAAMLSY-GGHEIHFELDKNCWF
M_US11_EuGene_00137091/1-76	MYFATAFAIAAVLAFGPDNVIA---SRRMCMLALAEETPNFEOEQARTAAISY-GGHYIHLKLDKNCWF
M_TH16_EuGene_00137221/1-76	MYFATAFAIAAVLAFGPDNVIA---SRRMCMLALAEETPNFEOEQARTAAISY-GGHYIHLKLDKNCWF
M_TH12_EuGene_00186401/1-76	MYFATAFAIAAVLAFGPDNVIA---SRRMCMLALAEETPNFEOEQARTAAISY-GGHYIHLKLDKNCWF
MGG_1806270/1-76	MYFATAFAIAAVLAFGPDNVIA---SRRMCMLALAEETPNFEOEQARTAAISY-GGHYIHLKLDKNCWF
M_BR29_EuGene_00117501/1-76	MRLDVFVLIATVLAFFGPNVIA---EVRICSLALGKING-EVLQQLPGPAGNVVLEDFSLQLSLDKNCWF
M_TH16_EuGene_00118821/1-76	MHFTNKIGVAVILAFGWGSAIV---SRRPCFVALSKDGEELWGGVHEIIGGVEVTMAHYKFKIKLDGNCWF
M_TH12_EuGene_00148581/1-76	MHFTNKIGVAVILAFGWGSAIV---SRRPCFVALSKDGEELWGGVHEIIGGVEVTMAHYKFKIKLDGNCWF
M_PH14_EuGene_00129441/1-76	MHFTNKIGVAVILAFGWGSAIV---SRRPCFVALSKDGEELWGGVHEIIGGVEVTMAHYKFKIKLDGNCWF
M_GY11_EuGene_00077511/1-76	MHFTNKIGVAVILAFGWGSAIV---SRRPCFVALSKDGEELWGGVHEIIGGVEVTMAHYKFKIKLDGNCWF
M_FR13_EuGene_00079141/1-76	MHFTNKIGVAVILAFGWGSAIV---SRRPCFVALSKDGEELWGGVHEIIGGVEVTMAHYKFKIKLDGNCWF
MGG_3724970/1-76	MHFTNKIGVAVILAFGWGSAIV---SRRPCFVALSKDGEELWGGVHEIIGGVEVTMAHYKFKIKLDGNCWF
M_US11_EuGene_00022721/1-76	MRFGSIFACTAAVLLAARPAVAE---SDPHCKVSVYYQNTPRYQSTKKFAVAVKPIVIGSRMHVEVTVDKACK-
M_TH16_EuGene_00019681/1-76	MRFGSIFACTAAVLLAARPAVAE---SDPHCKVSVYYQNTPRYQSTKKFAVAVKPIVIGSRMHVEVTVDKACK-
M_TH12_EuGene_00025641/1-76	MRFGSIFACTAAVLLAARPAVAE---SDPHCKVSVYYQNTPRYQSTKKFAVAVKPIVIGSRMHVEVTVDKACK-
M_PH14_EuGene_00076101/1-76	MRFGSIFACTAAVLLAARPAVAE---SDPHCKVSVYYQNTPRYQSTKKFAVAVKPIVIGSRMHVEVTVDKACK-
M_GY11_EuGene_00019701/1-76	MRFGSIFACTAAVLLAARPAVAE---SDPHCKVSVYYQNTPRYQSTKKFAVAVKPIVIGSRMHVEVTVDKACK-
M_FR13_EuGene_00013421/1-76	MRFGSIFACTAAVLLAARPAVAE---SDPHCKVSVYYQNTPRYQSTKKFAVAVKPIVIGSRMHVEVTVDKACK-
MGG_0082170/1-76	MRFGSIFACTAAVLLAARPAVAE---SDPHCKVSVYYQNTPRYQSTKKFAVAVKPIVIGSRMHVEVTVDKACK-
M_CD156_EuGene_00030201/1-76	MRFGSIFACTAAVLLAARPAVAE---SDPHCKVSVYYQNTPRYQSTKKFAVAVKPIVIGSRMHVEVTVDKACK-
M_BR32_EuGene_00124651/1-76	MRFGSIFACTAAVLLAARPAVAE---SDPHCKVSVYYQNTPRYQSTKKFAVAVKPIVIGSRMHVEVTVDKACK-
M_BR29_EuGene_00041131/1-76	MQLGSLIFETLAVLFLAIPFLVD---SNPHCEVSVYENKQKLIAYVYVPPGSDIVGSGMVAATVDEKACK-
M_BR29_EuGene_00114891/1-76	MRFHALLFLLGVAAAPPVSLVE---DGLCKVTVKCKREKLIEMKPKVHVSKEFKIDRTSFLFVNGCGR-
M_CD156_EuGene_00086271/1-76	MQASTLFLTAIVTLLSLNOAYA---HGWCIVTVFYQAKKVGRRKYEFPGEGRY---APYKFCVDDCKCK-
M_TH16_EuGene_00120741/1-76	MQASTLFLTAIVTLLSLNOAYA---HGWCIVTVFYQAKKVGRRKYEFPGEGRY---APYKFCVDDCKCK-
M_TH12_EuGene_00159981/1-76	MQASTLFLTAIVTLLSLNOAYA---HGWCIVTVFYQAKKVGRRKYEFPGEGRY---APYKFCVDDCKCK-
M_PH14_EuGene_00008071/1-76	MQASTLFLTAIVTLLSLNOAYA---HGWCIVTVFYQAKKVGRRKYEFPGEGRY---APYKFCVDDCKCK-
M_GY11_EuGene_00110061/1-76	MQASTLFLTAIVTLLSLNOAYA---HGWCIVTVFYQAKKVGRRKYEFPGEGRY---APYKFCVDDCKCK-
M_FR13_EuGene_00057131/1-76	MQASTLFLTAIVTLLSLNOAYA---HGWCIVTVFYQAKKVGRRKYEFPGEGRY---APYKFCVDDCKCK-
MGG_1562370/1-76	MQASTLFLTAIVTLLSLNOAYA---HGWCIVTVFYQAKKVGRRKYEFPGEGRY---APYKFCVDDCKCK-
M_BR32_EuGene_00006431/1-76	MQASTLFLTAIVTLLSLNOAYA---HGWCIVTVFYQAKKVGRRKYEFPGEGRY---APYKFCVDDCKCK-
M_BR29_EuGene_00004011/1-76	MQMPTLSAPILIAFLSLSEVVA---WEGTVTVFYQIAKVGIVLVSFPGEGRY---ADYQVILDEDCGR-
M_CD156_EuGene_00133981/1-73	MQLNNTFAIVVACAAYLPVAVMA---AFDCTIERRLGTQTKAEMIAPESGKHVTI---AGTRVWVYGDGR-
M_US11_EuGene_00107141/1-73	MQLNNTFAIVVACAAYLPVAVMA---AFDCTIERRLGTQTKAEMIAPESGKHVTI---AGTRVWVYGDGR-
M_TH16_EuGene_00127871/1-73	MQLNNTFAIVVACAAYLPVAVMA---AFDCTIERRLGTQTKAEMIAPESGKHVTI---AGTRVWVYGDGR-
M_TH12_EuGene_00148801/1-73	MQLNNTFAIVVACAAYLPVAVMA---AFDCTIERRLGTQTKAEMIAPESGKHVTI---AGTRVWVYGDGR-
M_PH14_EuGene_00129241/1-73	MQLNNTFAIVVACAAYLPVAVMA---AFDCTIERRLGTQTKAEMIAPESGKHVTI---AGTRVWVYGDGR-
M_GY11_EuGene_00077711/1-73	MQLNNTFAIVVACAAYLPVAVMA---AFDCTIERRLGTQTKAEMIAPESGKHVTI---AGTRVWVYGDGR-
M_FR13_EuGene_00079381/1-73	MQLNNTFAIVVACAAYLPVAVMA---AFDCTIERRLGTQTKAEMIAPESGKHVTI---AGTRVWVYGDGR-
M_BR32_EuGene_00114491/1-73	MQLNNTFAIVVACAAYLPVAVMA---AFDCTIERRLGTQTKAEMIAPESGKHVTI---AGTRVWVYGDGR-
M_BR29_EuGene_00119491/1-63	MQCHPFAIFVAVTAYIPEVHA---DFDCTIERRLGKEH-MLQGMKAEPMKKAEWI---DGFELWVYNSCR-
M_CD156_EuGene_00119511/1-62	MKVNITFNLVLAALAYIPTAEHA---HKCVIKRVFSDG-SSVKIPAPASTIIT---EYKAWVYNSCR-
M_US11_EuGene_00114961/1-63	MKQVATFALILAAAYPAAINA---WKCILSRKQGD-VHNMXYTARNNEETI---EYKAVFVNEACK-
M_CD156_EuGene_00072751/1-67	MMYSKIFAAILLAAALPTAFVA---EYKCHTIGCAFEGOWHYVVTGNIIGDTIHI---NDHSTVILGDCCK-
MGG_1617570/3-63	---SKIFAIAFAANA-YLPTAFVA---ADCTIGCKVLRNRWVSVKSANIGDTLYI---MGHSTKIRGRCCK-
M_CD156_EuGene_00121391/2-64	---YKRTFAILAFAVAYLPTAFVA---ADCTIGCKVLRNRWVSVKSANIGDTLYI---MGHSTKIRGRCCK-
M_BR32_EuGene_00135751/2-64	MQFGTAFVVFVAT-SFSTVLA---GLGCTYSLYNNGRQLVSVAPVSPGTDIDYIGRTHLTKLKFADTCE-
M_TH16_EuGene_001781/1-67	MQFGTAFVVFVAT-SFSTVLA---GLGCTYSLYNNGRQLVSVAPVSPGTDIDYIGRTHLTKLKFADTCE-
M_TH12_EuGene_00172291/1-67	MQFGTAFVVFVAT-SFSTVLA---GLGCTYSLYNNGRQLVSVAPVSPGTDIDYIGRTHLTKLKFADTCE-
M_PH14_EuGene_00121761/1-67	MQFGTAFVVFVAT-SFSTVLA---GLGCTYSLYNNGRQLVSVAPVSPGTDIDYIGRTHLTKLKFADTCE-
M_GY11_EuGene_00139121/1-67	MQFGTAFVVFVAT-SFSTVLA---GLGCTYSLYNNGRQLVSVAPVSPGTDIDYIGRTHLTKLKFADTCE-
M_FR13_EuGene_00076081/1-67	MQFGTAFVVFVAT-SFSTVLA---GLGCTYSLYNNGRQLVSVAPVSPGTDIDYIGRTHLTKLKFADTCE-
M_US11_EuGene_00135351/1-67	MQFGTAFVVFVAT-SFSTVLA---GLGCTYSLYNNGRQLVSVAPVSPGTDIDYIGRTHLTKLKFADTCE-
M_CD156_EuGene_00098011/1-67	MQFGTAFVVFVAT-SFSTVLA---GLGCTYSLYNNGRQLVSVAPVSPGTDIDYIGRTHLTKLKFADTCE-
M_BR32_EuGene_00116041/1-67	MQFGTAFVVFVAT-SFSTVLA---GLGCTYSLYNNGRQLVSVAPVSPGTDIDYIGRTHLTKLKFADTCE-
M_BR29_EuGene_00107481/1-67	MQFGTFTLVFIAAAMIPAVFA---RCTQSVVDSGKMVVEKDLPTIDDYITII-GKKKYITFYFVLEKCD-
MGG_1613370/3-73	MQLGKTFEVALVAARPTLVAGFP---IYEGCVLRIKGIIEYEGRMEEFGATVYFGGTVVILGDCCK-
M_FR13_EuGene_00052401/10-110	MLEQACNFILNALPANTVAGFTPLDTCCLIRICKGDK-TIHAKFAPAGTEFDWIDDRLLTVLRLRSDCV-
M_US11_EuGene_00021061/10-110	MLEQACNFILNALPANTVAGFTPLDTCCLIRICKGDK-TIHAKFAPAGTEFDWIDDRLLTVLRLRSDCV-
M_TH16_EuGene_00069311/10-110	MLEQACNFILNALPANTVAGFTPLDTCCLIRICKGDK-TIHAKFAPAGTEFDWIDDRLLTVLRLRSDCV-
M_TH12_EuGene_00080341/10-110	MLEQACNFILNALPANTVAGFTPLDTCCLIRICKGDK-TIHAKFAPAGTEFDWIDDRLLTVLRLRSDCV-
M_GY11_EuGene_00064691/10-110	MLEQACNFILNALPANTVAGFTPLDTCCLIRICKGDK-TIHAKFAPAGTEFDWIDDRLLTVLRLRSDCV-
MGG_0095470/10-110	MLEQACNFILNALPANTVAGFTPLDTCCLIRICKGDK-TIHAKFAPAGTEFDWIDDRLLTVLRLRSDCV-
M_CD156_EuGene_00028531/10-104	MLEQACNFMLLAA-LPANTVAGFTPLDTCCLIRICKGDK-TIHAKFAPAGTEFDWIDDRLLTVLRLRSDCV-
M_BR32_EuGene_00028251/10-104	MLEQACNFMLLAA-LPANTVAGFTPLDTCCLIRICKGDK-TIHAKFAPAGTEFDWIDDRLLTVLRLRSDCV-
MGG_0093170/10-110	MHFKQVCHLTIIVLVALPANTVAGFTPLDTCCLIRICKGDK-TIHAKFAPAGTEFDWIDDRLLTVLRLRSDCV-
M_US11_EuGene_00016281/1-70	MHLAKPFLKLLTLVALDGIAAA-KGQPEHCFVRILKYEI-PHIGDGYLVGSDVNWVGNQKVDFFIINDCTI-
M_TH16_EuGene_0003721/1-70	MHLAKPFLKLLTLVALDGIAAA-KGQPEHCFVRILKYEI-PHIGDGYLVGSDVNWVGNQKVDFFIINDCTI-
M_TH12_EuGene_00036661/1-70	MHLAKPFLKLLTLVALDGIAAA-KGQPEHCFVRILKYEI-PHIGDGYLVGSDVNWVGNQKVDFFIINDCTI-
M_PH14_EuGene_00010991/1-70	MHLAKPFLKLLTLVALDGIAAA-KGQPEHCFVRILKYEI-PHIGDGYLVGSDVNWVGNQKVDFFIINDCTI-
M_GY11_EuGene_00081081/1-70	MHLAKPFLKLLTLVALDGIAAA-KGQPEHCFVRILKYEI-PHIGDGYLVGSDVNWVGNQKVDFFIINDCTI-
M_FR13_EuGene_00056271/1-70	MHLAKPFLKLLTLVALDGIAAA-KGQPEHCFVRILKYEI-PHIGDGYLVGSDVNWVGNQKVDFFIINDCTI-
M_CD156_EuGene_00040101/1-70	MHLAKPFLKLLTLVALDGIAAA-KGQPEHCFVRILKYEI-PHIGDGYLVGSDVNWVGNQKVDFFIINDCTI-
MGG_1357570/1-70	MHLAKPFLKLLTLVALDGIAAA-KGQPEHCFVRILKYEI-PHIGDGYLVGSDVNWVGNQKVDFFIINDCTI-
M_BR32_EuGene_00003601/1-70	MQFQNIILSILLAFDASSVAA---NRRGCKVLYRNRKEAQAATYAEIDKETHWIDGFFIIVIKPDGDCI-
M_US11_EuGene_00106891/1-71	MQFQNIILSILLAFDASSVAA---NRRGCKVLYRNRKEAQAATYAEIDKETHWIDGFFIIVIKPDGDCI-
M_FR13_EuGene_0006611/1-71	MQFQNIILSILLAFDASSVAA---NRRGCKVLYRNRKEAQAATYAEIDKETHWIDGFFIIVIKPDGDCI-
M_TH16_EuGene_00079081/1-71	MQFQNIILSILLAFDASSVAA---NRRGCKVLYRNRKEAQAATYAEIDKETHWIDGFFIIVIKPDGDCI-
M_TH12_EuGene_00084101/1-71	MQFQNIILSILLAFDASSVAA---NRRGCKVLYRNRKEAQAATYAEIDKETHWIDGFFIIVIKPDGDCI-
M_PH14_EuGene_00078711/1-71	MQFQNIILSILLAFDASSVAA---NRRGCKVLYRNRKEAQAATYAEIDKETHWIDGFFIIVIKPDGDCI-
M_GY11_EuGene_00121251/1-71	MQFQNIILSILLAFDASSVAA---NRRGCKVLYRNRKEAQAATYAEIDKETHWIDGFFIIVIKPDGDCI-
M_FR13_EuGene_00121051/1-71	MQFQNIILSILLAFDASSVAA---NRRGCKVLYRNRKEAQAATYAEIDKETHWIDGFFIIVIKPDGDCI-
M_BR32_EuGene_00121051/1-71	MQFQNIILSILLAFDASSVAA---NRRGCKVLYRNRKEAQAATYAEIDKETHWIDGFFIIVIKPDGDCI-
MGG_1242670/30-95	MRFANAWLAFVAVTQALGVAA---ELCTIDILFRGR-RVTFRNIIPSGQTEV-GKRCIISVDDSCQP-
M_US11_EuGene_00129431/1-66	MRFANAWLAFVAVTQALGVAA---ELCTIDILFRGR-RVTFRNIIPSGQTEV-GKRCIISVDDSCQP-
M_TH16_EuGene_00027191/1-66	MRFANAWLAFVAVTQALGVAA---ELCTIDILFRGR-RVTFRNIIPSGQTEV-GKRCIISVDDSCQP-
M_TH12_EuGene_00047101/1-66	MRFANAWLAFVAVTQALGVAA---ELCTIDILFRGR-RVTFRNIIPSGQTEV-GKRCIISVDDSCQP-
M_PH14_EuGene_00084891/1-66	MRFANAWLAFVAVTQALGVAA---ELCTIDILFRGR-RVTFRNIIPSGQTEV-GKRCIISVDDSCQP-
M_GY11_EuGene_00137731/1-66	MRFANAWLAFVAVTQALGVAA---ELCTIDILFRGR-RVTFRNIIPSGQTEV-GKRCIISVDDSCQP-
M_FR13_EuGene_00093871/1-66	MRFANAWLAFVAVTQALGVAA---ELCTIDILFRGR-RVTFRNIIPSGQTEV-GKRCIISVDDSCQP-
M_BR32_EuGene_00142111/1-66	MRFANAWLAFVAVTQALGVAA---ELCTIDILFRGR-RVTFRNIIPSGQTEV-GKRCIISVDDSCQP-
M_CD156_EuGene_00122861/1-66	MRFANAWLAFVAVTQALGVAA---ELCTIDILFRGR-RVTFRNIIPSGQTEV-GKRCIISVDDSCQP-
M_BR29_EuGene_00091361/1-66	MHFCVTFEFLIALAIGTALGVAA---KNCCEIITLBSGE-TIGFVNVYPSGKRENRNWYITSCDEDCCK-
M_US11_EuGene_00034451/1-66	MHFNTOFFAIALVALDGTATA---VCFDIKONGT-DITSGNIWAGATANFVSGKWAVVSATSECKL
M_TH16_EuGene_00054231/1-66	MHFNTOFFAIALVALDGTATA---VCFDIKONGT-DITSGNIWAGATANFVSGKWAVVSATSECKL
M_TH12_EuGene_00010901/1-66	MHFNTOFFAIALVALDGTATA---VCFDIKONGT-DITSGNIWAGATANFVSGKWAVVSATSECKL
M_PH14_EuGene_00038181/1-66	MHFNTOFFAIALVALDGTATA---VCFDIKONGT-DITSGNIWAGATANFVSGKWAVVSATSECKL
M_GY11_EuGene_00019111/1-66	MHFNTOFFAIALVALDGTATA---VCFDIKONGT-DITSGNIWAGATANFVSGKWAVVSATSECKL
M_FR13_EuGene_00086971/1-66	MHFNTOFFAIALVALDGTATA---VCFDIKONGT-DITSGNIWAGATANFVSGKWAVVSATSECKL
MGG_1479370/1-66	MHFNTOFFAIALVALDGTATA---VCFDIKONGT-DITSGNIWAGATANFVSGKWAVVSATSECKL
M_BR32_EuGene_00011701/1-66	MHFNTOFFAIALVALDGTATA---VCFDIKONGT-DITSGNIWAGATANFVSGKWAVVSATSECKL
M_CD156_EuGene_0003391/1-66	MHFNTOFFAIALVALDGTATA---VCFDIKONGT-DITSGNIWAGATANFVSGKWAVVSATSECKL
M_BR29_EuGene_00048371/1-66	MRFNAEITFETLAVL-LPCTAADDIDPDHCDCEVILKSGKLIIGFVWSPGATEFDWIDDRLLTVLRLRSDCV-
M_BR32_EuGene_0011161/1-66	MRLDDTFFTLAFIAAHFSAAA---KGCPSVRRYDK-EVKWDIVEAGSHIRVDGYPATISADODCKL
M_PH14_EuGene_00174641/1-66	MRLDDTFFTLAFIAAHFSAAA---KGCPSVRRYDK-EVKWDIVEAGSHIRVDGYPATISADODCKL
MGG_1806070/1-66	MRLDDTFFTLAFIAAHFSAAA---KGCPSVRRYDK-EVKWDIVEAGSHIRVDGYPATISADODCKL
M_GY11_EuGene_00074021/1-66	MRLDDTFFTLAFIAAHFSAAA---KGCPSVRRYDK-EVKWDIVEAGSHIRVDGYPATISADODCKL
M_FR13_EuGene_00117281/1-66	MRLDDTFFTLAFIAAHFSAAA---KGCPSVRRYDK-EVKWDIVEAGSHIRVDGYPATISADODCKL
M_TH12_EuGene_00186341/1-66	MRLDDTFFTLAFIAAHFSAAA---KGCPSVRRYDK-EVKWDIVEAGSHIRVDGYPATISADODCKL
M_US11_EuGene_00106201/1-66	MRLDDTFFTLAFIAAHFSAAA---KGCPSVRRYDK-EVKWDIVEAGSHIRVDGYPATISADODCKL
M_TH16_EuGene_00135721/1-66	MRLDDTFFTLAFIAAHFSAAA---KGCPSVRRYDK-EVKWDIVEAGSHIRVDGYPATISADODCKL
M_CD156_EuGene_00125491/1-66	MRLDDTFFTLAFIAAHFSAAA---KGCPSVRRYDK-EVKWDIVEAGSHIRVDGYPATISADODCKL
M_BR29_EuGene_00087671/1-66	MRFDAITFETLALIMAYFGTAAA---KGCPSVRRYDK-EVKWDIVEAGSHIRVDGYPATISADODCKL
M_FR13_EuGene_00094051/1-66	MHFTSIFIPFALAALQVAAA---FCVNYDGLHLPALVLLMVRIGDTATIRGHFEVVEAKDKCV
M_BR29_EuGene_00106461/1-67	MHFTSILVLAALQVAAA---EFCAYFTGSNKSR-RRQVGVVIRIGDTIITADGHELVVHAADRQCV

S8 Fig.A continued

M_US71_EuGene_00064521/1-67	MQF	KSLF	AFAAL	TLL	IPAV	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	F	
M_FR13_EuGene_00001281/1-67	MQF	KSLF	AFAAL	TLL	IPAV	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	F	
MGG_17132/701-67	MQF	KSLF	AFAAL	TLL	IPAV	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	F	
M_TH16_EuGene_00016981/1-67	MQF	KSLF	AFAAL	TLL	IPAV	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	F	
M_GY11_EuGene_00010431/1-67	MQF	KSLF	AFAAL	TLL	IPAV	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	F	
M_PH14_EuGene_00001771/1-67	MQF	KSLF	AFAAL	TLL	IPAV	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	F	
M_CD156_EuGene_00015441/1-67	MQF	KSLF	AFAAL	TLL	IPAV	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	F	
M_BR32_EuGene_00038791/1-67	MQF	KSLF	AFAAL	TLL	IPAV	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	F	
M_BR29_EuGene_00094821/1-65	MRL	KL	LF	AFAAL	TLL	IPAV	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	
M_CD156_EuGene_00107581/1-66	MQF	NLL	LAAT	V	AM	LA	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	
M_BR32_EuGene_00128831/1-66	MQF	NLL	LAAT	V	AM	LA	SA	-----	LGCE	VLIS	SKKG	DPV	KSSC	IPK	SGSK	VIV	NGK	TVT	VTS	ADG	SCK	
M_US71_EuGene_00048581/1-74	MQF	K	S	A	L	A	L	S	A	L	S	A	L	S	A	L	S	A	L	S	A	L
M_TH16_EuGene_00045961/1-74	MQF	K	S	A	L	A	L	S	A	L	S	A	L	S	A	L	S	A	L	S	A	L
M_TH12_EuGene_00052861/1-74	MQF	K	S	A	L	A	L	S	A	L	S	A	L	S	A	L	S	A	L	S	A	L
M_PH14_EuGene_0010861/1-74	MQF	K	S	A	L	A	L	S	A	L	S	A	L	S	A	L	S	A	L	S	A	L
M_GY11_EuGene_00096861/1-74	MQF	K	S	A	L	A	L	S	A	L	S	A	L	S	A	L	S	A	L	S	A	L
M_FR13_EuGene_00093031/1-74	MQF	K	S	A	L	A	L	S	A	L	S	A	L	S	A	L	S	A	L	S	A	L
MGG_18019/701-74	MQF	K	S	A	L	A	L	S	A	L	S	A	L	S	A	L	S	A	L	S	A	L
M_BR32_EuGene_00060601/1-74	MQF	K	S	A	L	A	L	S	A	L	S	A	L	S	A	L	S	A	L	S	A	L
M_CD156_EuGene_00097391/1-74	MQF	K	S	A	L	A	L	S	A	L	S	A	L	S	A	L	S	A	L	S	A	L
M_BR29_EuGene_00082031/1-71	MQF	Q	H	K	F	I	F	I	A	I	T	Q	A	A	T	V	A	S	-----	RD	DC	V
M_BR29_EuGene_00049931/1-79	MH	F	Q	S	I	F	V	L	C	V	A	I	T	Q	A	A	T	V	A	S	-----	RD
M_US71_EuGene_00120631/1-68	MQL	S	L	L	P	S	A	M	L	V	L	S	A	S	A	N	N	W	A	-----	W	
M_CD156_EuGene_00009301/1-68	MQL	S	L	L	P	S	A	M	L	V	L	S	A	S	A	N	N	W	A	-----	W	
M_BR32_EuGene_00101931/1-68	MQL	S	L	L	P	S	A	M	L	V	L	S	A	S	A	N	N	W	A	-----	W	
M_BR29_EuGene_00121041/1-71	M	-----	L	F	L	F	I	A	T	A	M	L	S	A	S	A	N	N	W	A	-----	W
M_CD156_EuGene_00086281/1-83	M	-----	V	T	T	A	M	A	L	S	A	S	A	N	N	W	A	-----	W			
M_BR32_EuGene_00006631/1-83	M	-----	V	T	T	A	M	A	L	S	A	S	A	N	N	W	A	-----	W			
M_TH16_EuGene_00120731/1-83	M	-----	V	T	T	A	M	A	L	S	A	S	A	N	N	W	A	-----	W			
M_TH12_EuGene_00159991/1-83	M	-----	V	T	T	A	M	A	L	S	A	S	A	N	N	W	A	-----	W			
M_PH14_EuGene_00008801/1-83	M	-----	V	T	T	A	M	A	L	S	A	S	A	N	N	W	A	-----	W			
M_GY11_EuGene_00110071/1-83	M	-----	V	T	T	A	M	A	L	S	A	S	A	N	N	W	A	-----	W			
M_BR32_EuGene_0006101/1-82	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
M_US71_EuGene_00013991/1-82	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
M_CD156_EuGene_00086781/1-82	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
M_TH16_EuGene_0004011/1-82	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
M_TH12_EuGene_00039091/1-82	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
M_PH14_EuGene_00008531/1-82	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
M_GY11_EuGene_00148071/1-82	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
MGG_08944/701-82	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
CH063_14852/701-82	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
ENH48621_CO3-67	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
ELA28482_CO3-102	M	-----	F	T	T	V	A	S	I	L	A	L	P	A	I	L	A	F	K	K	P	
M_BR29_EuGene_00126081/1-82	M	-----	Q	L	T	F	C	V	I	A	L	F	P	A	T	F	A	-----	D			
M_BR32_EuGene_00113101/1-82	M	-----	L	A	S	I	V	A	A	L	P	A	T	L	A	-----	D					
M_US71_EuGene_0013581/1-82	M	-----	F	L	A	S	I	V	A	A	L	P	A	T	L	A	-----	D				
M_TH16_EuGene_00135161/1-82	M	-----	F	L	A	S	I	V	A	A	L	P	A	T	L	A	-----	D				
M_TH12_EuGene_00187421/1-82	M	-----	F	L	A	S	I	V	A	A	L	P	A	T	L	A	-----	D				
M_GY11_EuGene_00148731/1-82	M	-----	F	L	A	S	I	V	A	A	L	P	A	T	L	A	-----	D				
M_CD156_EuGene_00095713/1-82	M	-----	F	L	A	S	I	V	A	A	L	P	A	T	L	A	-----	D				
ENH89188_CO14-76	M	-----	F	L	A	S	I	V	A	A	L	P	A	T	L	A	-----	D				
ELA38490_CO3-70	M	-----	V	I	L	S	A	C	F	A	G	V	T	A	-----	D						
ENH8800_CO20-77	M	-----	L	R	C	T	I	E	R	N	S	V	G	A	V	V	S	T	I	A	G	
M_BR32_EuGene_00085071/1-80	M	-----	F	A	A	L	F	A	A	L	F	V	T	G	A	A	-----	I				
ENH83538_CO5-68	M	-----	V	L	-----	A	L	A	L	S	A	M	A	A	P	Q	A	A	-----	I		
ENH79583_CO5-67	M	-----	F	-----	F	V	L	S	A	F	L	A	A	T	A	-----	I					
ENH81700_CO5-64	M	-----	F	-----	A	F	I	L	T	A	V	F	A	S	T	A	-----	I				
CH063_01399/701-64	M	-----	F	-----	F	T	F	V	-----	L	T	A	V	F	A	T	A	-----	I			
M_GY11_EuGene_00148091/1-53	M	-----	L	L	A	F	V	T	L	S	L	S	A	M	Q	V	S	A	-----	A		
M_US71_EuGene_00013911/1-67	M	-----	L	L	A	F	V	T	L	S	L	S	A	M	Q	V	S	A	-----	A		
M_TH16_EuGene_00040131/1-67	M	-----	L	L	A	F	V	T	L	S	L	S	A	M	Q	V	S	A	-----	A		
M_TH12_EuGene_00039111/1-67	M	-----	L	L	A	F	V	T	L	S	L	S	A	M	Q	V	S	A	-----	A		
M_PH14_EuGene_00008511/1-67	M	-----	L	L	A	F	V	T	L	S	L	S	A	M	Q	V	S	A	-----	A		
M_FR13_EuGene_00044251/1-67	M	-----	L	L	A	F	V	T	L	S	L	S	A	M	Q	V	S	A	-----	A		
M_BR32_EuGene_0006121/1-67	M	-----	L	L	A	F	V	T	L	S	L	S	A	M	Q	V	S	A	-----	A		
M_CD156_EuGene_00086761/1-71	M	-----	L	L	A	F	V	T	L	S	L	S	A	M	Q	V	S	A	-----	A		
M_BR32_EuGene_00143071/1-95	M	-----	P	A	S	V	O	E	C	F	W	F	M	F	V	N	G	-----	R			
M_BR32_EuGene_0007481/1-100	M	-----	H	I	G	P	C	I	L	N	S	G	T	P	V	A	I	D	A	P	K	
MGG_15922/701-61	M	-----	H	I	G	P	C	I	L	N	S	G	T	P	V	A	I	D	A	P	K	
M_CD156_EuGene_00138501/1-64	M	-----	F	T	T	I	A	A	C	L	S	Q	V	A	F	A	-----	L				
M_BR32_EuGene_0007481/1-64	M	-----	F	T	T	I	A	A	C	L	S	Q	V	A	F	A	-----	L				
M_BR29_EuGene_00117591/1-64	M	-----	M	R	S	F	T	T	I	A	L	G	S	T	V	A	-----	R				
M_CD156_EuGene_0013171/1-66	M	-----	M	R	S	F	T	T	I	A	L	G	S	T	V	A	-----	R				
M_BR29_EuGene_00117791/1-66	M	-----	M	R	S	F	T	T	I	A	L	G	S	T	V	A	-----	R				
M_BR32_EuGene_00136841/1-72	M	-----	M	R	S	F	T	T	I	A	L	G	S	T	V	A	-----	R				
M_BR29_EuGene_00096771/1-67	M	-----	M	H	F	K	G	L	A	V	A	T	A	V	P	P	S	T	L	G	-----	I
MGG_17267/701-95	M	-----	M	L	S	A	L	A	G	A	A	L	F	S	G	A	A	H	O	T	P	

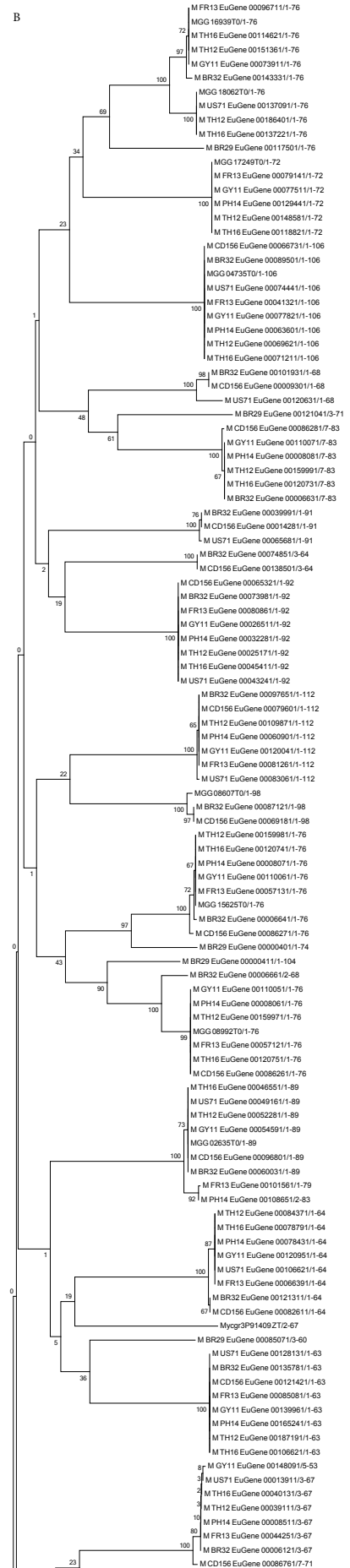
S8 Fig.A continued

M_BR29_EuGene_00014781/1-77	MQLQTLTILMAVALALLPDAALG---LTKCIIITIRKAGQTLVDLVSAPNSDLTF--GKTEVTFGVDEKCOA
M_BR29_EuGene_00043011/1-84	MKNQNLFAIITAAAMFPAVQAIINNNRRCCYITIFHRAGREWRRIQRTIITIPNCVNIIPKFWDDDDCTP
MGG_154597/1-83	MHIQNLFAIITASAALFPTVSAPPATEKCAVITIRVASKNGLKALVREGDYMAY--KHHGVVHFDDGECNPF
M_BR32_EuGene_00102861/1-83	MHIQNLFAIITASAALFPTVSAPPATEKCAVITIRVASKNGLKALVREGDYMAY--KHHGVVHFDDGECNPF
M_US11_EuGene_00072311/1-83	MHIQNLFAIITASAALFPTVSAPPATEKCAVITIRVASKNGLKALVREGDYMAY--KHHGVVHFDDGECNPF
M_FR13_EuGene_00080121/1-83	MHIQNLFAIITASAALFPTVSAPPATEKCAVITIRVASKNGLKALVREGDYMAY--KHHGVVHFDDGECNPF
M_TH12_EuGene_00159531/1-83	MHIQNLFAIITASAALFPTVSAPPATEKCAVITIRVASKNGLKALVREGDYMAY--KHHGVVHFDDGECNPF
M_GY11_EuGene_0012351/1-83	MHIQNLFAIITASAALFPTVSAPPATEKCAVITIRVASKNGLKALVREGDYMAY--KHHGVVHFDDGECNPF
MGG_11967/1-83	MHIQNLFAIITASAALFPTVSAPPATEKCAVITIRVASKNGLKALVREGDYMAY--KHHGVVHFDDGECNPF
M_TH16_EuGene_00118381/1-83	MHIQNLFAIITASAALFPTVSAPPATEKCAVITIRVASKNGLKALVREGDYMAY--KHHGVVHFDDGECNPF
M_PH14_EuGene_00124021/1-83	MHIQNLFAIITASAALFPTVSAPPATEKCAVITIRVASKNGLKALVREGDYMAY--KHHGVVHFDDGECNPF
M_CD156_EuGene_00117261/1-83	MHIQNLFAIITASAALFPTVSAPPATEKCAVITIRVASKNGLKALVREGDYIAY--KHHGVVHFDDGECNPF
M_BR29_EuGene_00112111/1-82	MQLQGLFAIITASAALFPTIYAPPSEQKCFYITVYRNGEKVGVVDHYSYTGWDAEIT---KKIIVHVFDSKCNPF
M_CD156_EuGene_00082141/1-74	MQLNLLFFAITGAVALLPTAYAESYKVKKCHYSIQKLHGTWGTGQADAGKDTHTI---EGTKVFFSNQCLPF
M_BR32_EuGene_00120811/1-74	MQLNLLFFAITGAVALLPTAYAESYKVKKCHYSIQKLHGTWGTGQADAGKDTHTI---EGTKVFFSNQCLPF
M_US11_EuGene_00139531/1-74	MQLNLLFFAITGAVALLPTAYAESYKVKKCHYSIQKLHGTWGTGQADAGKDTHTI---EGTKVFFSNQCLPF
M_TH16_EuGene_00079311/1-74	MQLNLLFFAITGAVALLPTAYAESYKVKKCHYSIQKLHGTWGTGQADAGKDTHTI---EGTKVFFSNQCLPF
M_TH12_EuGene_00083891/1-74	MQLNLLFFAITGAVALLPTAYAESYKVKKCHYSIQKLHGTWGTGQADAGKDTHTI---EGTKVFFSNQCLPF
M_PH14_EuGene_00078951/1-74	MQLNLLFFAITGAVALLPTAYAESYKVKKCHYSIQKLHGTWGTGQADAGKDTHTI---EGTKVFFSNQCLPF
M_GY11_EuGene_00049391/1-74	MQLNLLFFAITGAVALLPTAYAESYKVKKCHYSIQKLHGTWGTGQADAGKDTHTI---EGTKVFFSNQCLPF
M_FR13_EuGene_00089161/1-74	MQLNLLFFAITGAVALLPTAYAESYKVKKCHYSIQKLHGTWGTGQADAGKDTHTI---EGTKVFFSNQCLPF
M_BR29_EuGene_00060181/1-73	MQPIGLFTIITATAALPTVYASFKAGKICHTILKMYKNIRKQSDAPADDTITII---EGIKVFFSKDCITP
M_BR32_EuGene_00117301/1-70	MRIQDFALFIVTVALLPIVVA--GDVEQDCYITVWTSKGNRSDKIRSDHLLGI---HGASVYDFDDEDCTP
M_US11_EuGene_00123911/1-70	MRIQDFALFIVTVALLPIVVA--GDVEQDCYITVWTSKGNRSDKIRSDHLLGI---HGASVYDFDDEDCTP
M_TH16_EuGene_0009371/1-70	MRIQDFALFIVTVALLPIVVA--GDVEQDCYITVWTSKGNRSDKIRSDHLLGI---HGASVYDFDDEDCTP
M_FR13_EuGene_00094721/1-70	MRIQDFALFIVTVALLPIVVA--GDVEQDCYITVWTSKGNRSDKIRSDHLLGI---HGASVYDFDDEDCTP
M_CD156_EuGene_0005041/1-124	MRIQVCTVVIAPLALVFNVIASPPPAECEVTVYRNGEDARRKHSQEGTYSFHGKFGVVKVRFDRRCRPF
M_TH12_EuGene_00089291/1-124	MRIQVCTVVIAPLALVFNVIASPPPAECEVTVYRNGEDARRKHSQEGTYSFHGKFGVVKVRFDRRCRPF
M_TH12_EuGene_00102651/1-124	MRIQVCTVVIAPLALVFNVIASPPPAECEVTVYRNGEDARRKHSQEGTYSFHGKFGVVKVRFDRRCRPF
M_PH14_EuGene_00106641/1-124	MRIQVCTVVIAPLALVFNVIASPPPAECEVTVYRNGEDARRKHSQEGTYSFHGKFGVVKVRFDRRCRPF
M_GY11_EuGene_00121071/1-124	MRIQVCTVVIAPLALVFNVIASPPPAECEVTVYRNGEDARRKHSQEGTYSFHGKFGVVKVRFDRRCRPF
M_US11_EuGene_00071771/1-131	MRIQVCTVVIAPLALVFNVIASPPPAECEVTVYRNGEDARRKHSQEGTYSFHGKFGVVKVRFDRRCRPF
MGG_09425/1-124	MRIQVCTVVIAPLALVFNVIASPPPAECEVTVYRNGEDARRKHSQEGTYSFHGKFGVVKVRFDRRCRPF
M_BR32_EuGene_00107091/1-124	MRIQVCTVVIAPLALVFNVIASPPPLECEVTVYRNS---DHHGKAANKSKQNI---LGVVRFDRRCRPF
M_US11_EuGene_00000151/1-78	MHIQNLFIALSAALLLSVHAPGAIIVKACPYAFRHGKRPVYGEADANSSEEEI---LFAVKRFAPDCKI
M_CD156_EuGene_00141051/1-78	MHIQNLFIALSAALLLSVYAPGAIIVKACPYAFRHGKRPVYGEADANSSEEOI---LFAVKRFAPDCKI
M_BR32_EuGene_00128081/1-78	MHIQNLFIALSAALLLSVYAPGAIIVKACPYAFRHGKRPVYGEADANSSEEOI---LFAVKRFAPDCKI
M_BR29_EuGene_00076521/1-96	MYTHSLFTIAAIALSPTVYAVVAVVGGSCFETLIIKKNKPKMPSGGLDShKSIKI---PFGGMTFNWACNPF
MGG_13133/1-105	MYTHSLFTVIAATVALSPTVYAFNKRVPGCCFIFYKGDTLIGTOVERVGSKYSW--KFEFAGCKWDKCEKPF
M_TH16_EuGene_00060141/1-105	MYTHSLFTVIAATVALSPTVYAFNKRVPGCCFIFYKGDTLIGTOVERVGSKYSW--KFEFAGCKWDKCEKPF
M_GY11_EuGene_00067931/1-105	MYTHSLFTVIAATVALSPTVYAFNKRVPGCCFIFYKGDTLIGTOVERVGSKYSW--KFEFAGCKWDKCEKPF
M_FR13_EuGene_00003001/1-105	MYTHSLFTVIAATVALSPTVYAFNKRVPGCCFIFYKGDTLIGTOVERVGSKYSW--KFEFAGCKWDKCEKPF
M_CD156_EuGene_00076681/1-85	MKLATGFLPYIGIALSATSITKDRKEPACHVVIVKPP---LTFAWWKAGVVEWPKYSGVVKVTDDEDCHL
M_BR32_EuGene_00075311/1-85	MKLATGFLPYIGIALSATSITKDRKEPACHVVIVKPP---LTFAWWKAGVVEWPKYSGVVKVTDDEDCHL
M_TH16_EuGene_00009641/1-85	MKLATGFLPYIGIALSATSITKDRKEPACHVVIVKPP---LTFAWWKAGVVEWPKYSGVVKVTDDEDCHL
M_PH14_EuGene_00130281/1-85	MKLATGFLPYIGIALSATSITKDRKEPACHVVIVKPP---LTFAWWKAGVVEWPKYSGVVKVTDDEDCHL
M_US11_EuGene_00000861/1-85	MKLATGFLPYIGIALSATSITKDRKEPACHVVIVKPP---LTFAWWKAGVVEWPKYSGVVKVTDDEDCHL
M_TH12_EuGene_00135851/1-85	MKLATGFLPYIGIALSATSITKDRKEPACHVVIVKPP---LTFAWWKAGVVEWPKYSGVVKVTDDEDCHL
M_GY11_EuGene_00158751/1-85	MKLATGFLPYIGIALSATSITKDRKEPACHVVIVKPP---LTFAWWKAGVVEWPKYSGVVKVTDDEDCHL
MGG_05896/1-85	MKLATGFLPYIGIALSATSITKDRKEPACHVVIVKPP---LTFAWWKAGVVEWPKYSGVVKVTDDEDCHL
CCT62538_GFI-100	MKFSITIFASLVAYCIGVHSMPSDKGWCWKCDVKTDSVRSRGTWCWAKNWN--GWDGHWVSKHDDCK
M_CD156_EuGene_00113751/1-95	MRFVTFNTVFLTLGLISLVNAQPGIPPSDQCFWFLKKNVROHAGVCCSDWSFDMKI--GQYVWITHTDDEGRLL
M_BR32_EuGene_00122611/1-95	MRFVTFNTVFLTLGLISLVNAQPGIPPSDQCFWFLKKNVROHAGVCCSDWSFDMKI--GQYVWITHTDDEGRLL
M_TH12_EuGene_00195311/1-95	---TTFNTFLLLTLG--TVAVVNAAGIPVPECFWFMFNKNNVQDAGTCYSWVKMMDKVK--GPNVWHIKSDDNCNLL
M_FR13_EuGene_00129011/1-95	---TTFNTFLLLTLG--TVAVVNAAGIPVPECFWFMFNKNNVQDAGTCYSWVKMMDKVK--GPNVWHIKSDDNCNLL
M_PH14_EuGene_00183451/1-95	---TTFNTFLLLTLG--TVAVVNAAGIPVPECFWFMFNKNNV--RODDGTCYSWVKMMDKVK--GPNVWHIKSDDNCNLL
M_TH16_EuGene_00136901/1-95	---TTFNTFLLLTLG--TVAVVNAAGIPVPECFWFMFNKNNV--RODDGTCYSWVKMMDKVK--GPNVWHIKSDDNCNLL
M_GY11_EuGene_00191731/1-95	---TTFNTFLLLTLG--TVAVVNAAGIPVPECFWFMFNKNNV--RODDGTCYSWVKMMDKVK--GPNVWHIKSDDNCNLL
MGG_15972/1-95	---TTFNTFLLLTLG--TVAVVNAAGIPVPECFWFMFNKNNV--RODDGTCYSWVKMMDKVK--GPNVWHIKSDDNCNLL
MGG_17799/1-95	---FKACLAMTAALALYVQTVAAV--KTVYPCRDLIYDLA--RNLVFDGSEEGARGIT--LEPKYKLYARDCK
MGG_16416/1-117	---INHLAAITAAAMFFAVIIVAHVKGORLCEYRKRNOHQNLQIRHSTEVILPETMGIALYVGRDC
M_TH12_EuGene_00135081/1-68	MQLSALAAVAALALFPEATVA---AGAKCRVDLKHGGA--VVVSFEVLVLAGSRKTVEEGFSWTCSTERTCN
M_TH12_EuGene_00187351/1-68	MQLSALAAVAALALFPEATVA---AGAKCRVDLKHGGA--VVVSFEVLVLAGSRKTVEEGFSWTCSTERTCN
M_PH14_EuGene_00150811/1-68	MQLSALAAVAALALFPEATVA---AGAKCRVDLKHGGA--VVVSFEVLVLAGSRKTVEEGFSWTCSTERTCN
M_GY11_EuGene_00130211/1-68	MQLSALAAVAALALFPEATVA---AGAKCRVDLKHGGA--VVVSFEVLVLAGSRKTVEEGFSWTCSTERTCN
M_FR13_EuGene_00102441/1-68	MQLSALAAVAALALFPEATVA---AGAKCRVDLKHGGA--VVVSFEVLVLAGSRKTVEEGFSWTCSTERTCN
MGG_09675/1-68	MQLSALAAVAALALFPEATVA---AGAKCRVDLKHGGA--VVVSFEVLVLAGSRKTVEEGFSWTCSTERTCN
M_US11_EuGene_00131061/1-68	MQLSALAAVAALALFPEATVA---AGAKCRVDLKHGGA--VVVSFEVLVLAGSRKTVEEGFSWTCSTERTCN
M_CD156_EuGene_00125371/1-67	MQLSALAAVAALALFPEATVA---AGAKCRVDLKHGGA--VVVSFEVLVLAGSRKTVEEGFSWTCSTERTCN
M_BR32_EuGene_00111041/1-67	MQLSALAAVAALALFPEATVA---AGAKCRVDLKHGGA--VVVSFEVLVLAGSRKTVEEGFSWTCSTERTCN
M_BR29_EuGene_0008741/1-69	MQLSALLAAAALALFPEFTVA---AGGECIVKLDHPNGKVVSWFIALAGKREVK--KEFSWBCSTNSTICE
M_BR32_EuGene_00141031/1-98	MVLRKILAIATILALSPPGALADADSPARCKVITSMAFRQIYGGELTTRGGVVAI---ENFKCTVDETCNPF
M_US11_EuGene_00110221/1-98	MVLRKILAIATILALSPPGALADADSPARCKVITSMAFRQIYGGELTTRGGVVAI---ENFKCTVDETCNPF
M_PH14_EuGene_00122181/1-98	MVLRKILAIATILALSPPGALADADSPARCKVITSMAFRQIYGGELTTRGGVVAI---ENFKCTVDETCNPF
M_GY11_EuGene_00091931/1-98	MVLRKILAIATILALSPPGALADADSPARCKVITSMAFRQIYGGELTTRGGVVAI---ENFKCTVDETCNPF
M_FR13_EuGene_00130761/1-98	MVLRKILAIATILALSPPGALADADSPARCKVITSMAFRQIYGGELTTRGGVVAI---ENFKCTVDETCNPF
MGG_08469/1-98	MVLRKILAIATILALSPPGALADADSPARCKVITSMAFRQIYGGELTTRGGVVAI---ENFKCTVDETCNPF
M_CD156_EuGene_00130031/1-91	MVLRKILAIATILALSPPGALADADSPARCKVITSMAFRQIYGGELTTRGGVVAI---ENFKCTVDETCNPF
M_BR29_EuGene_00089621/1-97	MVLRKILAMAAILLSPAGALADHEGGDCVYITEMGRFOLYGGDYATRGGTVYI---NOYQCTVDVVMCNPF
M_BR29_EuGene_00011811/1-70	MQLKFFPTVFCATLFLPAMVLA-----VGCVITIDGPPYNYKXVGRGVTPGHQVVI---GDFICTVGTACNPF
M_US11_EuGene_00119301/1-85	MHYQNCASLIAAIVLFPAMAFANTPAVGCCYITIDGPPFNTKYGHAEISPGGSVVI---EOPKCTVDAICR
M_BR32_EuGene_00143081/1-85	MHYQNCASLIAAIVLFPAMAFANTPAVGCCYITIDGPPFNTKYGHAEISPGGSVVI---EOPKCTVDAICR
M_US11_EuGene_00000761/1-81	MHPENLFAIILATATLFPASAWADLRRPLCDVYIEVPSGVFGHDIVPAGETAIEI---GFRCKTVRITC
M_TH16_EuGene_00000541/1-81	MHPENLFAIILATATLFPASAWADLRRPLCDVYIEVPSGVFGHDIVPAGETAIEI---GFRCKTVRITC
M_TH12_EuGene_00135761/1-81	MHPENLFAIILATATLFPASAWADLRRPLCDVYIEVPSGVFGHDIVPAGETAIEI---GFRCKTVRITC
M_PH14_EuGene_00113591/1-81	MHPENLFAIILATATLFPASAWADLRRPLCDVYIEVPSGVFGHDIVPAGETAIEI---GFRCKTVRITC
M_GY11_EuGene_00094941/1-81	MHPENLFAIILATATLFPASAWADLRRPLCDVYIEVPSGVFGHDIVPAGETAIEI---GFRCKTVRITC
M_CD156_EuGene_00132961/1-81	MHPENLFAIILATATLFPASAWADLRRPLCDVYIEVPSGVFGHDIVPAGETAIEI---GFRCKTVRITC
M_BR32_EuGene_00075401/1-81	MHPENLFAIILATATLFPASAWADLRRPLCDVYIEVPSGVFGHDIVPAGETAIEI---GFRCKTVRITC
M_BR29_EuGene_00095641/1-77	MHPENLFAIFATAALFPAVYARLKRVVICVEFIRMSSGLPLGKDFVPAGETTII---HAYTCEIDTMCGA
M_CD156_EuGene_00129861/1-101	MQLPLTVVAATLTLAPLTPGAFAEVLRNPPCKVYIYAPGRTLGFASAGKNEDEVVSTGHTYQCRDATIC
M_TH12_EuGene_00141221/1-101	MQLPLTVVAATLTLAPLTPGAFAEVLRNPPCKVYIYAPGRTLGFASAGKNEDEVVSTGHTYQCRDATIC
M_PH14_EuGene_00054331/1-101	MQLPLTVVAATLTLAPLTPGAFAEVLRNPPCKVYIYAPGRTLGFASAGKNEDEVVSTGHTYQCRDATIC
M_GY11_EuGene_00134561/1-101	MQLPLTVVAATLTLAPLTPGAFAEVLRNPPCKVYIYAPGRTLGFASAGKNEDEVVSTGHTYQCRDATIC
M_FR13_EuGene_00128431/1-101	MQLPLTVVAATLTLAPLTPGAFAEVLRNPPCKVYIYAPGRTLGFASAGKNEDEVVSTGHTYQCRDATIC
MGG_08482/1-101	MQLPLTVVAATLTLAPLTPGAFAEVLRNPPCKVYIYAPGRTLGFASAGKNEDEVVSTGHTYQCRDATIC
M_TH16_EuGene_00107161/1-101	MQLPLTVVAATLTLAPLTPGAFAEVLRNPPCKVYIYAPGRTLGFASAGKNEDEVVSTGHTYQCRDATIC
M_US11_EuGene_00132991/1-101	MQLPLTVVAATLTLAPLTPGAFAEVLRNPPCKVYIYAPGRTLGFASAGKNEDEVVSTGHTYQCRDATIC
M_BR32_EuGene_00141191/1-101	MQLPLTVVAATLTLAPLTPGAFAEVLRNPPCKVYIYAPGRTLGFASAGKNEDEVVSTGHTYQCRDATIC
M_BR29_EuGene_00081821/1-102	MQLGNTLVAATLTLAPLTPGAFAEVLRNPPCKVYIYAPGRTLGFASAGKNEDEVVSTGHTYQCRDATIC
M_CD156_EuGene_00103411/1-82	MQLATVFAAAATLAVLPTGLVARYIHPHECEVITIMKPPTPSPDRGYPGSDSKTKI--KVGSKFCVADSHCTP
M_US11_EuGene_00116781/1-82	MQLATVFAAAATLAVLPTGLVARYIHPHECEVITIMKPPTPSPDRGYPGSDSKTKI--KVGSKFCVADSHCTP
M_BR32_EuGene_00133701/1-82	MQLATVFAAAATLAVLPTGLVARYIHPHECEVITIMKPPTPSPDRGYPGSDSKTKI--KVGSKFCVADSHCTP
M_GY11_EuGene_00134531/1-94	---FEFLAAAATLVALSPFLVFGDPTVYDGGIIFRFPNGALAEVIVYNGGETFKI--IEGRYTCVSEONCR
M_TH12_EuGene_00141231/1-96	MRFEFLAAAATLVALSPFLVFGDPTVYDGGIIFRFPNGALAEVIVYNGGETFKI--IEGRYTCVSEONCR
M_PH14_EuGene_0004431/1-96	MRFEFLAAAATLVALSPFLVFGDPTVYDGGIIFRFPNGALAEVIVYNGGETFKI--IEGRYTCVSEONCR
M_FR13_EuGene_00025011/1-96	MRFEFLAAAATLVALSPFLVFGDPTVYDGGIIFRFPNGALAEVIVYNGGETFKI--IEGRYTCVSEONCR
CB01159_LM1-70	MGEFTIFAAVAAALLPRLVLAAGDIEDCCCKDQKQ--HKNVDVCGKSHAEI--CENYKVSQCADGCK
M_TH16_EuGene_00114921/1-102	MCPGTFLASALAALVAPAGALAAPRRQCFLSADLGGPITLLEATPFVNDMDFD---HGRCQMGKDC
M_TH12_EuGene_00168001/1-102	MCPGTFLASALAALVAPAGALAAPRRQCFLSADLGGPITLLEATPFVNDMDFD---HGRCQMGKDC
M_PH14_EuGene_00112331/1-102	MCPGTFLASALAALVAPAGALAAPRRQCFLSADLGGPITLLEATPFVNDMDFD---HGRCQMGKDC
M_GY11_EuGene_00112871/1-102	MCPGTFLASALAALVAPAGALAAPRRQCFLSADLGGPITLLEATPFVNDMDFD---HGRCQMGKDC
M_FR13_EuGene_00070821/1-102	MCPGTFLASALAALVAPAGALAAPRRQCFLSADLGGPITLLEATPFVNDMDFD---HGRCQMGKDC
MGG_02546/1-102	MCPGTFLASALAALVAPAGALAAPRRQCFLSADLGGPITLLEATPFVNDMDFD---HGRCQMGKDC
M_US11_EuGene_00108991/1-102	MCPGTFLASALAALVAPAGALAAPRRQCFLSADLGGPITLLEATPFVNDMDFD---HGRCQMGKDC
M_BR32_EuGene_00061371/1-102	MCPGTFLASALAALVAPAGALAAPRRQCFLSADLGGPITLLEATPFVNDMDFD---HGRCQMGKDC
M_CD156_EuGene_00127331/1-102	MCPGTFLASALAALVAPAGALAAPRRQCFLSADLGGPITLLEATPFVNDMDFD---HGRCQMGKDC

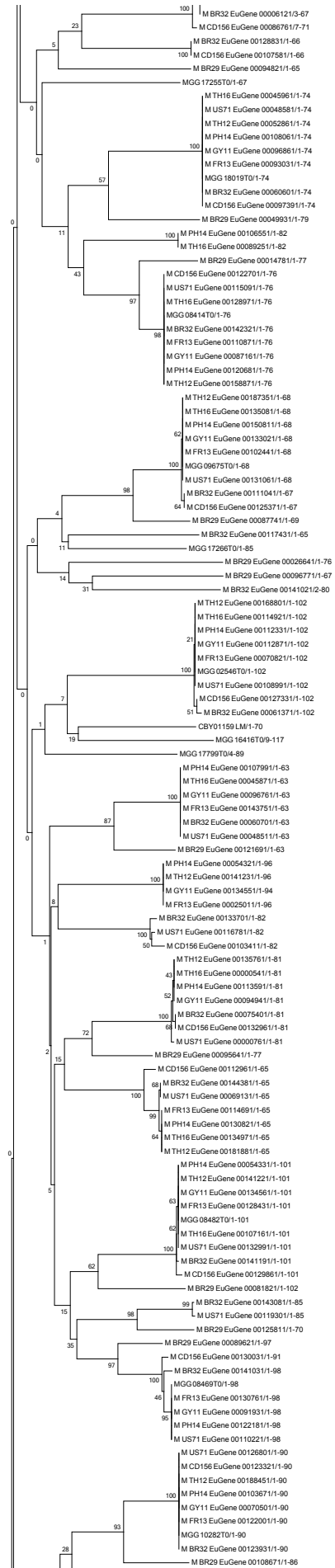
S8 Fig.A continued

M_TH16_EuGene_00106621/1-63	MKFSAIYLAALT--LVSGAAAA--DPKNCKFNVLNGSGINVANGCCDVTKQVTV--SPFFTVSCTE--SCG--
M_TH12_EuGene_00187191/1-63	MKFSAIYLAALT--LVSGAAAA--DPKNCKFNVLNGSGINVANGCCDVTKQVTV--SPFFTVSCTE--SCG--
M_PH14_EuGene_00165241/1-63	MKFSAIYLAALT--LVSGAAAA--DPKNCKFNVLNGSGINVANGCCDVTKQVTV--SPFFTVSCTE--SCG--
M_GY11_EuGene_00139961/1-63	MKFSAIYLAALT--LVSGAAAA--DPKNCKFNVLNGSGINVANGCCDVTKQVTV--SPFFTVSCTE--SCG--
M_FR13_EuGene_00085081/1-63	MKFSAIYLAALT--LVSGAAAA--DPKNCKFNVLNGSGINVANGCCDVTKQVTV--SPFFTVSCTE--SCG--
M_CD156_EuGene_00121421/1-63	MKFSAIYLAALT--LVSGAAAA--DPKNCKFNVLNGSGINVANGCCDVTKQVTV--SPFFTVSCTE--SCG--
M_BR32_EuGene_00135781/1-63	MKFSAIYLAALT--LVSGAAAA--DPKNCKFNVLNGSGINVANGCCDVTKQVTV--SPFFTVSCTE--SCG--
M_US71_EuGene_00128131/1-63	MKFSAIYLAALT--LVSGAAAA--DPKNCKFNVLNGSGINVANGCCDVTKQVTV--SPFFTVSCTE--SCG--
M_BR32_EuGene_00117431/1-65	MQLKALVVAVAAAAIIGCATIPV-----VCKIVVVDANTIVILRGEAPAGESHNI---LGVWVITFESDCEA
M_CD156_EuGene_00082611/1-64	MHWEKNLALAIATSTCISVVNA---AKLCKWTVLNSEGEIEKSGGCLNTEYMKH---GDIHVEFENDCTF
M_BR32_EuGene_00121311/1-64	MHWEKNLALAIATSTCISVVNA---AKLCKWTVLNSEGEIEKSGGCLNTEYMKH---GDIHVEFENDCTF
M_TH16_EuGene_00078791/1-64	MHWEKNLALAIATSTCISVVNA---AKLCKWTVLNSEGEIEKSGGCLNTEYMKH---GDIHVEFENDCTF
M_TH12_EuGene_00084371/1-64	MHWEKNLALAIATSTCISVVNA---AKLCKWTVLNSEGEIEKSGGCLNTEYMKH---GDIHVEFENDCTF
M_PH14_EuGene_00078431/1-64	MHWEKNLALAIATSTCISVVNA---AKLCKWTVLNSEGEIEKSGGCLNTEYMKH---GDIHVEFENDCTF
M_GY11_EuGene_00120951/1-64	MHWEKNLALAIATSTCISVVNA---AKLCKWTVLNSEGEIEKSGGCLNTEYMKH---GDIHVEFENDCTF
M_US71_EuGene_00106621/1-64	MHWEKNLALAIATSTCISVVNA---AKLCKWTVLNSEGEIEKSGGCLNTEYMKH---GDIHVEFENDCTF
M_FR13_EuGene_00066391/1-64	MHWEKNLALAIATSTCISVVNA---AKLCKWTVLNSEGEIEKSGGCLNTEYMKH---GDIHVEFENDCTF
M_CD156_EuGene_00009311/1-69	MQLKSLFAFAALAVVPAVSVVFFGSGNTFNCCEVSVRDKSGTEVAHACIKAGGSKNV---GGAVVSTSNKCVV
M_BR32_EuGene_00101951/1-69	MQLKSLFAFAALAVVPAVSVVFFGSGNTFNCCEVSVRDKSGTEVAHACIKAGGSKNV---GGAVVSTSNKCVV
M_US71_EuGene_00117631/1-89	MOVGQLVFVAIGLSLLHVVIAADTRLVPGCLVKLYRPRRILVGAACIPVOTTWKI---KGHDISVWNGCTV
M_TH16_EuGene_00112201/1-89	MOVGQLVFVAIGLSLLHVVIAADTRLVPGCLVKLYRPRRILVGAACIPVOTTWKI---KGHDISVWNGCTV
M_TH12_EuGene_00039321/1-89	MOVGQLVFVAIGLSLLHVVIAADTRLVPGCLVKLYRPRRILVGAACIPVOTTWKI---KGHDISVWNGCTV
M_PH14_EuGene_00102791/1-89	MOVGQLVFVAIGLSLLHVVIAADTRLVPGCLVKLYRPRRILVGAACIPVOTTWKI---KGHDISVWNGCTV
M_GY11_EuGene_00069561/1-89	MOVGQLVFVAIGLSLLHVVIAADTRLVPGCLVKLYRPRRILVGAACIPVOTTWKI---KGHDISVWNGCTV
M_FR13_EuGene_00088361/1-89	MOVGQLVFVAIGLSLLHVVIAADTRLVPGCLVKLYRPRRILVGAACIPVOTTWKI---KGHDISVWNGCTV
MGG_103357D/1-89	MOVGQLVFVAIGLSLLHVVIAADTRLVPGCLVKLYRPRRILVGAACIPVOTTWKI---KGHDISVWNGCTV
M_BR32_EuGene_001123101/1-89	MOVGQLVFVAIGLSLLHVVIAADTRLVPGCLVKLYRPRRILVGAACIPVOTTWKI---KGHDISVWNGCTV
M_CD156_EuGene_00128071/1-89	MOVGQLVFVAIGLSLLHVVIAADTRLVPGCLVKLYRPRRILVGAACIPVOTTWKI---KGHDISVWNGCTV
M_BR32_EuGene_00131871/1-94	MOFKQALSVPILCALATTTISAASDOTINGCLVCFIFAGT---LGAACIKFHQSDYI---TGIKRVSVGEICN-
M_US71_EuGene_00134321/1-94	MOFKQALSVPILCALATTTISAASDOTINGCLVCFIFAGT---LGAACIKFHQSDYI---TGIKRVSVGEICN-
M_TH16_EuGene_00101931/1-94	MOFKQALSVPILCALATTTISAASDOTINGCLVCFIFAGT---LGAACIKFHQSDYI---TGIKRVSVGEICN-
M_PH14_EuGene_00104981/1-94	MOFKQALSVPILCALATTTISAASDOTINGCLVCFIFAGT---LGAACIKFHQSDYI---TGIKRVSVGEICN-
M_GY11_EuGene_00138881/1-94	MOFKQALSVPILCALATTTISAASDOTINGCLVCFIFAGT---LGAACIKFHQSDYI---TGIKRVSVGEICN-
M_TH16_EuGene_00045871/1-63	MKYGSILAALISLA-APASVLA---VWECHVTWVEGTPRGTEMLKEGGDDKTI---AGYTCRAGAGCKA
M_PH14_EuGene_00107991/1-63	MKYGSILAALISLA-APASVLA---VWECHVTWVEGTPRGTEMLKEGGDDKTI---AGYTCRAGAGCKA
M_GY11_EuGene_00096761/1-63	MKYGSILAALISLA-APASVLA---VWECHVTWVEGTPRGTEMLKEGGDDKTI---AGYTCRAGAGCKA
M_FR13_EuGene_00143751/1-63	MKYGSILAALISLA-APASVLA---VWECHVTWVEGTPRGTEMLKEGGDDKTI---AGYTCRAGAGCKA
M_BR32_EuGene_00067011/1-63	MKYGSILAALISLA-APASVLA---VWECHVTWVEGTPRGTEMLKEGGDDKTI---AGYTCRAGAGCKA
M_US71_EuGene_00048511/1-63	MKYGSILAALISLA-APASVLA---VWECHVTWVEGTPRGTEMLKEGGDDKTI---AGYTCRAGAGCKA
M_BR29_EuGene_00121691/1-63	MRFGSILAFAAALSP-IATVLA---VGECHITIWEGKT-PIGTETIAIEGEYVTI---AGYQCNAGAGCTA
M_TH12_EuGene_00181881/1-65	MKKSATIFATVATLALLPFGALA---GKWRCNVKIYNDKRYQGGASDDWGETLTI---RGYTCYIDSYCKA
M_TH16_EuGene_00134971/1-65	MKKSATIFATVATLALLPFGALA---GKWRCNVKIYNDKRYQGGASDDWGETLTI---RGYTCYIDSYCKA
M_PH14_EuGene_00130821/1-65	MKKSATIFATVATLALLPFGALA---GKWRCNVKIYNDKRYQGGASDDWGETLTI---RGYTCYIDSYCKA
M_FR13_EuGene_00114691/1-65	MKKSATIFATVATLALLPFGALA---GKWRCNVKIYNDKRYQGGASDDWGETLTI---RGYTCYIDSYCKA
M_US71_EuGene_00069131/1-65	MKKSATIFATVATLALLPFGALA---GKWRCNVKIYNDKRYQGGASDDWGETLTI---RGYTCYIDSYCKA
M_BR32_EuGene_00144381/1-65	MKKSATIFATVATLALLPFGALA---GKWRCNVKIYNDKRYQGGASDDWGETLTI---RGYTCYIDSYCKA
M_CD156_EuGene_00112961/1-65	MKKSATIFATVATLALLPFGALA---GKWRCNVKIYNDKRYQGGASDDWGETLTI---RGYTCYIDSYCKA
M_TH16_EuGene_00101881/1-75	MOFTAAVFTIAAALLPSALGAPAAVLPCKVEMFDNNARFVGMARGNWDKDFVI---KGHTIKCNTRCEI
M_TH12_EuGene_001176051/1-75	MOFTAAVFTIAAALLPSALGAPAAVLPCKVEMFDNNARFVGMARGNWDKDFVI---KGHTIKCNTRCEI
M_PH14_EuGene_00105041/1-75	MOFTAAVFTIAAALLPSALGAPAAVLPCKVEMFDNNARFVGMARGNWDKDFVI---KGHTIKCNTRCEI
M_FR13_EuGene_00104931/1-75	MOFTAAVFTIAAALLPSALGAPAAVLPCKVEMFDNNARFVGMARGNWDKDFVI---KGHTIKCNTRCEI
M_US71_EuGene_00134971/1-75	MOFTAAVFTIAAALLPSALGAPAAVLPCKVEMFDNNARFVGMARGNWDKDFVI---KGHTIKCNTRCEI
M_BR32_EuGene_00131911/1-75	MOFTAAVFTIAAALLPSALGAPAAVLPCKVEMFDNNARFVGMARGNWDKDFVI---KGHTIKCNTRCEI
M_US71_EuGene_00069131/1-75	MOFTAAVFTIAAALLPSALGAPAAVLPCKVEMFDNNARFVGMARGNWDKDFVI---KGHTIKCNTRCEI
M_BR29_EuGene_00124951/1-138	MKVATILTAVTITLVGATIPVPEHHIKGCKVEMGKGN-PLGAKHEWNTPII---AGWTFICNTDCTV
MGG_170777D/9-80	-----RAAIALTIFINAMVAEQSYRWDGTRMLFYS--QMDYIHVPPOTLVMNYASLRWGIEMTPGCG--

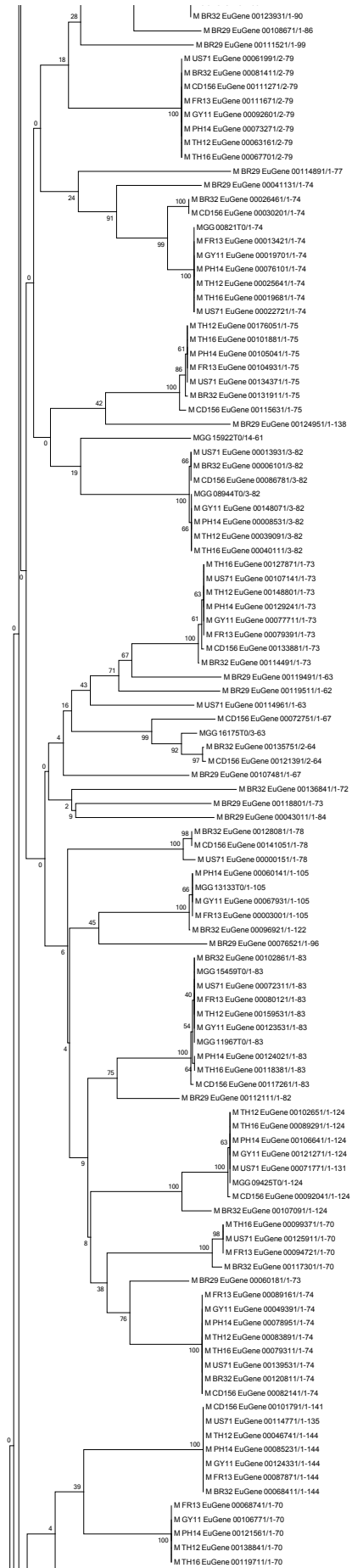
S8 Fig.C continued



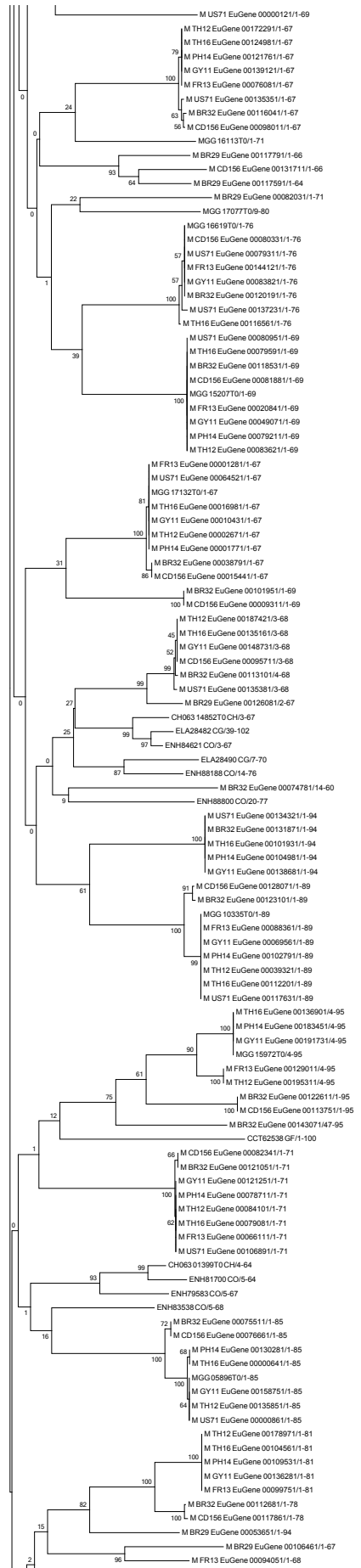
S8 Fig.C continued



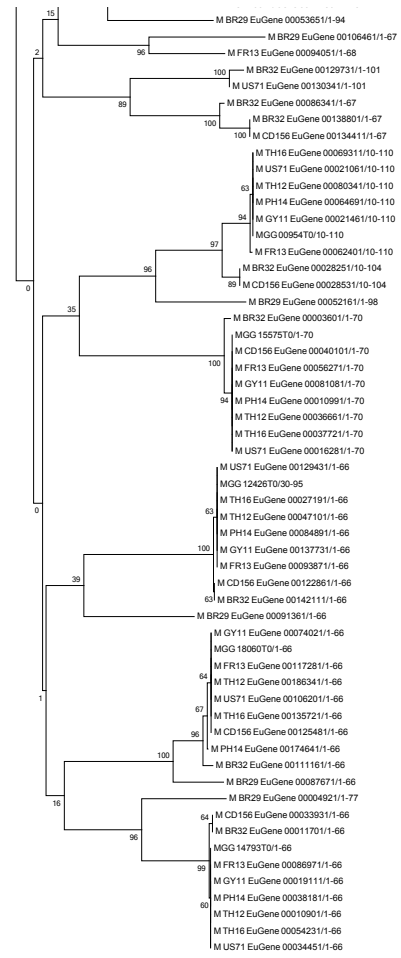
S8 Fig.C continued



S8 Fig.C continued



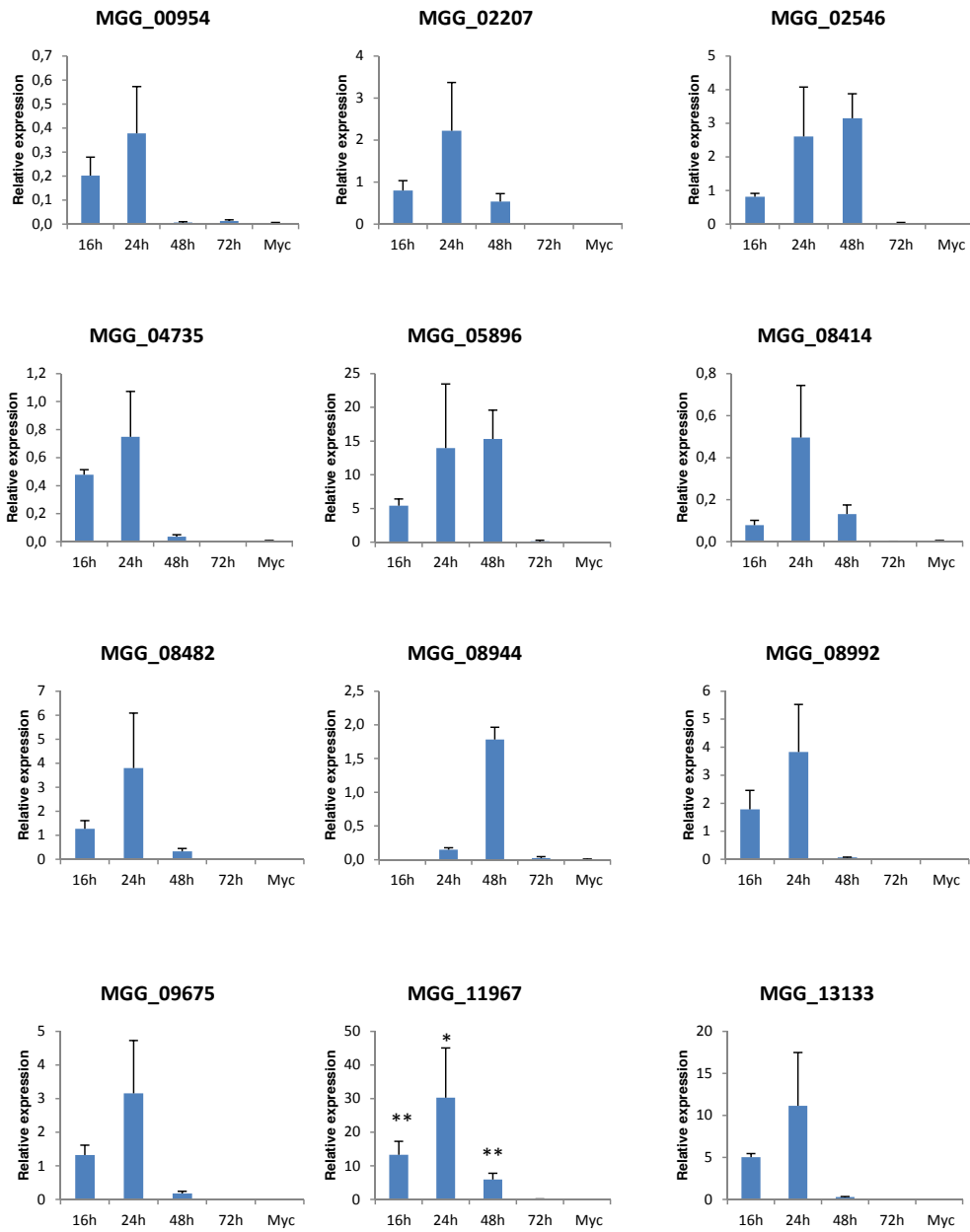
S8 Fig.C continued



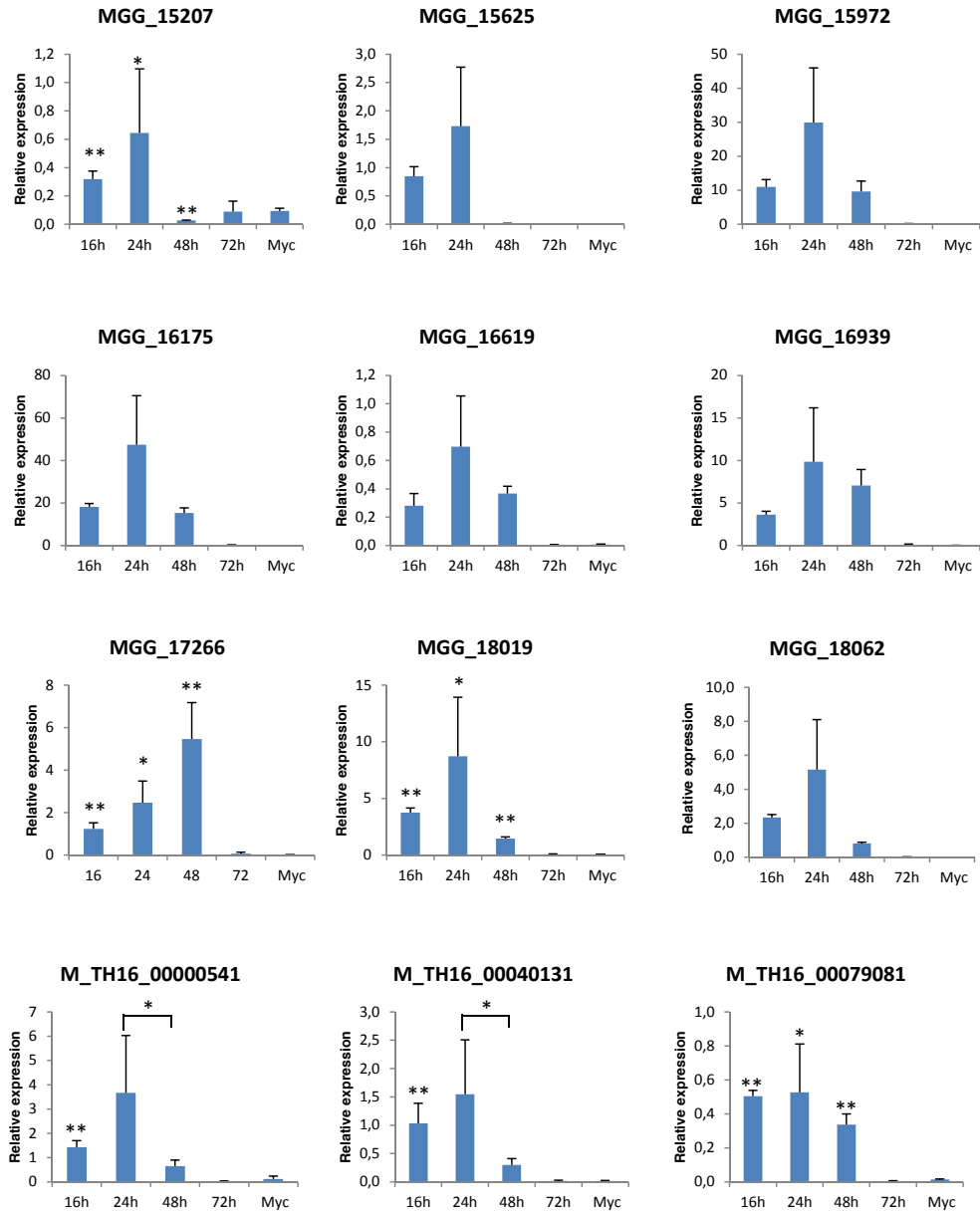
S9 Fig. Expression of *M. oryzae* MAX-effector candidates and marker genes during rice infection and in in vitro grown mycelium. mRNA levels of *M. oryzae* genes coding for MAX-effectors (A, B and C) and marker genes (D) was determined by q-RT-PCR in rice leaf samples harvested 16, 24, 48 or 72 h after inoculation and mycelium grown liquid medium for 72 hours. (A) Infection specific MAX-effectors identified in the HMM search, (B) infection specific MAX-effectors identified in the Psi Blast search but nor in the HMM search, (C) constitutively expressed MAX-effectors identified in the HMM search and (D) marker genes for appressorium and very early infection (*ORF3* of the ACE1 cluster, *MGG_08381*), biotrophic infection (*BAS3*, *MGG_11610*), late infection (*MGG_01147*), constitutive expression (*EF1 α* , *MGG_03641*). Relative expression levels were calculated by using expression of a constitutively expressed Actin (*MGG_03982*) as a reference. Mean values and standard deviation were calculated from three independent biological samples. The analyzed genes, were in most cases not or extremely weakly expressed in the mycelium. For genes with significant expression in the mycelium (ratio gene versus actine > 0,01) a T-test was performed to determine if in planta expression was significantly different from expression in the mycelium. In these cases (*MGG_11967*, *MGG_14793*, *MGG_15207*, *MGG_17266*, *MGG_18019*, *M_TH16_00000541*, *M_TH16_00040131*, *M_TH16_00079081*, *M_TH16_00104561*, *M_TH16_00120731*, *M_TH16_00124981*), a star or two stars (* or **) mark conditions where the expression was different from expression in the mycelium at respectively $p < 0,05$ or $p < 0,005$.

S9 Fig. continued

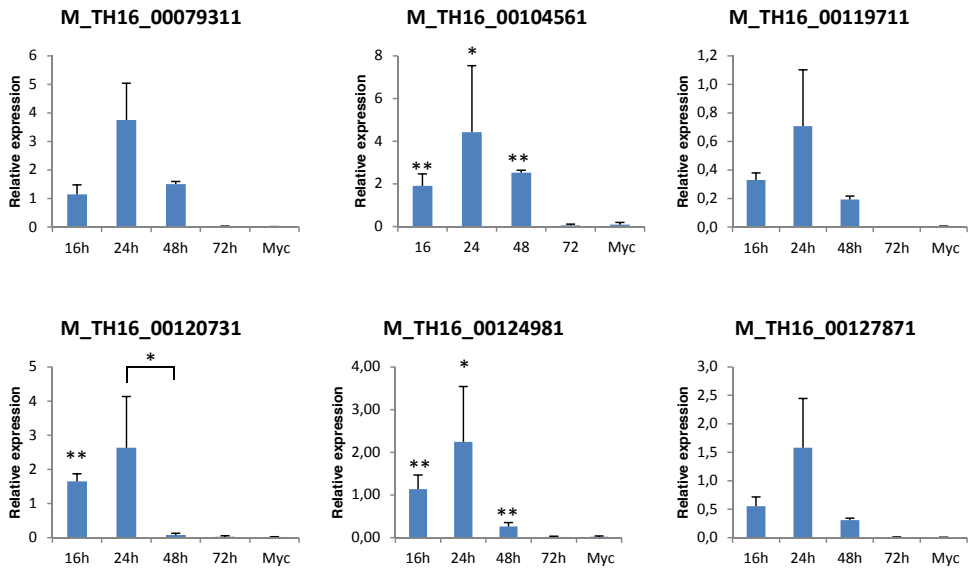
A



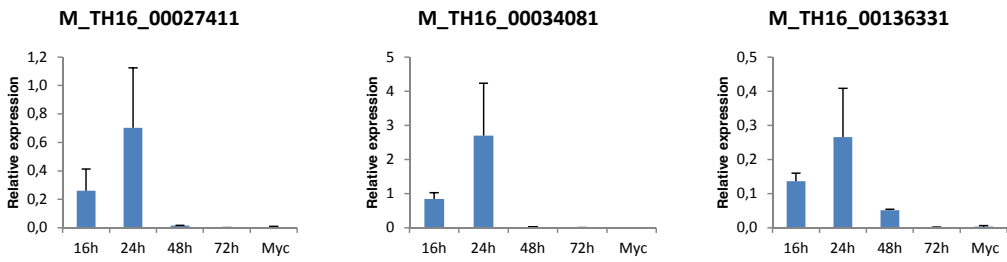
S9 Fig. continued



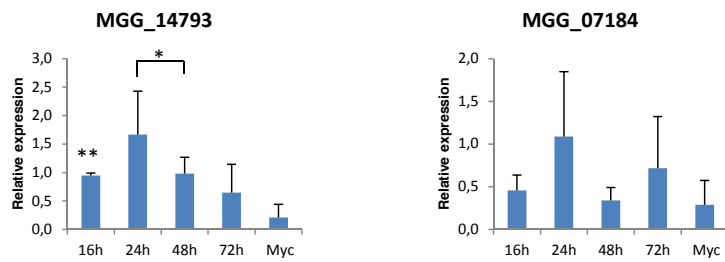
S9 Fig. continued



B

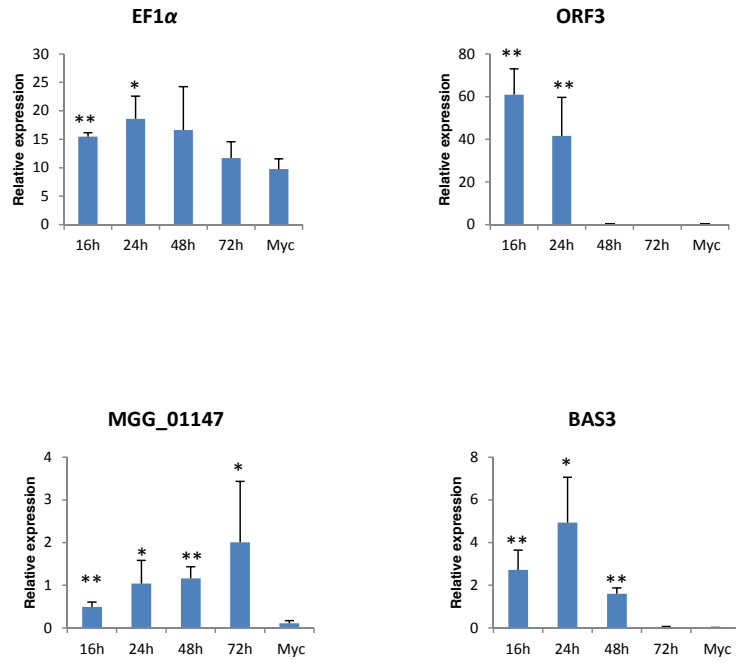


C

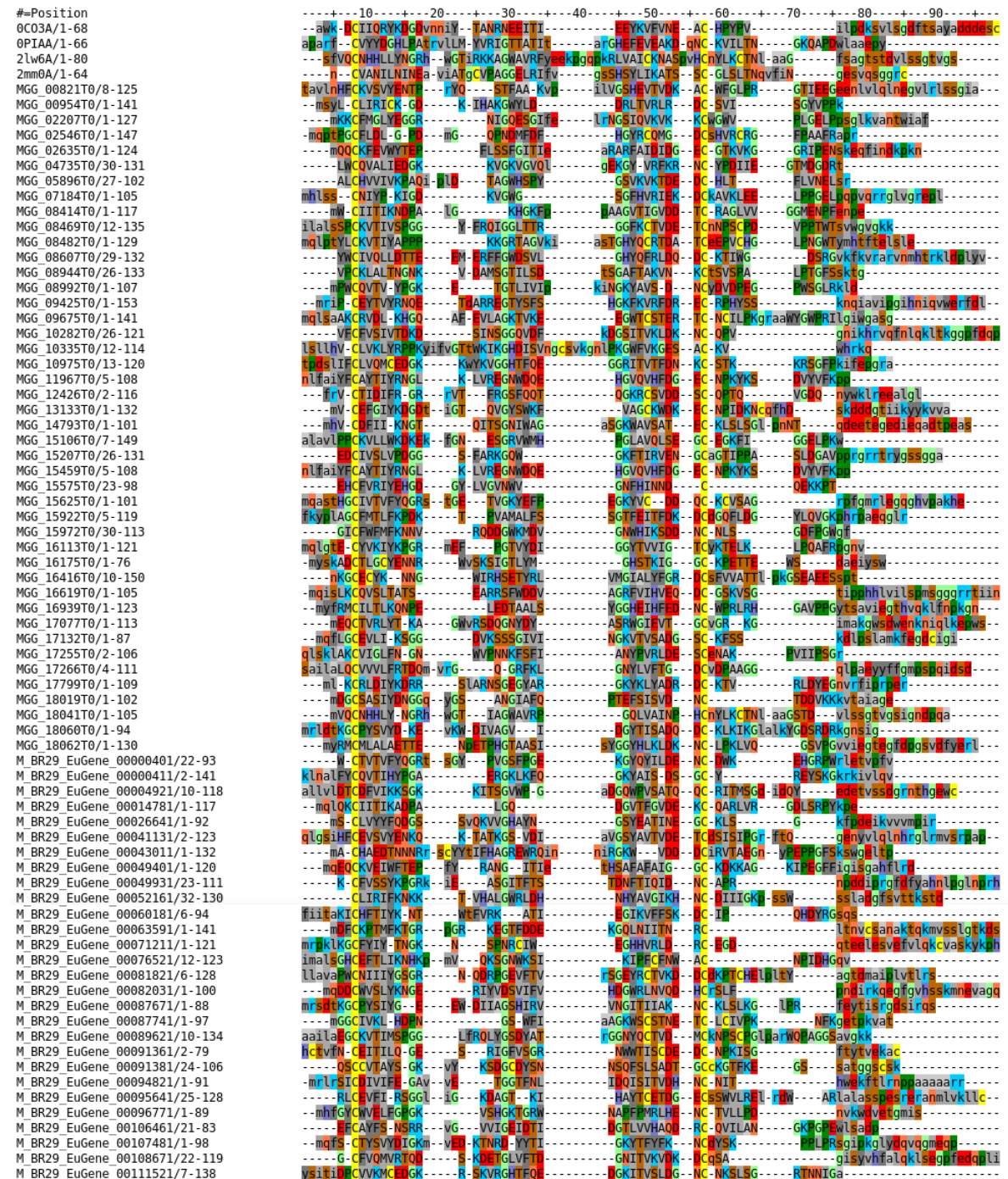


S9 Fig. continued

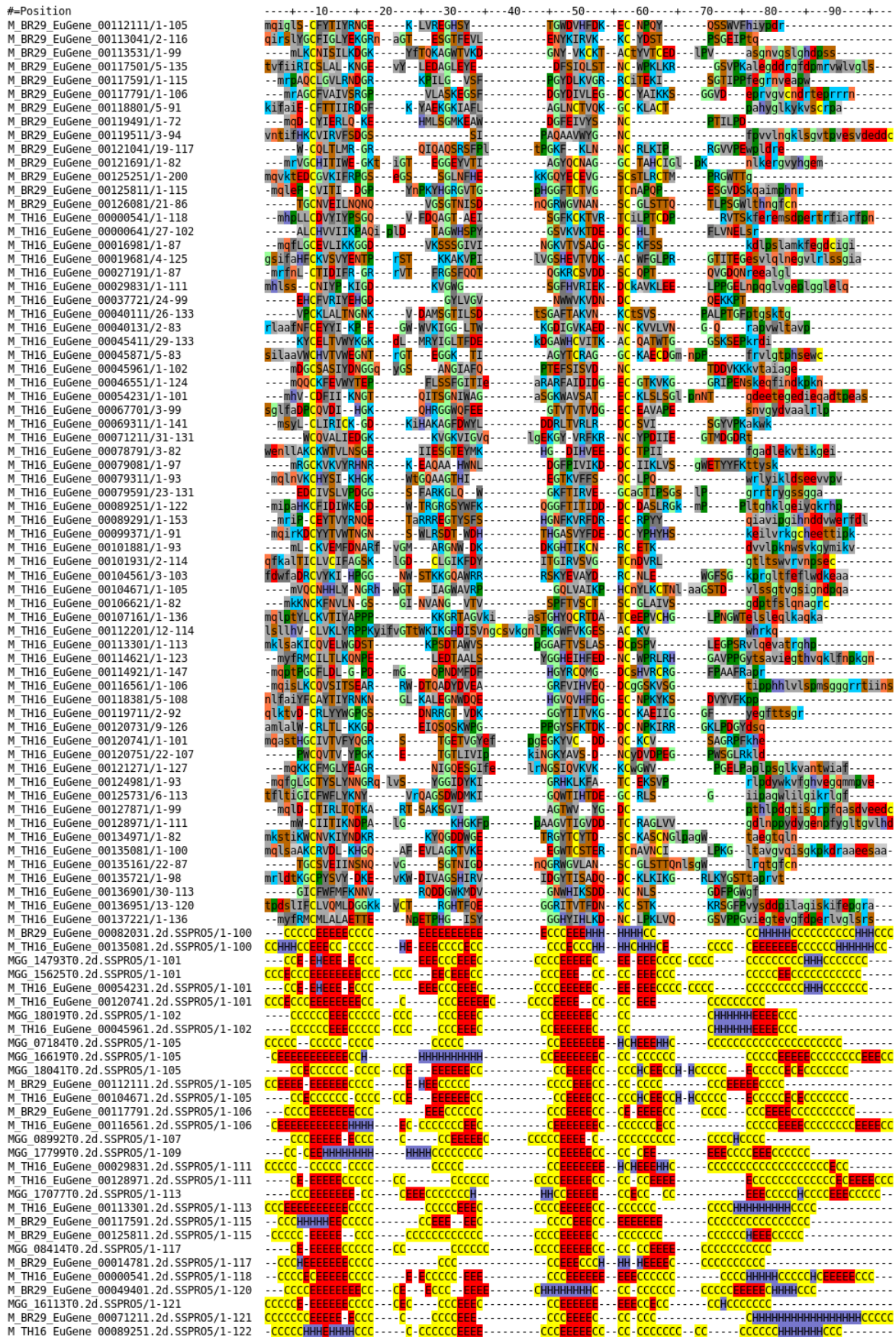
D



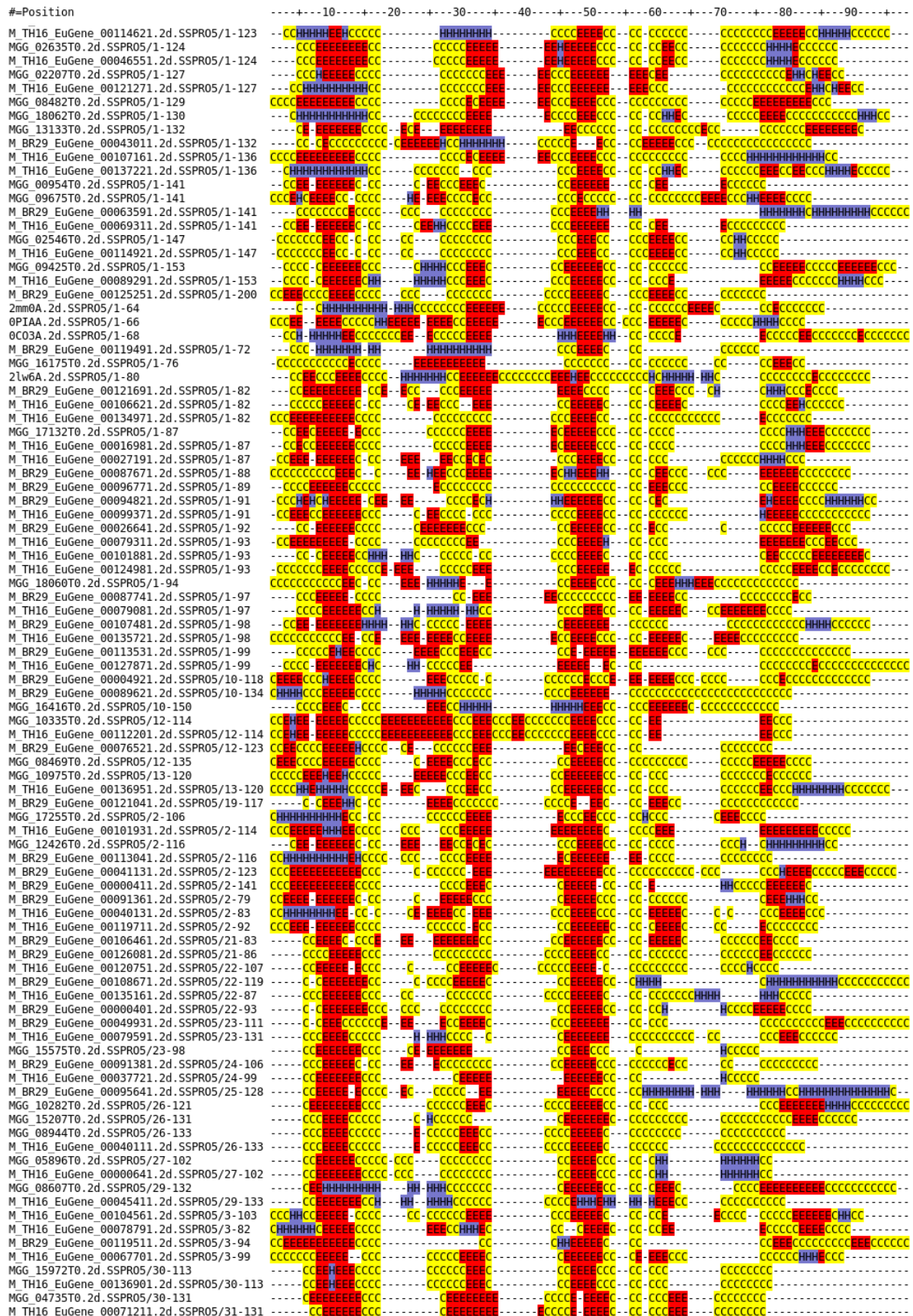
S10 Fig. Prediction of the secondary structure of *M. oryzae* MAX effectors. The secondary structures of the MAX-effectors from the 70–155, TH16 and BR29 genomes were predicted with SPRO5 [93]. The predictions are shown at the bottom of the figure and are aligned onto the corresponding primary sequence alignment shown at the top of the figure. Sequence identifiers for the secondary structure predictions are suffixed with « .2d.SSPRO5 ». Blue »H«, red »E« and yellow »C« correspond respectively to helix, extended sheet and coil predictions. The sequences of the 4 MAX effectors with experimentally determined structures are displayed at the top of the multiple sequence alignment and, for clarity, the alignment positions corresponding to shared gaps in the known structures were removed.



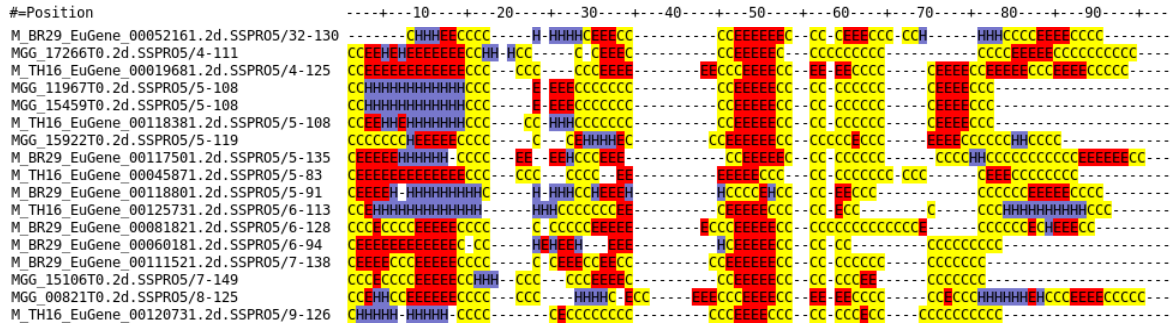
S10 Fig. continued



S10 Fig. continued



S10 Fig. continued



CHAPITRE 2

Cartographie de la surface d'interaction d'AVR-Pia et du domaine RGA5-RATX1

1. Introduction

La grande différence entre la susceptibilité et la résistance des plantes face aux pathogènes est majoritairement déterminée par les différences génétiques retrouvées dans les gènes de résistance (R), qui appartiennent à une famille de gènes codant des protéines de la classe des NB-LRR (Okuyama et al., 2011). Les gènes de NB-LRR sont très étudiés dans la génétique des plantes, d'une part parce qu'ils jouent un rôle clé dans la résistance de la plante face aux maladies et d'autre part du fait de leur mode unique d'évolution; les gènes de NB-LRR coévoluent avec les gènes du pathogène dans une course à l'armement incessante (Jones and Dangl, 2006).

1.1 Le couple de récepteur immunitaire RGA4 - RGA5

Les protéines de résistance multi-domaines (NB-LRR) produites par les gènes de résistance (R) sont impliquées dans la reconnaissance des effecteurs ou facteur d'avirulence (Avr) de pathogène. Dans le Chapitre 1, nous nous sommes particulièrement intéressées aux effecteurs AVR1-CO39 et AVR-Pia de *M. oryzae*. Au sein du génome du riz, les loci *Pi-CO39* et *Pia* confèrent respectivement la résistance envers les souches de *M. oryzae* exprimant le gène d'avirulence AVR1-CO39 ou AVR-Pia (Chauhan et al., 2002; Okuyama et al., 2011). Des analyses génétiques ont montré la colocalisation de locus *Pia* et *Pi-CO39* sur le chromosome n°11 du riz (Chauhan et al., 2002; Okuyama et al., 2011). C'est au niveau de ces loci que sont situés les gènes de résistance R. Dans les deux cas, sur ces endroits précis du chromosome sont retrouvés deux gènes de résistance adjacents RGA4 et RGA5, orientés dans des directions opposées et codant pour deux protéines NB-LRR de Type CNL (CC-NB-LRR). Il est intéressant de noter que la même paire de NB-LRR, formée par les protéines RGA4 et RGA5 est impliquée dans la reconnaissance des deux effecteurs AVR1-CO39 et AVR-Pia (Cesari et al., 2013; Okuyama et al., 2011).

L'analyse de différentes souches de *M. oryzae* a permis d'identifier un allèle inactif du gène AVR-Pia, nommé AVR-Pia-H3. Le polymorphisme d'AVR-Pia-H3 provoque une double mutation de la protéine, la phénylalanine 24 est changée en sérine et la tyrosine 46 en asparagine (F24S, T46N). Ces mutations induisent une perte de la reconnaissance par *Pia* (Cesari et al., 2013). A l'heure actuelle, il n'y a pas de données expérimentales expliquant pourquoi les deux gènes de NB-LRR sont indispensables au mécanisme de reconnaissance conduisant à la réponse hypersensible. L'une des

hypothèses postulée est que l'un des CNL soit un senseur pouvant interagir avec l'effecteur alors que l'autre CNL serait impliqué dans la transduction du signal (Cf Fig 2.1).

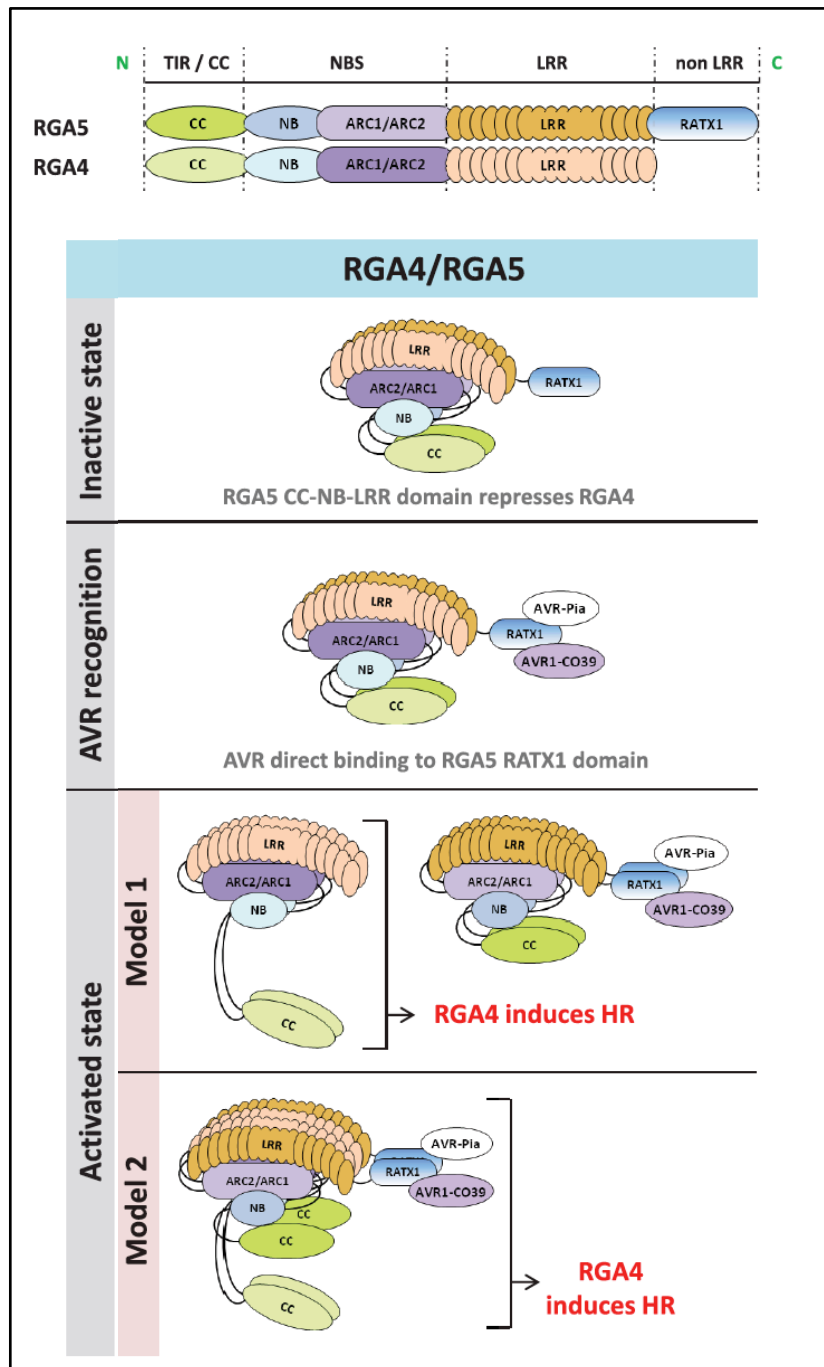


Fig 2.1 modèle fonctionnel de RGA4/RGA5 :

Lors de l'état inactif, l'hétéro-complexe RGA-4 (NLR acteur de la signalisation), RGA-5 (NLR senseur des effecteurs) adopte une conformation où RGA-5 réprime RGA-4. La reconnaissance des effecteurs AVR1-CO39 et AVR-Pia se fait par le domaine RATX1 situé à l'extrémité C-terminale du RGA-5. Après reconnaissance, l'hétéro-complexe est toujours présent. L'induction de la réponse hypersensible est présentée selon deux modèles hypothétiques :

Dans le modèle 1, l'hétéro-complexe se dissocie, la formation d'un dimère de RGA-4 induit la réponse hypersensible (HR).

Dans le modèle 2, la formation d'un tétramère permettrait l'induction de la HR par l'homo-complexe RGA4 toujours associé au complexe RGA5/AVR. (Cesari et al., 2014a)

1.2 Le domaine RATX1

De récentes études menées par nos collaborateurs (Cesari et al., 2013) ont démontré que les effecteurs AVR1-CO39 et AVR-Pia interagissent spécifiquement avec la partie C-terminale non-LRR de la protéine RGA5 (Cf fig 2.1). Des recherches par BLAST (Basic Local Alignment Search Tool) sur ce domaine ont permis d'identifier un domaine RATX1 homologue dans la protéine NB-LRR de riz PIKp-1 (Cesari et al., 2013). Contrairement à RGA5, ce domaine est situé en N-terminal de la protéine PIK, entre le Coiled-coil et le Nucleotide-Binding domain. Le domaine RATX1 est également appelé HMA pour Heavy-Metal associated domain. La caractéristique de ce domaine est de lier des métaux lourds ou des ions comme le Cu^{2+} ou le Zn^{2+} , par l'intermédiaire de deux résidus cystéines conservés. Ce domaine est impliqué dans le transport du métal ou les voies de détoxification (Bull and Cox, 1994). Cependant ces deux résidus ne sont ni conservés dans RATX1_{Pia}, ni dans RATX1_{Pik}.

D'une façon intéressante, comme pour RGA5, le domaine RATX1 de Pikp-1 est impliqué dans la reconnaissance de son effecteur AVR-Pik (Kanzaki et al., 2012). Cependant des études ont montré que chacun des domaines RATX1 est spécifique à son effecteur (Cesari et al., 2013). Ainsi une analyse structure-fonction permettrait d'identifier les mécanismes moléculaires impliqués dans la spécificité de cette reconnaissance. Dans cette optique, nous avons focalisé notre étude sur l'interaction entre AVR-Pia et RATX1_{Pia} *in vitro* par différentes techniques de biophysique (anisotropie de fluorescence, RMN et ITC) décrites dans la partie 2 de ce chapitre. Pour une meilleure compréhension du système, des études *in vivo* ont été menées en parallèle par nos collaborateurs. L'ensemble de ces résultats est présenté dans l'article en 3^{ème} partie de ce chapitre.

Récemment, la structure d'AVR-Pik en complexe avec le domaine RAXT1_{Pik} a été résolue par cristallographie (Maqbool et al., 2015). La structure d'AVR-Pik est commentée dans la première partie de la discussion de cette thèse. La comparaison entre les systèmes AVR-Pik/RAXT1_{Pik} et AVR-Pia/RATX1_{Pia} est abordée dans l'article présenté dans la partie 3.

2. Méthodes

L'étude de cette interaction *in vitro* nécessite de produire et de purifier les protéines recombinantes. La protéine RGA5 est une protéine multi domaine d'environ 120kDa. Pour cette étude, la protéine est découpée en plusieurs domaines au niveau de l'extrémité C-terminale, située après le domaine LRR et préalablement caractérisée comme étant impliquée dans l'interaction avec AVR1-CO39 et AVR-Pia.

La première stratégie est d'utiliser un tag MBP (maltose binding protein) en N-terminal des protéines pour faciliter leur solubilité (Cf Fig 2.2). La technique de production et de purification est décrite dans la section B du Matériels et Méthodes.

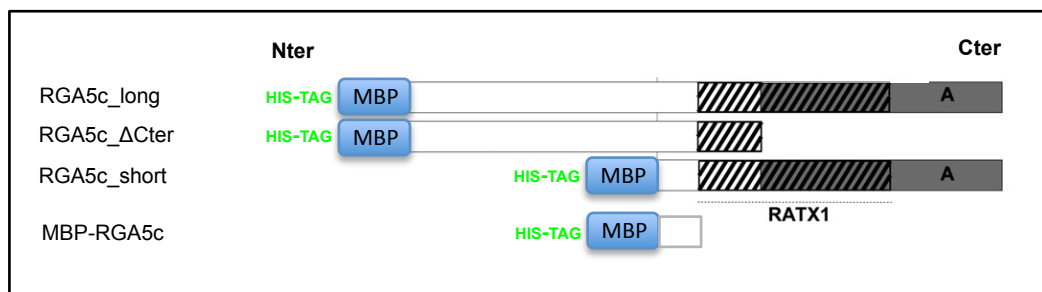


Fig 2.2 Schéma représentant les protéines recombinantes des différents domaines du C-ter de RGA5 :

Quatre protéines recombinantes ont été conçues selon la même stratégie de clonage, comprenant un tag histidine et un tag MBP (Matose Binding Protein) en N-terminal suivi de la protéine d'intérêt. Le détail des constructions est disponible dans la figure M.3 du matériels et méthodes.

Les découpages sont basés sur l'extrémité C-terminale de la protéine RGA5 exon A, dont l'interaction a été mise en évidence par double hybride (le domaine d'activation étant remplacé par les tag dans notre cas). RGA5c-long correspond à la partie C-terminale de RGA5-A caractérisée en double hybride comme impliquée dans l'interaction.

RGA5c_ΔCter correspond à la protéine RGA5c-long tronquée en C-terminal mais incluant une partie du N-terminal du domaine RATX1, n'interagissant pas en double hybride.

RGA5c_short correspond au N-terminal de la protéine RGA5c-long, incluant le domaine RATX1, et interagissant en double hybride.

La protéine MBP-RGA5 est un témoin négatif comprenant une soixantaine d'acides aminés constituant la partie C-terminale de RGA5.

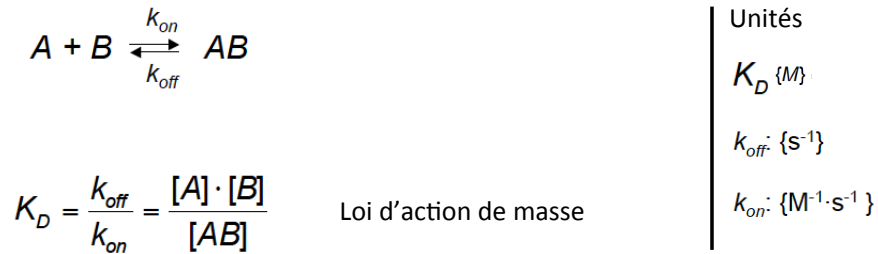
L'obtention de ces protéines a permis d'effectuer des tests d'interaction par anisotropie de fluorescence dont les résultats sont présentés dans la partie 2.2

Par la suite, l'utilisation d'une autre construction va se révéler nécessaire. En effet, la coupure enzymatique pour enlever le tag MBP n'est pas fructueuse, et les protéines ne semblent pas repliées. Aussi après vérification de l'intégrité des domaines par bioinformatique, une construction réduite au domaine RATX1 semble la meilleure stratégie pour obtenir une protéine repliée et soluble. La construction est réalisée à partir d'un gène synthétique possédant des codons optimisés pour la production dans *E.coli* et inséré dans un plasmide de type pET15b. Les détails de la production et de la purification de la protéine sont présentés dans la section C du matériels et méthodes.

2.1 Les Interactions biologiques

L'étude des interactions est capitale pour mieux comprendre le fonctionnement d'un système faisant intervenir plusieurs partenaires. En biologie structurale, plusieurs critères rentrent en compte pour caractériser une interaction. Ainsi l'affinité mesure l'intensité des interactions intermoléculaires

non covalentes entre deux molécules (protéine-protéine, protéine-ligand, protéine-ADN). La constante d'équilibre de dissociation, également appelée constante K_D , permet de décrire ce phénomène. Cette constante est décrite à l'équilibre, et à une température donnée par la loi d'action de masse et traduit l'affinité entre deux macromolécules A et B. Ainsi plus l'affinité est grande plus le K_D sera petit.



Cette constante est exprimée en molaire (M ou mol.l⁻¹), elle est reliée à deux paramètres cinétiques le k_{on} (M⁻¹ · s⁻¹) pour la formation et le k_{off} (s⁻¹) pour la dissociation.

La spécificité des interactions biologiques repose sur la structure tridimensionnelle, la nature physicochimique et la quantité d'interactions physiques (interactions électrostatiques, interactions de van der Waals, liaisons hydrogène) des deux partenaires.

La mesure des paramètres d'affinité peut être faite par différentes techniques. Dans notre cas nous avons utilisé l'anisotropie de fluorescence, la RMN et la calorimétrie pour réaliser plusieurs titrages.

2.2 Titration par anisotropie de fluorescence

L'avantage de cette technique est sa rapidité (quelques minutes) et sa sensibilité (des concentrations de l'ordre de quelques micro-molaires). L'anisotropie de fluorescence permet de mesurer des interactions protéine-protéine ou protéine-ligand. L'une des molécules est marquée par une sonde fluorescente et l'échantillon est excité par de la lumière polarisée. La liaison entre la molécule marquée et son partenaire va augmenter la polarisation de fluorescence, la mesure de cette augmentation permet de déterminer l'affinité de liaison.

La description de la méthode et du protocole utilisé est réalisée dans la partie B.3 du Matériel et Méthodes.

Le titrage d'AVR-Pia et d'Avr1-CO39 marqués à l'Alexa 488 par les différentes protéines dérivées de RGA5 montre qu'il existe une interaction entre les AVR et RGA5c_long et RGA5c_ΔCter, une

interaction est toujours visible avec RGA5c_short. Le témoin négatif MBP-RGA5c n'interagit plus. (Cf fig 2.3)

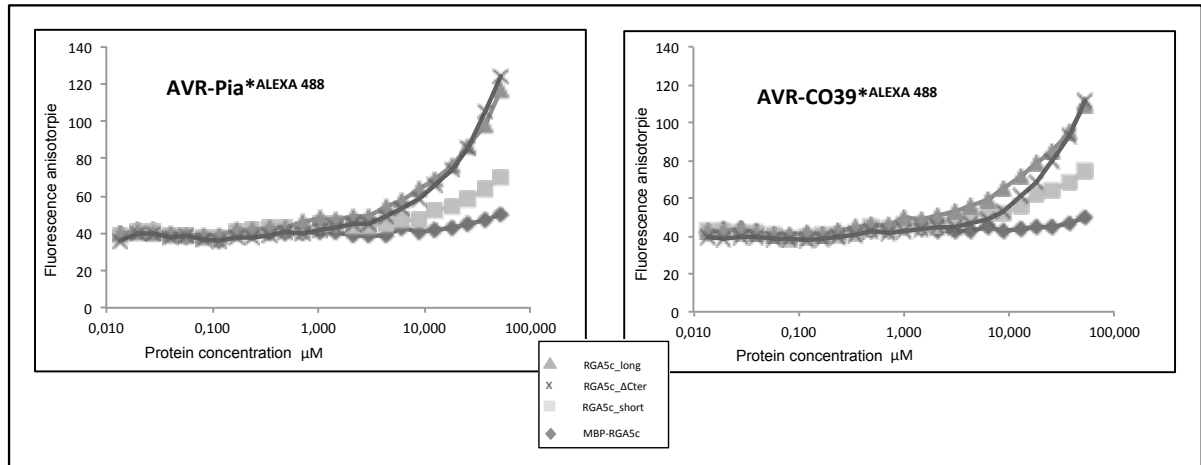


Fig 2.3 Mesure d'anisotropie de fluorescence des protéines AVR-Pia et AVR1-CO39 marquées à l'Alexa 488 en fonction de concentrations croissantes des différents constructions de protéine RGA5.

Cependant ces expériences ne nous donnent que des résultats partiels et plutôt qualitatifs. En effet l'augmentation de l'intensité de fluorescence à la fin des titrages est due à la forte concentration en protéines RGA5, et implique que le plateau de fin de réaction n'a pas pu être observé. Par conséquent il est impossible de déterminer une constante d'affinité. De plus il a été impossible de réaliser ces expériences avec des protéines clivées issues de la fusion avec la MBP car elles se dégradent trop vite suggérant qu'elles n'étaient pas repliées.

2.3 Titrage par RMN

La RMN peut être utilisée pour caractériser l'interface de l'interaction entre deux protéines. L'expérience à deux dimensions ^{15}N - ^1H -HSQC (Heteronuclear Single Quantum Correlation) permet d'obtenir une empreinte où chaque pic représente un résidu. Le titrage est réalisé en enregistrant une série d'HSQC, d'abord sur une protéine A marquée ^{15}N en absence et ensuite en présence d'une quantité croissante de protéine B non marquée. La présence du partenaire d'interaction B induit des variations des déplacements chimiques des noyaux ^1H et ^{15}N des acides aminés de la protéine A situés dans la zone d'interaction. L'avantage de cette technique, considérant que les pics de corrélation de l'HSQC sont attribués, est qu'elle donne une vision par acides aminés de la surface d'interaction. Le temps d'acquisition d'une expérience HSQC pour une protéine à 50µM est de l'ordre d'une heure. Le protocole utilisé est expliqué dans la partie expérimentale de l'article.

2.4 Titrage par calorimétrie

La titration par calorimétrie isotherme (ITC pour l'anglais *Isothermal Titration Calorimetry*) est une méthode de choix pour mesurer les propriétés thermodynamiques des interactions protéine-protéine. Cette méthode permet de mesurer la variation de température lors des interactions protéine-protéine en solution, et sans marquage préalable. Cette technique permet une détermination précise de la constante de dissociation (K_D), de la stœchiométrie (n) et de l'enthalpie (ΔH) de la réaction.

La description de la méthode et le protocole utilisé sont donnés dans la partie C.4 du Matériels et Méthodes et dans l'article suivant.

3. Article 2 : Recognition of *Magnaporthe oryzae* effector AVR-Pia by the decoy domain of the rice NLR immune receptor RGA5

Soumis à « The Plant Cell » le 09 juin 2016.

1 **Recognition of the *Magnaporthe oryzae* effector AVR-Pia by the decoy domain of**
2 **the rice NLR immune receptor RGA5**

3

4

5 Diana Ortiz^{a*}, Karine de Guillen^{b*}, Stella Cesari^c, Véronique Chalvon^a, Jérôme Gracy^b,
6 André Padilla^b, Thomas Kroj^{a,1}

7

8 ^a INRA, BGPI, Biology and Genetics of Plant-Pathogen Interactions, Campus
9 International de Baillarguet, Montpellier, France

10 ^b CNRS UMR 5048, INSERM U1054, Centre de Biochimie Structurale, Université
11 Montpellier, France

12 ^c CSIRO Agriculture Flagship, GPO Box 1600, Canberra ACT 2601, Australia

13

14 * These authors contributed equally to the work

15

16 ¹ **Corresponding author:** Thomas Kroj, thomas.kroj@supagro.inra.fr

17 The author responsible for distribution of materials integral to the findings presented in
18 this article in accordance with the policy described in the Instructions for Authors
19 (www.plantcell.org) is: Thomas Kroj (thomas.kroj@supagro.inra.fr).

20

21 **Short title:** Recognition of the effector AVR-Pia

22 Summary

23

24 Nucleotide-binding domain and leucine-rich repeat proteins (NLRs) are important
25 receptors in plant immunity that allow recognition of pathogen effectors. The rice NLR
26 RGA5 recognizes the *Magnaporthe oryzae* effector AVR-Pia through direct interaction.
27 Here, we gained detailed insights into the molecular and structural bases of AVR-Pia-
28 RGA5 interaction and the role of the RATX1 decoy domain of RGA5. NMR titration
29 combined with *in vitro* and *in vivo* protein-protein interaction analyses identified the AVR-
30 Pia interaction surface that binds to the RATX1 domain. Structure-informed AVR-Pia
31 mutants showed that, although AVR-Pia associates with additional sites in RGA5,
32 binding to the RATX1 domain is necessary for pathogen recognition, but can be of
33 moderate affinity. Therefore, RGA5-mediated resistance is highly resilient to mutations
34 in the effector. We propose a model that explains such robust effector recognition as a
35 consequence, and an advantage, of the combination of integrated decoy domains with
36 additional independent effector-NLR interactions.

37

38 INTRODUCTION

39 Plant disease resistance largely relies on inducible immune responses that are triggered
40 upon receptor-mediated recognition of pathogen molecules and that often involve a
41 localized programmed cell death called the hypersensitive response (HR). Particularly
42 important are cytoplasmic nucleotide-binding oligomerization domain-like receptors
43 (NLR) that present a multi-domain architecture composed of a C-terminal leucine-rich
44 repeat (LRR) domain and a central nucleotide-binding (NB-ARC) domain (Takken and
45 Goverse, 2012; Jacob et al., 2013; Qi and Innes, 2013). In most cases, they carry in
46 addition an N-terminal coiled-coil (CC) or TOLL/interleukin-1 receptor (TIR) domain that
47 have both been reported to mediate NLR homo complex formation and to be crucial for
48 the activation of downstream signalling (Bernoux et al., 2011; Maekawa et al., 2011).
49 Plant NLR proteins specifically recognize pathogen-derived effectors that act inside plant
50 cells (Cui et al., 2015).

51 Traditionally, both effector recognition and activation of resistance signalling are thought
52 to be mediated by single plant NLRs but recent studies revealed an increasing number
53 of cases where different NLRs cooperate in pathogen recognition and resistance (Eitas
54 and Dangl, 2010; Césari et al., 2014a). Frequently, the genes coding for these paired
55 NLRs occur in a paired, inverted tandem arrangement in the genome. In the few cases
56 that have been investigated in more detail, the NLR pairs seem to act as hetero-
57 complexes where only one of the paired NLRs acts directly in effector recognition while
58 the other is crucial for the activation of downstream signalling (Williams et al., 2014;
59 Césari et al., 2014b). In other cases, helper NLRs that act downstream of several NLRs
60 with different recognition specificities were shown to be required for resistance and
61 pathogen detection (Gabriëls et al., 2007; Bonardi et al., 2011; Wu et al., 2015).

62 In certain cases, plant NLRs recognize effectors in an indirect manner. They detect
63 either a modification of the effector's host target protein called 'guardee', or
64 modifications of a host protein that mimics the effector target and is called a 'decoy' (van
65 der Hoorn and Kamoun, 2008).

66 In other cases, effectors are recognized in a direct manner by binding, either alone, or in
67 complex with a cofactor that may be a guardee or adecoy, to the NLRs (Takken and
68 Goverse, 2012; Collier and Moffett, 2009). In these cases, the LRR domain plays a

69 crucial role in recognition specificity and was frequently shown to mediate direct effector
70 binding (Ellis et al., 2007; Krasileva et al., 2010; Jia et al., 2000). Alternatively, direct
71 effector recognition can also be mediated by non-canonical domains integrated into
72 NLRs at low frequencies (Kanzaki et al., 2012; Sarris et al., 2015; Maqbool et al., 2015;
73 Le Roux et al., 2015; Césari et al., 2013). Recent work lead to the hypothesis that these
74 highly diverse integrated domains are mimics of effector targets and can therefore be
75 considered as integrated decoy domains (Le Roux et al., 2015; Sarris et al., 2015;
76 Césari et al., 2014a). However, the molecular mechanisms of effector recognition by
77 integrated domains and eventual advantages of this mode of recognition remain largely
78 unknown.

79 Rice blast, caused by the fungus *Magnaporthe oryzae* is a highly destructive crop
80 disease and a serious threat for food security (Pennisi, 2010; Dean et al., 2012;
81 Skamnioti and Gurr, 2009). NLR-mediated pathogen recognition is the major mechanism
82 in rice blast resistance. Among 25 different blast resistance genes cloned over the last
83 20 years, 24 code for NLRs (Liu et al., 2014). A particular feature is that blast-resistance
84 is frequently conferred by paired-NLRs with clustered tandem organisation in the
85 genome. Among them, the NLR pair RGA4/RGA5 conferring *Pi-CO39* and *Pia*
86 resistances has been developed as a model for the molecular understanding of paired
87 NLRs (Okuyama et al., 2011; Césari et al., 2013, 2014b). In this pair, RGA4 acts as a
88 constitutively active disease resistance and cell death inducer that is repressed by
89 RGA5 in the absence of pathogen (Césari et al., 2014b). In addition to its repressor
90 function, RGA5 acts as a receptor for the *M. oryzae* effectors AVR1-CO39 and AVR-Pia.
91 Direct binding of RGA5 to these effectors results in de-repression of RGA4 and
92 activation of resistance signalling. Effector binding involves the unconventional C-
93 terminal domain of RGA5 that is related to ATX1 (RATX1 domain), a heavy metal-
94 associated (HMA) domain protein from *Saccharomyces cerevisiae*, that acts as a
95 cytoplasmic copper chaperone (Césari et al., 2013). Interestingly, the RATX1 domain of
96 RGA5 is dispensable for RGA4 repression and seems exclusively dedicated to effector
97 binding (Césari et al., 2014b). Since the rice RATX1/HMA protein Pi21 is a blast
98 susceptibility factor required for full disease development (Fukuoka et al., 2009), it has
99 been hypothesised that AVR1-CO39 and AVR-Pia target RATX1/HMA proteins for

100 disease development and that the RATX1 domain is an integrated decoy domain (Césari
101 et al., 2014b, 2013).

102 Interestingly, an HMA domain 53% identical to the RGA5 RATX1 domain is also present
103 in another rice NLR, Pik-1 that acts together with the NLR Pik-2 in the specific
104 recognition of the *M. oryzae* effector AVR-Pik. Like in RGA5, this domain acts by binding
105 directly the effector and is crucial for its recognition (Kanzaki et al., 2012). However, on
106 the contrary to the C-terminal RATX1 domain of RGA5, the HMA domain of Pik-1 is
107 located between the CC and NB domains indicating independent integration of the same
108 domains in the two unrelated NLRs (Césari et al., 2013). Recently, the determination of
109 the crystal structure of the AVR-PikD/Pikp-1 HMA domain complex allowed the precise
110 identification of the AVR-PikD surface mediating binding to the Pikp-1 HMA domain
111 (Maqbool et al., 2015).

112 Although AVR-Pik, AVR-Pia and AVR1-CO39 do not share sequence similarities, they
113 share a highly similar three dimensional structure characterized by a six β -sandwich fold
114 also present in two other effectors: AvrPiz-t from *M. oryzae* and ToxB from the wheat
115 pathogenic fungus *Pyrenophora tritici repentis* (Zhang et al., 2013; Guillen et al., 2015;
116 Maqbool et al., 2015; Nyarko et al., 2014). The corresponding, structurally related
117 Magnaporthe Avr and ToxB effectors were termed MAX effectors. MAX effectors are
118 present in other, sometimes only distantly related phytopathogenic fungi and the MAX
119 effector family underwent strong expansion in *M. oryzae* where it accounts for roughly
120 10% of the effectors (Guillen et al., 2015).

121 In the present study, we investigated the molecular and structural bases of AVR-Pia
122 recognition by RGA5 with a focus on the role of the RATX1 domain in effector-binding
123 and recognition. We show that AVR-Pia interacts with RGA5_{RATX1} domain through a
124 precise surface that shares some similarity, but also important differences, with the
125 HMA-binding surface of AVR-Pik. We demonstrate that binding to the RATX1 domain is
126 required for effector recognition but that strong reduction in binding strength is tolerated.
127 We also provide evidence that the RATX1 domain is not required for association of
128 AVR-Pia with RGA5 and that it interacts with additional sites in the NLR which could
129 explain the high tolerance of recognition to reduced AVR-Pia-RATX1 binding strength.
130 Based on our results, we propose a model illustrating advantages of effector recognition

131 by integrated decoy domains as well as additional simultaneously occurring interactions
132 with NLR receptors.

133

7

129

134 **RESULTS**

135

136 **The F24S and T46N substitutions in the non-recognized AVR-Pia-H3 allele affect**
137 **surface properties but not structure**

138 We previously described the naturally occurring *AVR-Pia* allele *AVR-Pia-H3* that carries
139 two non-synonymous polymorphisms leading to the F24S and T46N substitutions
140 (Césari et al., 2013). *M. oryzae* isolates carrying the *AVR-Pia-H3* allele are virulent on
141 rice varieties carrying the *Pia* resistance locus and AVR-Pia-H3 does not interact in Y2H
142 assays with the C-terminal part of the rice NLR immune receptor RGA5 containing the
143 RATX1 domain (RGA5_{C-ter}). The NMR structure of AVR-Pia showed that both the F24
144 and T46 residues are surface-exposed and suggested that the corresponding
145 substitutions only affect AVR-Pia surface properties without major structural
146 rearrangements (de Guillen et al., 2015).

147 To test this hypothesis, the structures of AVR-Pia-H3 and the single mutants AVR-
148 Pia^{F24S} or AVR-Pia^{T46N} were analyzed by NMR spectroscopy. We performed sequential
149 assignments using ¹⁵N- labeled AVR-Pia samples and the ¹³C α and ¹³C β assignments
150 were performed using ¹³C-¹H 2D experiments with a ¹³C-natural abundance sample in
151 D₂O (Supplemental Table 1). When compared to AVR-Pia wild type, ¹H-¹⁵N chemical
152 shifts differed more in AVR-Pia-H3 than in AVR-Pia^{F24S} or AVR-Pia^{T46N} single mutants
153 (Figure 1A). The NMR structure of AVR-Pia-H3 proved to be very similar to the structure
154 of AVR-Pia (PDB code 5JHJ) (Figure 1B, Supplemental Table 2). The backbone RMSD
155 for superposition of the AVR-Pia and AVR-Pia-H3 structures is 1.53 Å and drops to 0.93
156 Å when the β 1- β 2 loop is excluded and the superposition starts at residue R23. Like the
157 AVR-Pia wild type protein, AVR-Pia-H3 shows the MAX-effector topology characterized
158 by 6 antiparallel β -strands (Figure 1C). The ¹H-¹⁵N chemical shift data for AVR-Pia^{F24S}
159 and AVR-Pia^{T46N} indicate that both single mutants probably also keep the MAX-effector
160 fold (Figure 1A). The analysis of the relaxation data indicates that both AVR-Pia^{wt} and
161 AVR-Pia-H3 have rigid structures with average S2 values of 0.8 and similar S2 profiles
162 indicating similar protein dynamics (Supplemental Figure 1).

163 The 3D structure of AVR-Pia-H3 therefore supports that the F24S and T46N
164 substitutions do not result in conformational changes but rather alter AVR-Pia surface

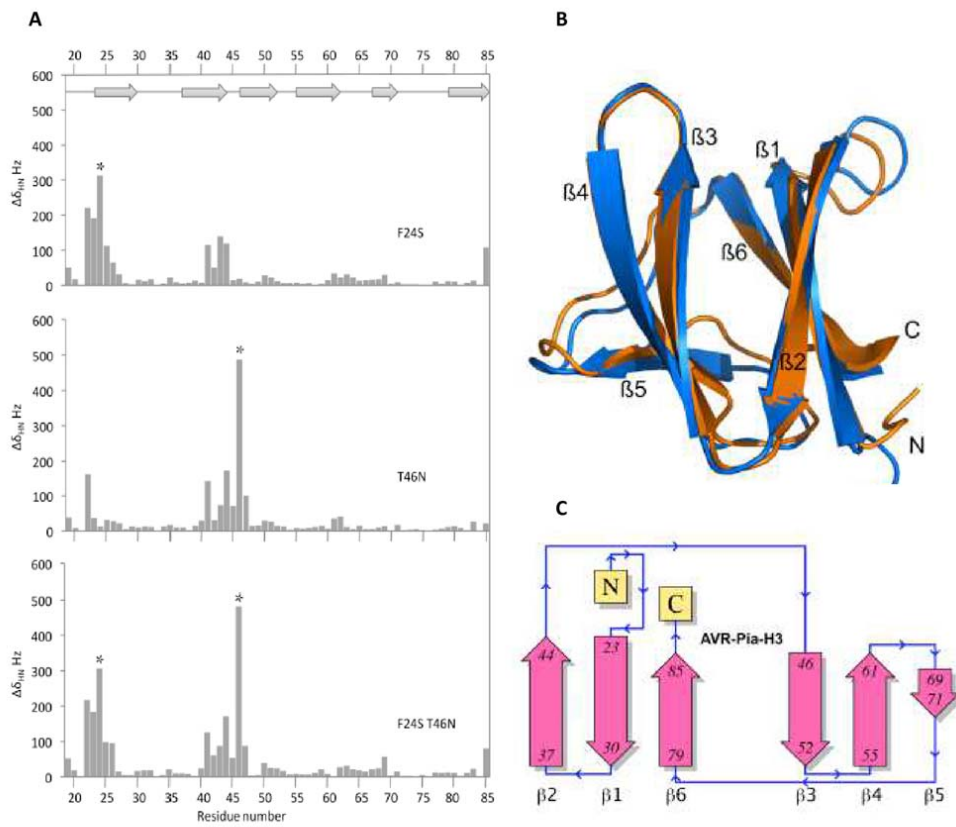


Figure 1. The AVR-Pia-H3 NMR structure is similar to the structure of wild-type AVR-Pia

165 properties required for strong interaction with the RATX1 domain of RGA5 (RGA5_{RATX1})
 166 and disease resistance activation.
 167

168 **AVR-Pia binds RGA5_{RATX1} with intermediate affinity**

169 To characterize the AVR-Pia/RGA5_{RATX1} interaction, *in vitro* binding assays with
170 recombinant RGA5_{RATX1} and AVR-Pia or AVR-Pia-H3 were performed using isothermal
171 calorimetry (ITC). For AVR-Pia, specific and direct binding to RGA5_{RATX1} with a one site
172 model and a K_d of 7,8 μM was detected (Figure 2A). For AVR-Pia-H3, no binding was
173 detected under identical conditions, suggesting that its affinity to RGA5_{RATX1} is at least
174 10 fold smaller than that of AVR-Pia.

175

176 **β-strands 2 and 3 and residues R23, F24, E56 and E58 constitute a candidate RGA5_{RATX1}-
177 interaction surface in AVR-Pia**

178 To test the hypothesis that the residues F24 and T46 are part of the AVR-Pia surface
179 mediating direct contacts to RGA5_{RATX1} and to identify other residues in direct contact to
180 RGA5_{RATX1} or located in the close vicinity of the binding interface, NMR-titration
181 experiments were performed. This technique consists in recording the ¹H-¹⁵N-HSQC
182 NMR spectra of ¹⁵N-labelled AVR-Pia in the presence of increasing amounts of
183 unlabeled RGA5_{RATX1}. When protein-protein binding occurs, it modifies the chemical
184 environment of the amino acids located on the binding surface and eventually mediating
185 the binding. This results in a change of the chemical shift in NMR experiments.
186 Depending on the transition rate of the complex formation, expressed by the exchange
187 rate constant k_{ex} , and the chemical shift difference $\Delta\omega$ between the unbound and bound
188 states ($\Delta\omega$ = difference between the resonance frequencies of the exchanging sites),
189 different exchange regimes occur. NMR titration shows that the AVR-Pia-RGA5_{RATX1}
190 complex is in slow exchange with $k_{ex} \ll \Delta\omega$ since separate resonances appear for
191 individual species (bound and unbound states) (Figure S2A). Residues with important
192 chemical shift changes between free AVR-Pia (R=0) and AVR-Pia bound to RGA5_{RATX1}
193 (molar ratio R=2) were almost exclusively surface-exposed and located in a region
194 formed essentially by β-strands 2 and 3 and including, in addition, residues R23 and F24
195 from β-strand 1 as well as E56 and E58 from β-strand 4 (Figures 2B and C). No peaks
196 were observed for residues Y27, V37, Y41, I44 and T51 in the complex. This candidate
197 interaction surface largely overlaps with an extended, solvent-exposed patch of
198 hydrophobic residues formed by F24, V26 and Y28 in β1, V37, L38 and Y41 in β2, and

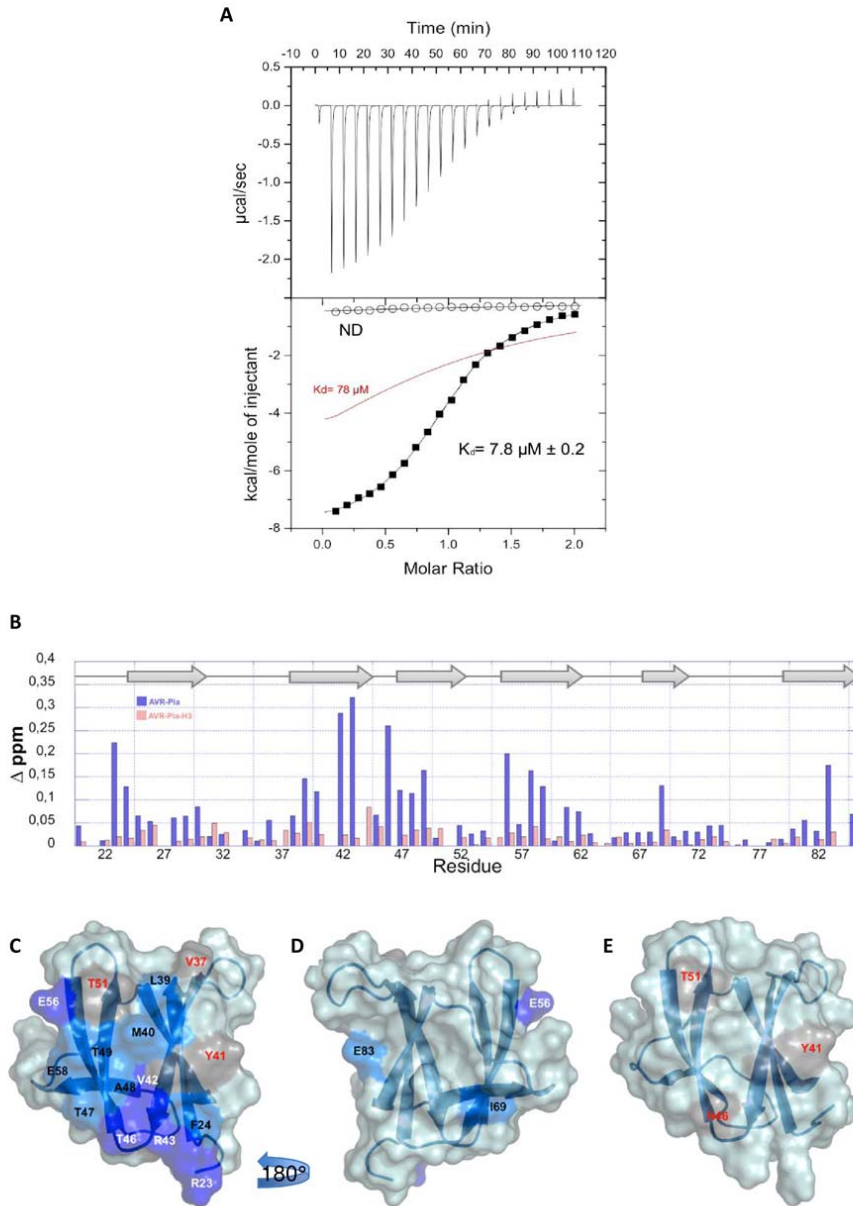


Figure 2. AVR-Pia binds RGA5_{RATX1} with intermediate affinity and a well-defined interaction surface

199 Y85 in β6. The residues on the other side of the AVR-Pia structure were not shifted in
 200 the NMR titration and seem therefore not involved in the interaction with RGA5_{RATX1}
 201 (Figure 2D). Two exceptions are E83, which probably senses a perturbation of the

202 residue Y41 that is close in space, and the I69 residue, which may be involved in local
203 conformational rearrangement of the short β 5 strand.

204 RGA5_{RATX1}-titration experiments were also performed with ¹⁵N-labelled AVR-Pia-H3 that
205 shows no binding in Y2H (Césari et al., 2013) and ITC analysis (Figure 2A). Spectral
206 perturbations were strongly reduced and only few and limited changes of chemical shifts
207 occurred when titrating AVR-Pia-H3 with RGA5_{RATX1} (Figure 2B and E and Figure S2B).
208 Signals for the R23, S24, V42, R43 and E83 residues were still observed at the end of
209 the titration while they were mostly lost at a molar ratio of 0.5 in the case of AVR-Pia
210 (Figure S2). Similarly, signals for E58, V59 and T47 were much less perturbed.
211 Nevertheless, the peaks for Y41, N46 and T51 were also perturbed indicating a weak
212 residual interaction between RGA5_{RATX1} and AVR-Pia-H3 (Figure 2B and E and
213 Supplemental Figure 2).

214 In summary, NMR titration identified a candidate interaction surface formed by β -strands
215 2 and 3 and including, in addition, residues R23, F24, E56 and E58 (Figure 2B). This
216 surface overlaps extensively with an extended hydrophobic patch on the AVR-Pia
217 surface that contains F24 and has T46 on its border and that may be crucial for
218 RGA5_{RATX1}-binding.

219

220 **Y2H experiments with structure-informed AVR-Pia mutants confirm an important** 221 **role of the candidate interaction surface in RGA5_{C-ter}-binding**

222 To test whether the AVR-Pia candidate interaction surface identified *in vitro* mediates
223 binding to RGA5_{C-ter} *in vivo*, we performed yeast two-hybrid (Y2H) assays using AVR-
224 Pia variants bearing point mutations in critical residues identified by NMR titration.
225 Individual surface-exposed hydrophobic (M40, Y41, Y85) or charged (R23, D29, R36,
226 E56, E58) amino acids, located in or at the border of the candidate interaction surface,
227 were replaced by alanine. In addition, naturally occurring AVR-Pia polymorphisms
228 located within the candidate interaction surface were tested: F24S and T46N from AVR-
229 Pia-H3 and R43G from AVR-Pia-H2 identified in *M. oryzae* isolates pathogenic on
230 *Setaria* species (Supplemental Figure 3A) (Césari et al., 2013). As controls, mutants
231 where surface-exposed charged residues located outside the candidate interaction
232 surface are replaced by alanine (D63A, K67A, K74A, D78A) were generated (Figure

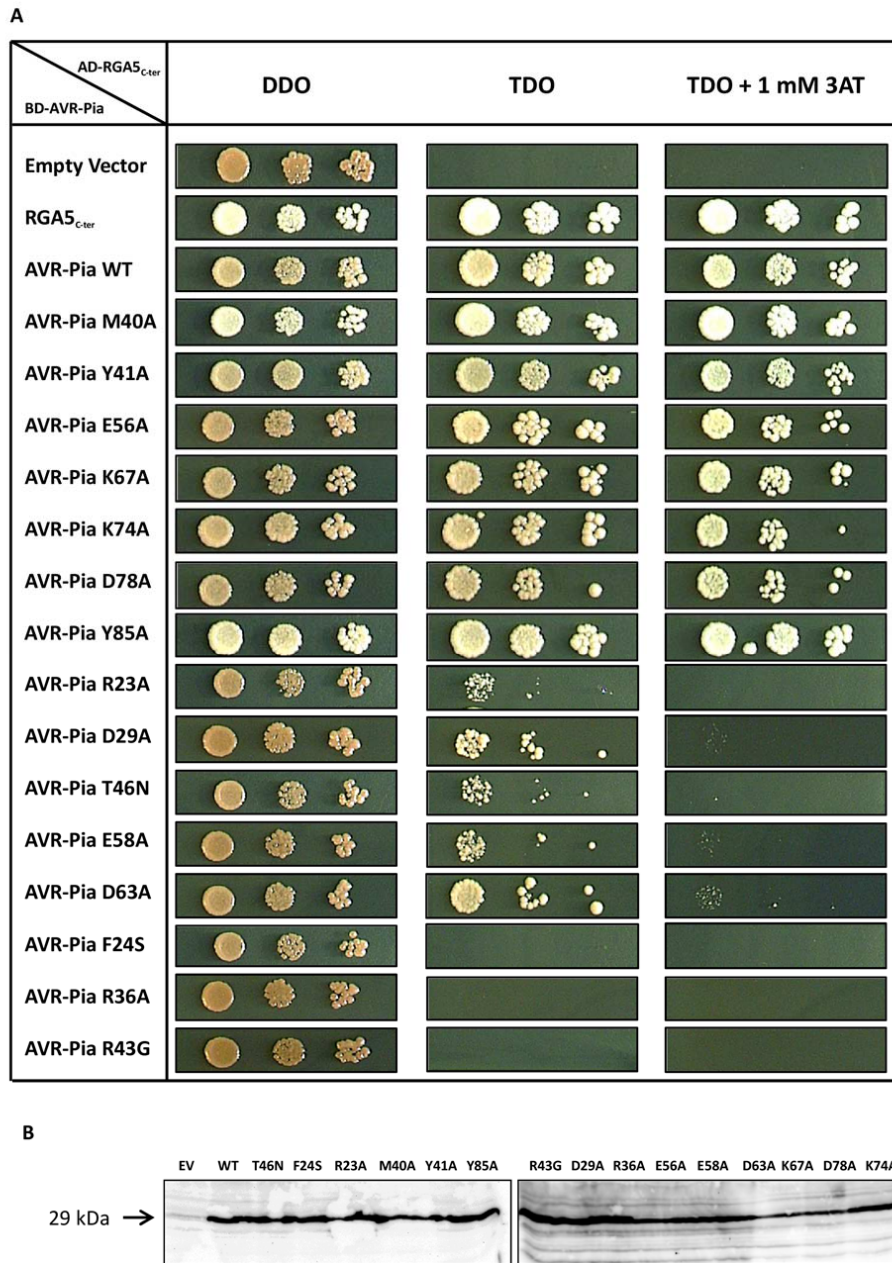


Figure 3. Mutations in the binding surface of AVR-Pia affect binding to RGA5_{C-ter} in yeast two hybrid assays.

233 3A).

234 As previously reported, yeasts co-expressing BD-AVR-Pia and AD-RGA5_{C-ter} or AD-
 235 RGA5_{C-ter} and BD-RGA5_{C-ter} grew on selective medium indicating physical binding

236 between AVR-Pia and RGA5_{C-ter}, and homo-interaction of the RGA5_{C-ter} domain (Figure
237 3A) (Cesari et al., 2013; Cesari et al., 2014). Yeasts co-expressing AD-RGA5_{C-ter} and
238 AVR-Pia^{F24S}, AVR-Pia^{R43G} or AVR-Pia^{R36A} fused to the BD domain did not grow on
239 selective medium indicating that these mutations abolish binding to RGA5_{C-ter}. Isolates
240 expressing BD-fusions of AVR-Pia variant carrying the mutation R23A, D29A, T46N,
241 E58A or D63A showed reduced growth compared to wild type BD-AVR-Pia indicating
242 that these mutations also affect AVR-Pia-RGA5_{C-ter} interaction. In contrast, yeast clones
243 expressing BD fusions of AVR-Pia^{Y41A}, AVR-Pia^{E56A}, AVR-Pia^{K67A}, AVR-Pia^{K74A} or AVR-
244 Pia^{D78A} showed similar growth as wild-type BD-AVR-Pia while AVR-Pia^{M40A} and AVR-
245 Pia^{Y85A} isolates showed stronger growth. All BD-AVR-Pia variants were expressed at
246 similar levels as the wild-type BD-AVR-Pia (Figure 3B). Taken together, these Y2H data
247 show that the replacement of all charged amino acids in the interaction surface, with the
248 exception of E56, either abolish or reduce binding of AVR-Pia to RGA5_{C-ter} while
249 exchanging hydrophobic residues within the interaction surface seems to abolish the
250 interaction in the case of F24S, or to increase the binding in the case of M40A and
251 Y85A.

252 To rule out that reduced binding of AVR-Pia mutants to RGA5_{C-ter} is due to major
253 changes in protein structure, the AVR-Pia mutants R23A, D29A, R36A, R43G and E58A
254 were expressed in *E. coli*, purified to homogeneity and analyzed by ¹H-1D-NMR
255 experiments (Supplemental Figure 3B). All mutant proteins showed similar spectra as
256 AVR-Pia wild-type, indicating that they are well structured and only locally disturbed.
257 Recombinant AVR-Pia-D63A could not be expressed.

258 Taken together, these results suggest that most residues of the AVR-Pia interaction
259 surface identified by NMR titration play an important role in RGA5_{C-ter}-binding.

260

261 **Co-IP experiments identify key residues in the AVR-Pia interaction surface that** 262 **are crucial for RGA5_{RATX1}-binding *in planta***

263 To investigate the role of the AVR-Pia interaction surface in *in planta* binding to RGA5_{C-ter},
264 co-immunoprecipitation (co-IP) experiments were performed. HA-tagged RGA5_{C-ter}
265 and YFP-tagged AVR-Pia mutants with reduced binding in Y2H were co-expressed in *N.*
266 *benthamiana* by *Agrobacterium tumefaciens*-mediated transient transformation. In

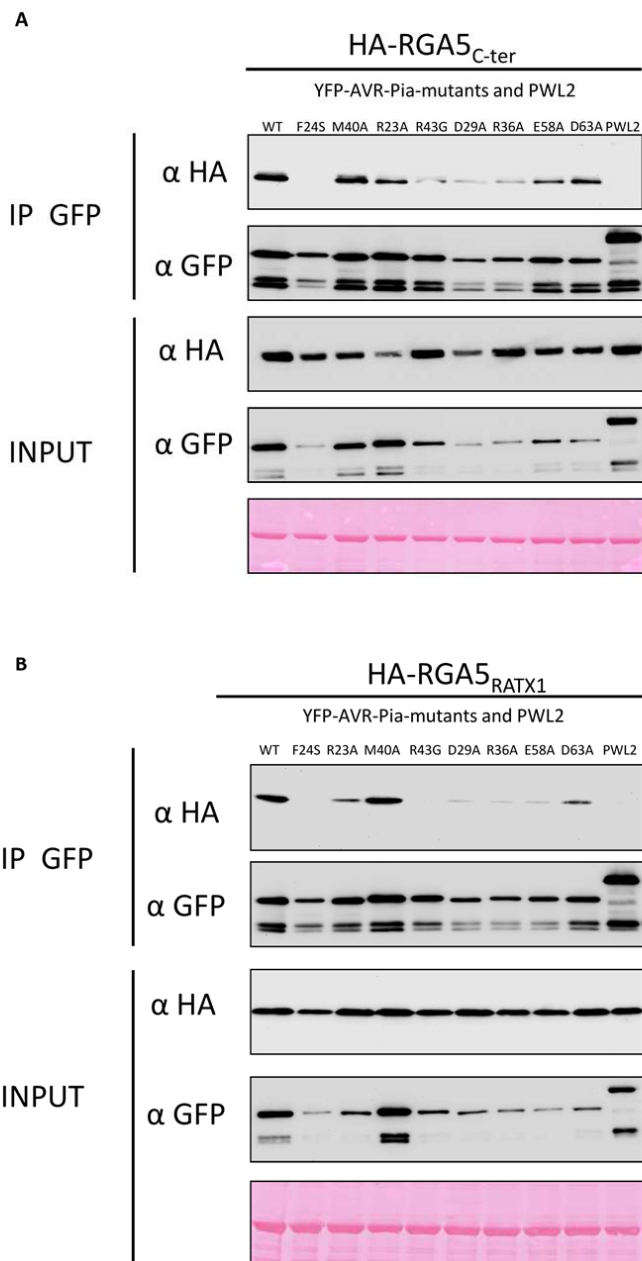


Figure 4. AVR-Pia mutants with reduced RGA5_{C-ter} binding in yeast are also impaired in binding to RGA5_{C-ter} and RGA5_{RATX1} in planta.

267 addition, AVR-Pia^{M40A} that, according to Y2H experiments, has increased affinity for
 268 RGA5_{C-ter} was also analyzed. As a negative control, a YFP fusion of the cytoplasmic *M.*
 269 *oryzae* effector PWL2 was used (Khang et al., 2010). Immunoblotting using anti-GFP

270 and anti-HA antibodies showed proper expression of all proteins (Figure 4). However,
271 AVR-Pia mutants with reduced binding to RGA5_{C-ter} in Y2H reproducibly accumulated at
272 lower levels than AVR-Pia^{wt}, while AVR-Pia^{M40A} was expressed at similar levels (Figure
273 4). All YFP fusion proteins were efficiently precipitated with anti-GFP antibodies but only
274 AVR-Pia^{M40A} co-precipitated RGA5_{C-ter} as strong as AVR-Pia^{wt}. The other mutants
275 showed various degrees of impairment ranging from slightly (AVR-Pia^{R23A}, AVR-Pia^{E58A}
276 and AVR-Pia^{D63A}) to strongly reduced (AVR-Pia^{D29A}, AVR-Pia^{R36A} and AVR-Pia^{R43G}) or
277 even completely abolished RGA5_{C-ter} co-precipitation (AVR-Pia^{F24S}) (Figure 4A). The
278 specificity of the interactions was confirmed with PWL2 that does not interact with
279 RGA5_{C-ter}.

280 It has previously been shown that the interaction of AVR-Pia with RGA5_{C-ter} relies on
281 interaction with the RATX1 domain (RGA5_{RATX1}) (Césari et al., 2013). To verify that
282 interaction specificities of the AVR-Pia mutants with RGA5_{C-ter} correlates with interaction
283 strength with RGA5_{RATX1}, co-IP experiments were performed using HA-tagged
284 RGA5_{RATX1}. AVR-Pia^{wt} and AVR-Pia^{M40A} strongly co-precipitated HA-RGA5_{RATX1}, while
285 the other mutants showed reduced (R23A and D63A), strongly reduced (D29A, R36A
286 and E58A) or no co-precipitation of RGA5_{RATX1} (F24S and R43G) (Figure 4B).

287 Taken together, these data indicate that AVR-Pia^{wt} and AVR-Pia^{M40A} strongly interact
288 with RGA5_{C-ter} and RGA5_{RATX1}, while mutants affected in direct binding to RGA5_{C-ter} in
289 Y2H showed reduced association with RGA5_{C-ter} and RGA5_{RATX1} *in planta*. Complete
290 absence of association with RGA5_{RATX1} for AVR-Pia^{F24S} and AVR-Pia^{R43G}, both *in planta*
291 and in Y2H, indicates a crucial role of these residues in the binding interface and
292 suggests that they are pivotal for AVR-Pia recognition.

293

294 **Direct binding to the RATX1 domain is required for AVR-Pia recognition**

295 To determine the role of the RATX1-binding surface of AVR-Pia in specific recognition
296 by the RGA4/RGA5 pair, AVR-Pia mutants were co-expressed in *N. benthamiana* with
297 RGA4/RGA5 and cell death activation was monitored. Since tagged versions of AVR-Pia
298 proved inactive in this assay, un-tagged AVR-Pia mutants were used. AVR-Pia mutants
299 with wildtype binding to RGA5_{RATX1} induced cell death indicating they are recognized by
300 RGA5/RGA4 (Supplemental Figure 4A and B). Weakly or non-binding mutants lost cell

301 death inducing activity but were also less abundant than AVR-Pia^{wt} or recognized AVR-
302 Pia mutants (Supplemental Figure 4A, B and C). Therefore, no clear conclusions can be
303 drawn for these mutants since lack of recognition may be due to low protein abundance.
304 Similar differences in the protein level of AVR-Pia mutants were previously observed
305 with YFP-tagged variants expressed in *N. benthamiana* (Figure 4) but not upon
306 expression in *E. coli* or yeast (Figure 3B). Therefore, differences in the accumulation of
307 AVR-Pia variants seem not related to an intrinsic destabilization of these proteins but
308 rather result from reduced stability in *N. benthamiana*.

309 Since transient heterologous experiments failed to determine the importance of the
310 binding of AVR-Pia to RGA5_{RATX1} for recognition and disease resistance, the biological
311 activity of AVR-Pia mutants was assayed in the homologous rice/*M. oryzae* system.
312 Transgenic *M. oryzae* isolates were generated that carried the different mutant alleles
313 under the control of the constitutive *RP27* promoter (*RP27*_{Pro}) (Bourett et al., 2002). As a
314 control, transgenic Guy11 isolates carrying a *RP27*_{Pro}:*mRFP* construct were generated
315 and proved to be fully virulent (Figures 5 and Supplemental Figure 5B). For three
316 different PCR-validated transgenic isolates per construct, the accumulation of AVR-Pia
317 variants was verified in culture filtrates by immunoblotting with anti AVR-Pia antibodies
318 (Supplemental Figure 5A). All AVR-Pia mutants were detected in at least one transgenic
319 isolate except AVR-Pia^{D63A} that may be instable in *M. oryzae*. For AVR-Pia^{D29A} and
320 AVR-Pia^{E58A}, only two and one isolate expressed the mutant protein (Figure S5A).

321 The transgenic isolates were analyzed on the rice cultivars Kitaake carrying the *Pia*
322 locus and Maratelli lacking *Pia*. All isolates were highly virulent on Maratelli indicating
323 that they were not affected in virulence (Supplemental Figure 5B). On Kitaake plants, the
324 isolates expressing AVR-Pia^{wt}, AVR-Pia^{R23A}, AVR-Pia^{D29A}, AVR-Pia^{R36A} or AVR-Pia^{E58A}
325 were completely avirulent and produced either no symptoms or small HR lesions
326 characteristic of resistance (Figure 5 and Supplemental Figure 5B). This indicates that
327 these AVR-Pia variants are fully active and recognized by RGA4/RGA5. Consistent with
328 the absence of protein expression, AVR-Pia^{D63A} isolates did not induce resistance and
329 were fully virulent on Kitaake plants. Isolates producing AVR-Pia^{R43G} were partially
330 virulent and formed disease lesions characterized by a grey center that were however
331 smaller and less frequent than those provoked by the control mRFP isolates. Isolates

A

	<i>M. oryzae</i> isolates								
	Controls		AVR-Pia mutants						
	AVR-Pia	mRFP	R23A	D29A	R36A	R43G	E58A	F24S	D63A
Kitaake (Pia)									
Avirulence activity	Active	Control	Active	Active	Active	Reduced activity		Inactive	---
α-AVR-Pia (7,5 KDa) Immunoblot									

B

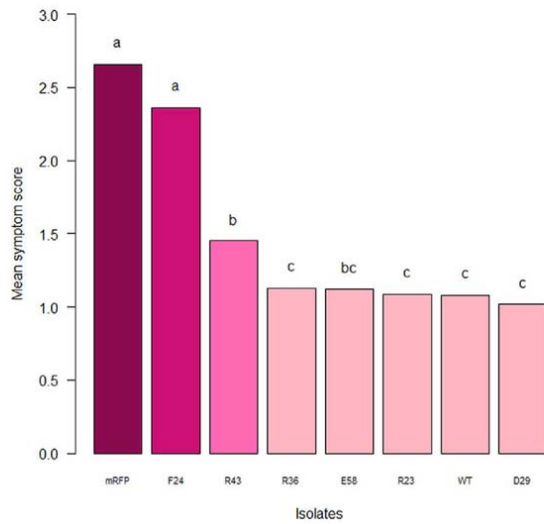


Figure 5. Effector recognition by RGA5 requires binding to the RATX1 domain

332 expressing AVR-Pia^{F24S} were highly virulent on Kitaake and produced large numbers of
 333 disease lesions (Figures 5 and Supplemental Figure5B).
 334 Taken together, these results indicate that interaction of AVR-Pia with the RGA5_{RATX1}

335 domain is required for recognition but that a reduction of this interaction as in AVR-
336 Pia^{R23A}, AVR-Pia^{D29A}, or AVR-Pia^{E58A} does not impair recognition. Only the R43G and
337 F24S polymorphisms that abolished RGA5_{RATX1} interaction both *in planta* and in yeast
338 affected AVR-Pia recognition, with AVR-Pia^{F24S} being completely inactive.

339

340 **AVR-Pia interacts with RGA5 outside of the RATX1 domain**

341 The high resilience of *RGA4/RGA5*-mediated AVR-Pia recognition to reduction of AVR-
342 Pia-RGA5_{RATX1} interaction strength suggested that AVR-Pia might interact with
343 additional sites in RGA5. To test this hypothesis, *in planta* association of the AVR-Pia
344 mutants with the RGA5 full-length protein was assayed by co-IP. All AVR-Pia mutants,
345 including AVR-Pia^{F24S} and AVR-Pia^{R43G} co-precipitated RGA5 as efficiently as AVR-Pia^{wt}
346 (Figure 6A). This indicates that lack of binding to RGA5_{RATX1} does not abolish
347 association with RGA5. To test whether binding of AVR-Pia to RGA5 is truly
348 independent of the RATX1 domain, association of AVR-Pia with an RGA5 construct
349 lacking the RATX1 domain (RGA5_{ΔRATX1}) was tested by co-IP. All AVR-Pia variants co-
350 precipitated RGA5_{ΔRATX1} (Figure 6B) and AVR-Pia mutants with reduced or no binding to
351 RGA5_{RATX1} interacted as strongly with RGA5_{ΔRATX1} as AVR-Pia^{wt} demonstrating that the
352 RATX1 domain is not necessary for formation of RGA5/AVR-Pia complexes. These
353 results suggest that AVR-Pia interacts with additional sites in RGA5 outside of the
354 RATX1 domain and that the region of AVR-Pia that mediates interaction with
355 RGA5_{ΔRATX1} lies outside of the RATX1-binding surface.

356 It was previously shown that RGA5_{ΔRATX1} inhibits RGA4-triggered cell death and that
357 therefore the RATX1 domain is not required for RGA5-mediated repression of RGA4
358 (Césari et al., 2014b). Since AVR-Pia still associates with RGA5_{ΔRATX1} *in planta*, we
359 tested whether AVR-Pia would be recognized by RGA5_{ΔRATX1}/RGA4 and trigger cell
360 death independently of the RATX1 domain. Neither co-expression of RGA4, RGA5_{ΔRATX1}
361 and AVR-Pia, nor expression of these three proteins together with the isolated RATX1
362 domain triggered cell death (Figure S6). This indicates that interaction of AVR-Pia with
363 regions outside of the RATX1 domain is not sufficient to release RGA5-mediated RGA4
364 repression and further confirms that binding of AVR-Pia to the RGA5_{RATX1} is required for
365 de-repression of RGA4. In addition, these results suggest that AVR-Pia has to interact

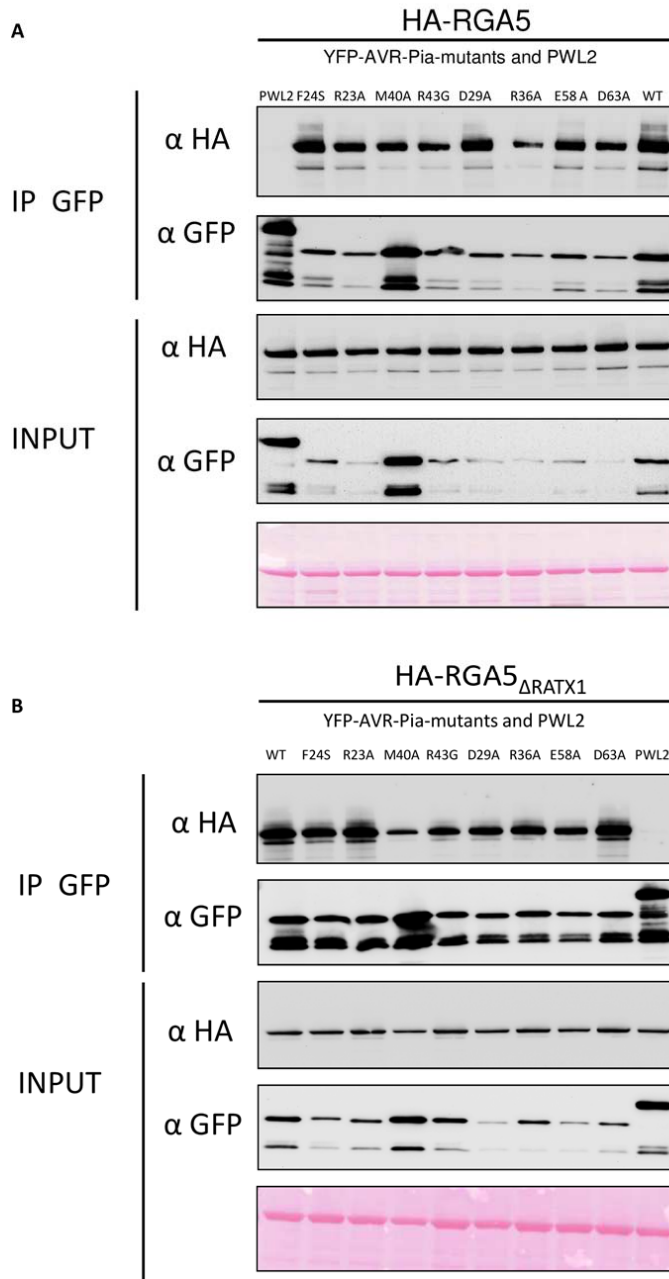


Figure 6. AVR-Pia interacts with RGA5 outside the RATX1 domain

366 with the RATX1 domain in the context of the full length RGA5 protein since an isolated
 367 RATX1 domain does not complement RGA5_{ΔRATX1} for AVR-Pia recognition.

368

21

143

369 DISCUSSION

370

371 Identification of a RGA5_{RATX1}-binding surface in AVR-Pia

372 In this study, we provide evidence that AVR-Pia interacts with the RATX1 domain of
373 RGA5 through a precise binding surface consisting of β -strands 2 and 3, residues R23
374 and F24 from β -strand 1, and residues E56 and E58 from β -strand 4 (Figure 2C). This
375 interaction surface, identified by NMR titration experiments with recombinant AVR-Pia
376 and the RATX1 domain, was confirmed by mutant analysis. Indeed, replacement of
377 residues R23, F24, D29, R43, T46 or E58 strongly reduced or abolished binding to
378 RGA5_{RATX1} in Y2H and Co-IP experiments while replacement of residues M40 and Y85
379 increased interaction in Y2H (Figures 3 and 4). Substitutions outside of the candidate
380 interaction surface had no impact on binding with the exception of residues R36 and
381 D63. Residue R36 is located in the loop joining β 1 and β 2 and might also be involved in
382 RATX1-binding. Alternatively, it may play an important role in defining the positions of β -
383 strands 1, 2 and 6 through the salt bridge it forms with residue E83 in β 6. The D63A
384 polymorphism seems to destabilize the overall structure since AVR-Pia^{D63A} could not be
385 expressed in *E. coli* or *M. oryzae*. Actually, D63 seems important to structure the loop
386 β 4- β 5 as its side-chain carboxyl group forms a hydrogen bond with the side-chain amid
387 group of N65. This may be required to position properly cysteine C66 that forms a
388 disulfide bridge with C25 linking the two β sheets β 1, β 2, β 6 and β 3, β 4, β 5 and thereby
389 impact global folding.

390

**391 AVR-Pia and AVR-PikD have distinct RATX1/HMA-binding surfaces that are
392 situated at similar positions**

393 The 3 dimensional structure of a AVR-PikD-Pikp-1_{HMA} complex was determined by
394 crystallography and showed that, like the formation of the AVR-Pia-RGA5_{RATX1} complex,
395 binding of AVR-PikD to Pikp-1_{HMA} involves β -strands 2 and 3 (Supplemental Figure 7A)
396 (Maqbool et al., 2015). However, in AVR-PikD, the residues of β -strand 2 that are crucial
397 for Pikp-1_{HMA}-binding, R64 and D66, are charged and establish, respectively, hydrogen
398 bonds and salt bridge interactions. In contrast, in AVR-Pia, surface-exposed residues of
399 β -strand 2 are hydrophobic and probably establish hydrophobic interactions (Figure 2C).

400 In addition, unlike AVR-Pia, AVR-PikD possesses an N-terminal extension of 32 amino
401 acids that is crucial for physical binding to Pikp-1_{HMA} and recognition by Pikp-1/Pikp-2
402 (Supplemental Figure7B). In particular, residue H46 from this extension establishes
403 important interactions with matching residues in Pikp-1_{HMA}. These interactions are
404 necessary for binding and, together with the neighboring residues P47 and G48, for
405 matching specificities with alleles of Pikp-1 (Kanzaki et al., 2012; Maqbool et al., 2015).
406 These residues are missing in AVR-Pia but similarly important interactions are
407 established with the amino acid F24 from the very hydrophobic β -strand 1. Therefore,
408 recognition of the sequence-unrelated, but structurally similar, effectors AVR-Pia and
409 AVR-PikD seems to involve similar structural elements but relies on distinct and highly
410 specific mechanisms.

411 Whether the effector interaction surfaces of the RATX1/HMA domains of RGA5 and
412 Pikp-1 are similar or completely different remains an open question. Crystal structures
413 show that Pikp-1_{HMA}, has a typical HMA α/β - sandwich fold composed of two α -helices
414 and a 4-stranded anti-parallel β -sheet, that mediates interaction with AVR-PikD
415 (Supplemental Figure 7C) (Maqbool et al., 2015). We used molecular modeling to
416 evaluate whether AVR-Pia-binding may involve similar regions in RGA5_{RATX1} but no
417 consensus docking model could be generated for the AVR-Pia-RGA5_{RATX1} complex
418 (Supplemental Figure 7C). Interestingly, none of the docking models predicted an
419 interaction surface in RGA5_{RATX1} similar to the effector-binding surface of Pikp-1_{HMA}.
420 This suggests that the RGA5_{RATX1}-AVR-Pia complex differs significantly from the Pikp-
421 1_{HMA}-AVR-Pik-D complex.

422 Taken together, recognition of the structurally similar MAX effectors AVR-Pia and AVR-
423 Pik by independently acquired NLR-integrated HMA domains seems to rely on distinct
424 molecular mechanisms. Future work is required to test this hypothesis through functional
425 studies of the Pikp-1_{HMA} interaction surface identified by structural analysis and the
426 identification of the surface that mediates effector binding in RGA5_{RATX1}.

427

428 **Binding of AVR-Pia to the integrated RATX1 domain is required for recognition** 429 **but of moderate affinity**

430 The mutants, AVR-Pia^{F24S} and AVR-Pia^{R43A}, showed drastically reduced RGA5_{RATX1}-

431 binding and triggered, respectively, no or reduced resistance indicating that the AVR-
432 Pia-RGA5_{RATX1} interaction is required for RGA4/RGA5-mediated recognition. The
433 presence of these polymorphisms in naturally occurring AVR-Pia alleles (Ribot et al.,
434 2013) suggests that *AVR-Pia* is undergoing selection for mutations in the RATX1-
435 interaction surface and escape from RGA4/RGA5-mediated recognition. These results
436 therefore provide further support for a crucial role of non-conventional, integrated decoy
437 domains in effector recognition and NLR specificity.

438 However, we also found high resilience of AVR-Pia recognition to a reduction in
439 RGA5_{RATX1}-binding strength since weakly-binding AVR-Pia mutants were still able to
440 trigger resistance. Similar observation was made in the case of AVR-PikD, the only other
441 case where the affinity of an effector to the integrated decoy domain of its NLR receptor
442 has been determined (Maqbool et al., 2015). Indeed, AVR-PikD^{A67D} and AVR-PikD^{P47A}
443 ^{G48D} mutants showed drastically reduced binding to Pikp-1_{HMA} but were nevertheless
444 perfectly well recognized by Pik-1/Pik-2.

445 A possible explanation for this tolerance to a reduction in the affinity between effectors
446 and integrated decoys could be that effectors interact with multiple independent sites in
447 NLR receptors. Indeed, our study suggests that, besides the RATX1 domain, AVR-Pia
448 interacts with other, not yet defined, regions in RGA5. In the simplest case, this
449 interaction relies on direct physical binding, but since it was solely detected by co-IP
450 experiments, it cannot be excluded to be indirect and to involve additional co-factors.
451 This interaction seems mediated by other AVR-Pia surfaces than those involved in
452 RGA5_{RATX1}-binding since mutants with reduced binding to RGA5_{RATX1} are not affected in
453 interaction with RGA5_{ΔRATX1}. Since the RATX1 domain is covalently linked to the rest of
454 the RGA5 receptor, AVR-Pia-binding to these other sites has the potential to increase
455 the overall effector binding affinity to RGA5 despite the low affinity binding to RGA5_{RATX1}
456 ($K_d=7\mu\text{M}$). In this context, further mutation-induced reduction of AVR-Pia-affinity toward
457 the RATX1 domain may not have a dramatic effect unless it completely abolishes AVR-
458 Pia/RGA5_{RATX1} interaction. This situation highlights an advantage of the integration of
459 the decoy domain into the NLR receptor over a situation where the decoy is a separate
460 molecule and has to bind to the effector before subsequent binding to the NLR receptor.
461 In the latter case, low affinity of the effector-decoy interaction leads to drastically

462 reduced receptor occupancy and renders the corresponding resistance very vulnerable
463 to effector mutations affecting decoy binding.

464

465 **Interaction of effectors with multiple independent sites is a hallmark of NLR**
466 **receptor activation**

467 Effector recognition by RGA4/RGA5 differs from other well-studied NLR models. Indeed,
468 RGA5 has no inherent signaling activity and functions, on the one hand, by repressing
469 RGA4 signaling activity and, on the other, by releasing repression upon AVR-Pia binding
470 (Césari et al., 2014a). Interestingly, the RATX1 domain is only required for de-repression
471 and not for repression (Supplemental Figure 6) (Césari et al., 2014b). Providing the
472 RATX1 domain separately in the presence of RGA5 Δ RATX1 and AVR-Pia does not relieve
473 repression despite the fact that AVR-Pia interacts with the separate partners,
474 RGA5 Δ RATX1 and RGA5^{RATX1}. To explain this result, we propose the hypothesis that
475 simultaneous binding of AVR-Pia to different sites in RGA5, including the RATX1
476 domain, is required to trap RGA5 in a conformation unable to repress RGA4 (Figure 7).
477 That effectors have to establish simultaneously several independent interactions with
478 NLRs or NLR_S and co-factors to be recognized and trigger resistance has been
479 frequently observed with effectors from various origins (Collier and Moffett, 2009). In the
480 *Pseudomonas syringae* effector AvrRPS4 two different surface areas on opposite and
481 distant sites of the molecule are required for recognition by the RRS1/RPS4 pair (Sohn
482 et al., 2012). One of these sites is crucial for binding to the integrated WRKY decoy
483 domain of RRS1 while the other seems to interact with other not yet identified regions in
484 RRS1 (Sarris et al., 2015). Similarly, recognition of the *Hyaloperonospora arabidopsidis*
485 effector ATR1-EMOY2 by the NLRs RPP1-NdA or RPP1-WsB from *A. thaliana* relies on
486 two different surface areas from two different domains and on opposite sides of the
487 molecule suggesting simultaneous interaction with independent binding sides in RPP1-
488 NdA and RPP1WsB (Chou et al., 2011; Steinbrenner et al., 2015). Also in NLRs that
489 recognize effector-co-factor complexes, simultaneous binding of these complexes to
490 different parts of the NLR, generally involving the N-terminus and the LRR have been
491 frequently described (Collier and Moffett, 2009). Therefore, we propose the hypothesis
492 that effectors or effector-co-factor complexes forcing or trapping NLRs in an activated

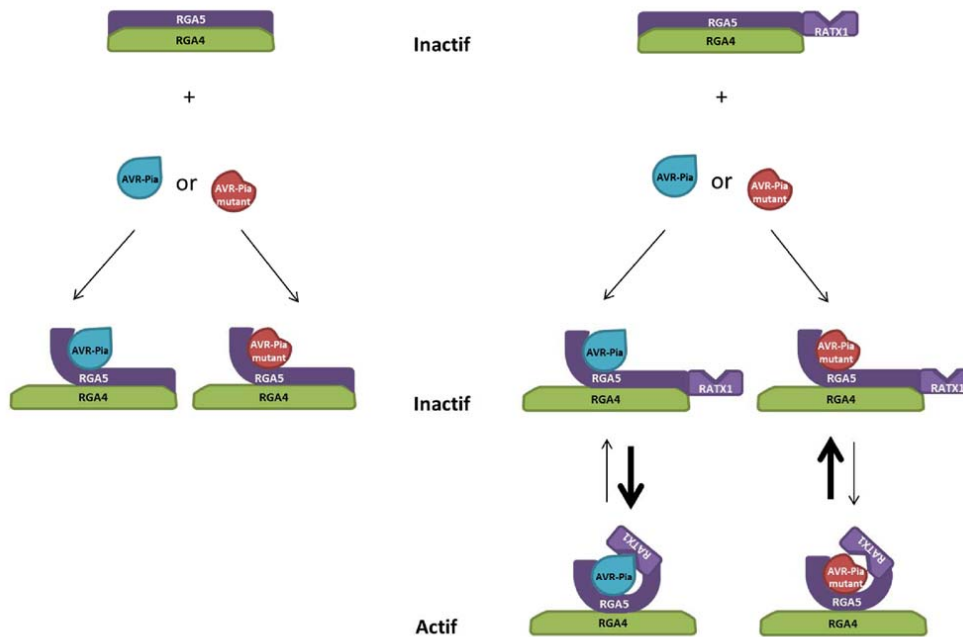


Figure 7. Model of AVR-Pia recognition by the RGA4/RGA5 receptor complex

493 state by simultaneously binding to multiple binding sites and inducing or stabilizing by
 494 this major conformational changes is a widespread mechanism in NLR activation and in
 495 particular NLRs with integrated domains. Future, structural and functional analysis will

496 be necessary to test this model and elucidate in more detail how activation occurs at the
497 molecular level.
498
499
500

501 **METHODS**

502

503 **Growth conditions of plants and fungi and infection assays**

504 *N. benthamiana* plants were grown in a growth chamber at 22°C with a 16-h light period.
505 Rice plants (*Oryza sativa*) were grown as described (Faivre-Rampant et al., 2008).
506 Transgenic *M. oryzae* GUY 11 strains were grown at 25°C during 5 days on rice flour
507 agar for spore production (Berruyer et al., 2003) and in Tanaka complete culture
508 medium (Villalba et al., 2008) agitated at 60 rpm and 25°C during 5 days for liquid
509 culture.

510 For the analysis of interaction phenotypes, a suspension of *M. oryzae* conidiospores in
511 water with 0.1% of gelatin and adjusted to 5×10^4 spores ml⁻¹ was sprayed on the
512 leaves of 3-week-old rice plants (Berruyer et al., 2003). Symptoms were analyzed 7
513 days after inoculation on the youngest leaf that was fully expanded at the time of
514 inoculation. For quantitative analysis lesions were classified and counted; resistant
515 lesions, visible as small brown spots (type 1); weakly susceptible/partially resistant
516 lesions characterized by a pronounced brown border and a small grey centre (type 2);
517 fully susceptible lesions characterized by a large grey centre (type 3).

518

519 **Constructs**

520 Plasmids were generated by Gateway cloning (ThermoFisher, Waltham, USA),
521 restriction/ligation, site-directed mutagenesis using the QuickChange Lightning kit
522 (Agilent, Santa Clara, USA) or gap-repair cloning in yeast (Bruno et al., 2004). Gateway
523 entry clones were generated using the pDONR207 plasmid (ThermoFisher, Waltham,
524 USA). Gateway destination vectors were modified pBIN19 plasmids for expression of
525 tagged proteins in *N. benthamiana* (Césari et al. 2013) or modified pGAD-T7 or pGBK-
526 T7 plasmids (Clontech, Mountain View, USA) for yeast two hybrid experiments (Bernoux
527 et al., 2011). For protein expression, the pET-15b vector (Merck-Millipore, Darmstadt
528 Germany) was used. For *M. oryzae* transformation constructs were based on the pDL02
529 plasmid (Bruno et al., 2004). For details on PCR and mutagenesis primers and
530 generation of plasmid refer to Supplemental Tables 3 and 4.

531

532 Nuclear magnetic resonance spectroscopy

533 Spectra were acquired on a 700 MHz Avance Bruker spectrometer equipped with triple-
534 resonance (^1H , ^{15}N , ^{13}C) z-gradient cryo-probe at 305 K. Experiments were recorded
535 using the TOPSPIN pulse sequence library (v. 2.1) (Supplemental Table 1). All spectra
536 are referenced to the internal reference DSS for the ^1H dimension and indirectly
537 referenced for the ^{15}N and ^{13}C dimensions (Wishart et al., 1995).

538 Spectra were processed using Topspin (v. 3.2) and analyzed using strip-plots with Cindy
539 in house software and CCPN (Vranken et al., 2005) [analysis v 2.3]. The ^{15}N and ^{13}C
540 assignments were derived from the 2D and 3D spectra at 700 MHz listed in Table S1.

541

542 NMR titration

543 For the assignments, protein samples (1mM) in 20 mM potassium-sodium phosphate,
544 pH 5.4 and 150 mM NaCl, were used. For the titrations of ^{15}N -labeled AVR-Pia proteins,
545 different samples with constant concentrations of AVR-Pia WT or H3 (50 μM) and
546 various concentrations of unlabeled RATX1 (ratios 2:1, 1:1, 0.5:1, 0.25:1, 0:1 for the
547 reference) were prepared. HSQC spectra were recorded at 305K on a Bruker Avance
548 700 MHz spectrometer. Chemical shift differences were measured from the HSQC
549 spectra of AVR-Pia or AVR-H3 alone and the AVR-RATX1 complex at R = 2. They are
550 reported as Hamming distance weighted by the magnetogyric ratios (Schumann et al.,
551 2007).

552

553 Coimmunoprecipitation and Yeast two hybrid interaction assays

554 Protein-protein interaction analyses by co-immunoprecipitation were performed with
555 protein extracts from *N.benthamina* leaf discs harvested 2 days after *Agrobacterium*
556 infiltration (Césari et al., 2013). For the interaction of AVR-Pia variants with RGA5_{C-ter}
557 and RGA5_{RATX1}, 5 leaf disks per sample were homogenized in extraction buffer (25mM
558 Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 10 mM DTT, 1 mM PMSF, 0.1% IGEPAL
559 CA-630 [NP-40]), supplemented with complete protease inhibitor cocktail (Roche) and
560 polyvinylpyrrolidone (PVPP 0.5%). After 2 centrifugations (30 minutes, 15000g) 5
561 μL magnetic GFP-trap_M beads ((Chromotek) per sample washed two times with
562 protein extraction buffer (without PVPP) were added to 500 μL protein extract and

563 incubated with gentle rotation for 2 h at 4°C. Beads were separated and washed three
564 times with 600 µL of protein extraction buffer (without PVVP).

565 For the interaction of AVR-Pia variants and RGA5 or RGA5_{ΔRATX1} a modified protein
566 extraction buffer was used (50mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 10 mM
567 DTT, 1 mM PMSF, 1.0% IGEPAL CA-630 [NP-40], 0.5% sodium deoxycholate, and
568 0.1% SDS, supplemented with a complete protease inhibitor cocktail (Roche) and
569 polyvinylpyrrolidone 0.5%). And co-IP was performed with 8 uL of agarose
570 GFP_trap_A suspension (Chromotek) and four washes with the modified protein
571 extraction buffer.

572 Bound proteins were eluted by boiling for 10 min at 70°C in 50 uL of Nupage sample
573 buffer, separated by polyacrylamide gel electrophoresis using NuPAGE 4-12% gels
574 (Invitrogen, Carlsbad, CA, USA), transferred to nitrocellulose membrane (Millipore), and
575 analyzed by immunoblotting. For immunodetection of proteins, rat anti-HA-horseradish
576 peroxidase (clone 3F10 ; Roche) or mouse anti-GFP (Roche) and goat anti-mouse-
577 horseradishperoxidase (Sigma-Aldrich) were used in combination with the Immobilon
578 western kit (Millipore).

579 Binding domain (BD) fusions of AVR-Pia variants in pGBKT7-53 and activation domain
580 (AD) fusions of RGA5_{C-ter} in pGADT7 were transformed in Gold and Y187 yeast strain
581 respectively. Interactions assays were performed according to the Matchmaker Gold
582 yeast two-hybrid system protocol (Clontech).

583

584 **Transient Protein Expression and HR assays in *N. benthamiana***

585 For agroinfiltration in *N. benthamiana* pBIN19 binary vectors containing either AVR-Pia,
586 PWL2 or RGA5 variants were transformed into *Agrobacterium tumefaciens* strains
587 GV3101 by electroporation. Individual clones were selected and grown in Luria-Bertani
588 liquid medium containing 50 mg ml⁻¹ rifampicin, 15 mg ml⁻¹ gentamycin, and 50 mg ml⁻¹
589 1 kanamycin at 28°C for 24h before agroinfiltration. Co-inoculation mixtures adjusted to
590 an OD600 of 1.0 were infiltrated in 4 weeks old *N. benthamiana* plants. The infiltrated
591 plants were incubated for 48 or 96 h in growth chambers under controlled conditions for
592 coimmunoprecipitations or cell death assays respectively. Three days post infiltration, *N.*
593 *benthamiana* leaves were scanned using a Typhoon FLA9000 fluorescence scanner

594 (GE Healthcare) with excitation at 635 nm and a long-pass red filter (LPR-665 nm) to
595 evaluate the HR response as a lack of red chlorophyll fluorescence.

596

597 **Accession Numbers**

598 Sequence data from this article correspond to those previously published by (Cesari, et
599 al., 2013) and can be found in the the GenBank/EMBL databases under the following
600 accession numbers: AVR-Pia (AB498873), AVR-Pia-H3 (KC777366), PWL2 (U26313),
601 RGA4 (AB604622), Sasanishiki RGA5-A (AB604627), Sasanishiki RGA5-B (KC777365).
602 The PDB accession number for the AVR-Pia_H3 structure is 5JHJ.

603

604 **Supplemental Data**

605

606 **Supplemental Table 1.** NMR experiments acquired for structure calculations and
607 chemical shift assignments of AVR-Pia-H3

608

609 **Supplemental Table 2.** Statistics for 20 NMR structures of AVR-Pia-H3

610

611 **Supplemental Table 3.** Primers

612

613 **Supplemental Table 4.** Plasmids

614

615 **Supplemental Methods**

616

617 **Supplemental Figure 1.** Comparison of NMR relaxation of AVR-Pia and AVR-Pia-H3

618

619 **Supplemental Figure 2.** HSQC spectra of AVR-Pia and AVR-Pia-H3 recorded upon
620 titration with RGA5_{RATX1}

621

622 **Supplemental Figure 3.** AVR-Pia mutants affected in RGA5_{RATX1}-binding are well-
623 structured

624

625 **Supplemental Figure 4.** AVR-Pia mutants not-affected in RGA5_{RATX1}-binding trigger HR
626 in *N.benthamiana*

627

594 (GE Healthcare) with excitation at 635 nm and a long-pass red filter (LPR-665 nm) to
595 evaluate the HR response as a lack of red chlorophyll fluorescence.

596

597 **Accession Numbers**

598 Sequence data from this article correspond to those previously published by (Cesari, et
599 al., 2013) and can be found in the the GenBank/EMBL databases under the following
600 accession numbers: AVR-Pia (AB498873), AVR-Pia-H3 (KC777366), PWL2 (U26313),
601 RGA4 (AB604622), Sasanishiki RGA5-A (AB604627), Sasanishiki RGA5-B (KC777365).
602 The PDB accession number for the AVR-Pia_H3 structure is 5JHJ.

603

604 **Supplemental Data**

605

606 **Supplemental Table 1.** NMR experiments acquired for structure calculations and
607 chemical shift assignments of AVR-Pia-H3

608

609 **Supplemental Table 2.** Statistics for 20 NMR structures of AVR-Pia-H3

610

611 **Supplemental Table 3.** Primers

612

613 **Supplemental Table 4.** Plasmids

614

615 **Supplemental Methods**

616

617 **Supplemental Figure 1.** Comparison of NMR relaxation of AVR-Pia and AVR-Pia-H3

618

619 **Supplemental Figure 2.** HSQC spectra of AVR-Pia and AVR-Pia-H3 recorded upon
620 titration with RGA5_{RATX1}

621

622 **Supplemental Figure 3.** AVR-Pia mutants affected in RGA5_{RATX1}-binding are well-
623 structured

624

625 **Supplemental Figure 4.** AVR-Pia mutants not-affected in RGA5_{RATX1}-binding trigger HR
626 in *N.benthamiana*

627

628 **Supplemental Figure 5.** Characterization of transgenic *M. oryzae* isolates carrying
629 *AVR-Pia* mutant constructs.

630

631 **Supplemental Figure 6.** RGA5_{ΔRATX1} represses RGA4-mediated cell death but does not
632 recognize AVR-Pia

633

634 **Supplemental Figure 7.** Comparison of the AVR-Pia and AVR-PikD structures and their
635 complexes with RAX1/HMA domains.

636

637

638 **Author contributions**

639 Conceptualization, T.K and A.P.; Investigation, D.O., J.G., K.G, V.C. and A.P.; Writing –
640 Original Draft, D.O, A.P and T.K.; Writing – Review & Editing, D.O, A.P, S.C. and T.K.;
641 Funding Acquisition, T.K and A.P.; Resources, D.O. and S.C.; Supervision, T.K. and
642 A.P.

643

644 T.K and A.P designed the research. D.O, S.C, K.G, V.C and A.P performed the
645 research. A.P and K.G performed NMR analyses. D.O and S.C generated vectors. D.O
646 and V.C produced transgenic *M. oryzae* strains. D.O performed protein-protein
647 interaction analyses. D.O, J.G, K.G, A.P and T.K analyzed data. D.O, TK and AP wrote
648 the article

649

650 **Acknowledgments**

651 This work was supported by the French Infrastructure for Integrated Structural Biology
652 (FRISBI) ANR-10-INSB-05-0 and the ANR project Immunereceptor (ANR-15-CE20-
653 0007). D. Ortiz-Vallejo was supported by a PhD grant from the ministry of research of
654 Colombia (Colciencias). This work benefited from interactions promoted by COST Action
655 FA 1208 <https://www.cost-sustain.org>. Christian Roumestand is acknowledged for fruitful
656 discussions.

657

658

659

660 **FIGURE LEGENDS**

661 **Figure 1. The AVR-Pia-H3 NMR structure is similar to the structure of wild-type**
 662 **AVR-Pia**

663 (A) Chemical shift differences ($\Delta\delta_{NH}$) from the comparison of ^{15}N -HSQC of AVR-Pia
 664 wild-type and mutants F24S, T46N or F24S T46N (AVR-Pia-H3). The β -strand
 665 assignments from the AVR-Pia^{wt} structure are indicated on the top and polymorphic
 666 residues by (*).

667 (B) Structure overlay of AVR-Pia (blue) and AVR-Pia-H3 (orange).

668 (C) Topology of the AVR-Pia-H3 structure.

669

670 **Figure 2. AVR-Pia binds RGA5_{RATX1} with intermediate affinity and a well-defined**
 671 **interaction surface**

672 (A) ITC curves for the titration of the RGA5_{RATX1} domain by AVR-Pia^{wt} (■) and AVR-Pia-
 673 H3 (●) at 25°C. For AVR-Pia^{wt} the fitting parameters were: $N = 0.994 \pm 0.004$, $K_a = 1.28$
 674 $\pm 0.04 \cdot 10^{-5} \text{ mol}^{-1}$, $\Delta H = -8179 \pm 47.95 \text{ cal.mol}^{-1}$, $\Delta S = -4.06 \text{ cal.K}^{-1}.\text{mol}^{-1}$. The red line
 675 shows a simulated curve for a 10x lower affinity ($K_a = 1.28 \cdot 10^{-4} \text{ mol}^{-1}$).

676 (B) NMR titration and surface mapping. Plot of the chemical shift differences (Δppm)
 677 between unbound and bound AVR-Pia (blue) or AVR-Pia-H3 (red). Chemical shift
 678 differences were calculated as the Hamming distance (Schumann et al., 2007), $\Delta\delta$ (ppm)
 679 $= |\Delta\delta(^1H)_{ij}| + 0.102 * |\Delta\delta(^{15}N)_{ij}|$, where $\Delta\delta(^1H)_{ij}$ and $\Delta\delta(^{15}N)_{ij}$ are the chemical shift
 680 differences observed at R=0 and R=2, respectively.

681 Structures of AVR-Pia (C and D) and AVR-Pia-H3 (E) with color-coded surfaces
 682 showing the differences in chemical shifts in the NMR titration (difference between free
 683 (R=0) and RGA5_{RATX1}-bound AVR-Pia or AVR-PiaH3 (R=2)). Surfaces of residues with

684 chemical shift differences $\Delta\delta(\text{ppm}) \geq 0.2$ are shown in dark blue (residues in white
685 letters), and in light blue for $0.2 > \Delta\delta(\text{ppm}) \geq 0.1$ ppm (residues in black letters).
686 Surfaces of residues not observed in the AVR-Pia-RGA5_{RATX1} complex (R=2) HSQC are
687 reported in grey (residues in red letters), and not perturbed residues are not highlighted
688 (residues are not indicated). The view in D is the opposite face of C, which has been
689 rotated 180° from the vertical axis.

690

691 **Figure 3. Mutations in the binding surface of AVR-Pia affect binding to RGA5_{C-ter} in**
692 **yeast two hybrid assays.**

693 (A) The interaction between AVR-Pia mutants (BD fusion) and RGA5_{C-ter} (AD fusion) was
694 assayed by a yeast two-hybrid experiment. Three dilutions (1/10, 1/100, 1/1000) of yeast
695 cultures adjusted to an OD of 0.2 were spotted on synthetic double drop out (DDO)
696 medium (-Trp/-Leu) to control proper growth and on synthetic TDO (-Trp/-Leu/-His)
697 either without or supplemented with 3-amino-1,2,4-triazole (3AT) to test for interaction.
698 Yeast transformations and interaction analyses were performed twice with identical
699 results. Photos were taken after 4 days of growth.

700 (B) Equal production of AVR-Pia mutant proteins was determined by immunoblotting
701 with anti-AVR-Pia antibodies.

702

703 **Figure 4. AVR-Pia mutants with reduced RGA5_{C-ter} binding in yeast are also**
704 **impaired in binding to RGA5_{C-ter} and RGA5_{RATX1} *in planta*.**

705 HA:RGA5_{C-ter} (A) or HA:RGA5_{RATX1} (B) were transiently expressed with YFP:AVR-Pia^{WT},
706 YFP:AVR-Pia mutants or YFP:PWL2 in *N. benthamiana*. Protein extracts were analysed
707 by immunoblotting with anti-HA (α -HA) and anti-GFP antibodies (α -GFP) (Input).

708 Immunoprecipitation (IP) was conducted with anti-GFP beads (IP GFP) and analysed by
709 immunoblotting with α -GFP for the detection of immunoprecipitated AVR-Pia variants.
710 Co-precipitated HA:RGA5_{C-ter} (A) or HA:RGA5_{RATX1} (B) proteins were detected using α -
711 HA antibody.

712

713 **Figure 5. Effector recognition by RGA5 requires binding to the RATX1 domain**

714 Transgenic *M. oryzae* isolates were analysed for the production of the AVR-Pia protein
715 by immunoblotting using culture filtrate and α -AVR-Pia antibodies (A lower panel) and
716 were sprayed on 3-week-old plants of the rice cultivar Kitaake possessing *Pia*
717 resistance. 7 days after inoculation, leaves were scanned (A lower panel) and 3 different
718 types of lesions (1= fully resistant, 2=partially resistant/weakly susceptible 3=fully
719 susceptible) were counted on leaves from 10 different plants per isolate to determine
720 mean symptom scores and significantly different classes of isolates using Kruskal-Wallis
721 analysis of variance combined with a multi- comparison Dunn test for non-parametric
722 data (B). The AVR-Pia variants grouped with respect to their avirulence activity in three
723 significantly different classes: a=inactive; b=partially active; c=active. Similar results
724 were obtained in two independent experiments and with additional transgenic isolates.

725

726 **Figure 6. AVR-Pia interacts with RGA5 outside the RATX1 domain**

727 HA:RGA5 (A) and HA:RGA5 Δ _{RATX1} (B) were expressed with YFP:AVR-Pia^{WT}, YFP:AVR-
728 Pia mutants and YFP:PWL2 in *N. benthamiana*. Protein extracts were analysed by
729 immunoblotting with anti-HA (α -HA) and anti-GFP antibodies (α -GFP) (Input).
730 Immunoprecipitation (IP) was conducted with anti-GFP beads (IP GFP) and analysed by

731 immunoblotting with α -GFP for the detection of immunoprecipitated AVR-Pia variants.

732 Co-precipitated RGA5 (A) or HA:RGA5 Δ RATX1 (B) were detected using α -HA antibody.

733

734 **Figure 7. Model of AVR-Pia recognition by the RGA4/RGA5 receptor complex**

735 AVR-Pia binds to the RATX1 domain of RGA5 with a defined interaction surface and

736 interacts, in addition, through independent surfaces with other sites in RGA5. These

737 additional interactions are not sufficient to relieve the repression RGA5 exerts on RGA4.

738 Indeed, AVR-Pia mutants that associate with RGA5, but do not bind RGA5_{RATX1} as well

739 as RGA5 mutants that lack the RATX1 domain, do not permit activation of resistance.

740 We propose that simultaneous interactions of AVR-Pia with different parts of RGA5,

741 including the RATX1 domain, stabilize conformational changes that activate the

742 RGA4/RGA5 complex.

Parsed Citations

Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., Zhang, X., Ellis, J.G., Kobe, B., and Dodds, P.N. (2011). Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell Host Microbe* 9: 200-11.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Bonardi, V., Tang, S., Stallmann, A., Roberts, M., Cherkis, K., and Dangl, J.L. (2011). Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc. Natl. Acad. Sci. U. S. A.* 108: 16463-8.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Bourett, T.M., Sweigard, J. a, Czymmek, K.J., Carroll, A., and Howard, R.J. (2002). Reef coral fluorescent proteins for visualizing fungal pathogens. *Fungal Genet. Biol.* 37: 211-20.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Césari, S. et al. (2013). The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell* 25: 1463-81.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Césari, S., Bernoux, M., Moncuquet, P., Kroj, T., and Dodds, P.N. (2014a). A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. *Plant-Microbe Interact.* 5: 606.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Césari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., Shimamoto, K., Dodds, P., Terauchi, R., and Kroj, T. (2014b). The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *EMBO J.* 33: 1941-1959.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Chou, S., Krasileva, K. V, Holton, J.M., Steinbrenner, A.D., Alber, T., and Staskawicz, B.J. (2011). *Hyaloperonospora arabidopsidis* ATR1 effector is a repeat protein with distributed recognition surfaces. *Proc. Natl. Acad. Sci. U. S. A.* 108: 13323-8.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Collier, S.M. and Moffett, P. (2009). NB-LRRs work a "bait and switch" on pathogens. *Trends Plant Sci.* 14: 521-9.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Cui, H., Tsuda, K., and Parker, J.E. (2015). Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. *Annu. Rev. Plant Biol.* 66: 487-511.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Dean, R., Kan, J.A.N.A.L.V.A.N., Pretorius, Z.A., Hammond-kosack, K.I.M.E., Pietro, A.D.I., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., Ellis, J., and Foster, G.D. (2012). The Top 10 fungal pathogens in molecular plant pathology. *13*: 414-430.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Eitas, T.K. and Dangl, J.L. (2010). NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr. Opin. Plant Biol.* 13: 472-7.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Ellis, J.G., Dodds, P.N., and Lawrence, G.J. (2007). Flax Rust Resistance Gene Specificity is Based on Direct Resistance-Avirulence Protein Interactions. *Annu. Rev. Phytopathol.* 45: 289-306.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)
 Google Scholar: [Author Only Title Only Author and Title](#)

Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., Hayashi, N., Takahashi, A., Hirochika, H., Okuno, K., and Yano, M. (2009). Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science* 325: 998-1001.

Pubmed: [Author and Title](#)
 CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Gabriëls, S.H.E.J., Vossen, J.H., Ekengren, S.K., van Ooijen, G., Abd-El-Halim, A.M., van den Berg, G.C.M., Rainey, D.Y., Martin, G.B., Takken, F.L.W., de Wit, P.J.G.M., and Joosten, M.H. a J. (2007). An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins. *Plant J.* 50: 14-28.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Guillen, K., Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., and Padilla, A. (2015). Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. *PLoS Pathog.* accepted f.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

van der Hoorn, R.A.L. and Kamoun, S. (2008). From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors. *Plant Cell* 20: 2009-2017.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Jacob, F., Vernaldi, S., and Maekawa, T. (2013). Evolution and Conservation of Plant NLR Functions. *Front. Immunol.* 4: 297.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Jia, Y., McAdams, S. a, Bryan, G.T., Hershey, H.P., and Valent, B. (2000). Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J.* 19: 4004-14.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kanzaki, H. et al. (2012). Arms race co-evolution of Magnaporthe oryzae AVR-Pik and rice Pik genes driven by their physical interactions. *Plant J.*: 894-907.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Khang, C.H., Berruyer, R., Giraldo, M.C., Kankanala, P., Park, S.-Y., Czymmek, K., Kang, S., and Valent, B. (2010). Translocation of Magnaporthe oryzae effectors into rice cells and their subsequent cell-to-cell movement. *Plant Cell* 22: 1388-403.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Krasileva, K. V, Dahlbeck, D., and Staskawicz, B.J. (2010). Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. *Plant Cell* 22: 2444-58.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Liu, W., Liu, J., Triplett, L., Leach, J.E., and Wang, G.-L. (2014). Novel Insights into Rice Innate Immunity Against Bacterial and Fungal Pathogens.

Maekawa, T. et al. (2011). Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. *Cell Host Microbe* 9: 187-99.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Maqbool, a, Saitoh, H., Franceschetti, M., Stevenson, C., Uemura, a, Kanzaki, H., Kamoun, S., Terauchi, R., and Banfield, M. (2015). Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *Elife* 4: 1-24.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Nyarko, A., Singarapu, K.K., Figueroa, M., Manning, V. a., Pandelova, I., Wolpert, T.J., Ciuffetti, L.M., and Barbar, E. (2014). Solution NMR Structures of Pyrenophora tritici-repentis ToxB and Its Inactive Homolog Reveal Potential Determinants of Toxin Activity. *J. Biol. Chem.* 289: 25946-25956.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Okuyama, Y. et al. (2011). A multifaceted genomics approach allows the isolation of the rice Pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes. *Plant J.* 66: 467-79.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Pennisi, E. (2010). Armed and Dangerous. *Science* (80-). 327: 804-805.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

- Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Qi, D. and Innes, R. (2013).** Recent Advances in Plant NLR Structure, Function, Localization, and Signaling. *Front. Immunol.*: 221.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Ribot, C., Césari, S., Abidi, I., Chalvon, V., Bournaud, C., Vallet, J., Lebrun, M.-H., Morel, J.-B., and Kroj, T. (2013).** The Magnaporthe oryzae effector AVR1-CO39 is translocated into rice cells independently of a fungal-derived machinery. *Plant J.* 74: 1-12.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Le Roux, C. et al. (2015).** A Receptor Pair with an Integrated Decoy Converts Pathogen Disabling of Transcription Factors to Immunity. *Cell* 161: 1074-1088.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Sarris, P.F.F. et al. (2015).** A Plant Immune Receptor Detects Pathogen Effectors that Target WRKY Transcription Factors. *Cell* 161: 1089-1100.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Schumann, F.H., Riepl, H., Maurer, T., Gronwald, W., Neidig, K.P., and Kalbitzer, H.R. (2007).** Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J. Biomol. NMR* 39: 275-289.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Skamnioti, P. and Gurr, S.J. (2009).** Against the grain: safeguarding rice from rice blast disease. *Trends Biotechnol.* 27: 141-50.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Sohn, K.H., Hughes, R.K., Piquerez, S.J., Jones, J.D.G., and Banfield, M.J. (2012).** Distinct regions of the Pseudomonas syringae coiled-coil effector AvrRps4 are required for activation of immunity. *Proc. Natl. Acad. Sci. U. S. A.* 109: 16371-6.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Steinbrenner, A.D., Goritschnig, S., and Staskawicz, B.J. (2015).** Recognition and Activation Domains Contribute to Allele-Specific Responses of an Arabidopsis NLR Receptor to an Oomycete Effector Protein. *PLOS Pathog.* 11: e1004665.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Takken, F.L.W. and Govere, A. (2012).** How to build a pathogen detector: structural basis of NB-LRR function. *Curr. Opin. Plant Biol.* 15: 375-384.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005).** The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Bioinforma.* 59: 687-696.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Williams, S.J. et al. (2014).** Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science* 344: 299-303.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., and Sykes, B.D. (1995).** 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* 6: 135-140.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Wu, C.-H., Belhaj, K., Bozkurt, T.O., Birk, M.S., and Kamoun, S. (2015).** Helper NLR proteins NRC2a/b and NRC3 but not NRC1 are required for Pto-mediated cell death and resistance in Nicotiana benthamiana. *New Phytol.*: doi: <http://dx.doi.org/10.1101/015479>.
Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Zhang, Z.-M., Zhang, X., Zhou, Z.-R., Hu, H.-Y., Liu, M., Zhou, B., and Zhou, J. (2013).** Solution structure of the Magnaporthe oryzae avirulence protein AvrPiz-t. *J. Biomol. NMR* 55: 219-23.

Pubmed: [Author and Title](#)
CrossRef: [Author and Title](#)
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Supplemental Table 1. NMR experiments acquired for structure calculations and chemical shift assignments of AVR-Pia-H3

Experiments	nuclei	Size			Sweep width (ppm)			Mix(ms)	NS	D1(s)	B ₀ (MHz)
		F3	F2	F1	F3	F2	F1				
¹⁵ N-HSQC	¹ H, ¹⁵ N	1500	256	-	-	-	-	-	8	1	700
¹⁵ N-NOESY-HSQC	¹ H, ¹⁵ N, ¹ H	1500	64	280	14	40	14	150	8	1	700
NOESY (H ₂ O)	¹ H, ¹ H	1500	512	-	15.95	15.95	-	150	96	1	700
TOCSY (H ₂ O)	¹ H, ¹ H	1500	512	-	15.95	15.95	-	41.2	64	1	700
¹³ C-HSQC(D ₂ O)	¹ H, ¹³ C	2048	400	-	14	160	-	-	240	1	700
¹³ C-TOCSY(D ₂ O)	¹ H, ¹³ C	2048	400	-	12.03	100	-	40.25	320	1	700
¹ H, ¹⁵ N NOE	¹ H, ¹⁵ N	1024	128	-	14	35	-	Sat. 3s	64	6	500
R ₁	¹ H, ¹⁵ N	1024	128	-	14	35	-		16	2.5	500
R ₂	¹ H, ¹⁵ N	1024	128	-	14	35	-		16	2.5	500

Supplemental Table 2. Statistics for 20 NMR structures of AVR-Pia-H3

	AVR-Pia-H3
NOE restraints	1366
Short range (i-j ≤1)	891
Medium range (1< i-j <5)	133
Long range (i-j ≥5)	342
H-bond restraints	20
Dihedral restraints (a)	100
Average Number of NOE Violations per structure	
> 0.1 Å	23.75
> 0.2 Å	1.0
> 0.3 Å	0.05
> 0.4 Å	0
Dihedral violations	
> 2°	0
Ramachandran plot statistics	
most favourable regions (%)	84.3
additionally allowed regions (%)	15.7
generously allowed regions (%)	0.0
disallowed regions (%)	0.0
Pairwise RMSD (Å) (b)	
Backbone	0.77 ± 0.19
Heavy atoms	1.43 ± 0.16

Structures were calculated using CYANA, refined using CNS, and analysed using PROCHECK.

(a) Residues in regular secondary structures were derived from the chemical shifts using TALOS+ software.

(b) Main chain atoms (N, C α , C) over the residues 20-85.

Supplemental Table 3. Primer

oDO01	gcttaatggcgccagctgcttttgctctattacgac
oDO02	gtcgtaatagacgcaaaaagcagctggcgccattaagc
oDO03	gacacgtgtcctgctttagcttagaatacgca
oDO04	tgccgattctaacgtaagcaagcaggacacgtgtc
oDO05	acgtgtcctgcttatggctgttagaatcggcacta
oDO06	tagtgccgattctaacagccataagcaggacacgt
oDO07	ggcttgctgccgagcctgcttaggacccagctttctt
oDO08	aagaaagctgggtcctaagcaggctcggcagcaagcc
oDO17	atthtgcgtctattacgctggccaccttcccgcga
oDO18	tcgcgggaaggtggccagcgtaatagacgcaaat
oDO19	ccaccttcccgcgacagctgtcctgcttatgtacgt
oDO20	acgtacataagcaggacagctgtcgcgggaaggtgg
oDO21	tacggcccgtgggcacgctttcgaagtgaagcaa
oDO22	ttgctcaacttcgaaagcgtgccacgggcccgtta
oDO23	ccgtgggcacgaattcgctgttgaagcaaaagacca
oDO24	tggtctttgctcaacagcgaattcggtcccacgg
oDO25	tcgaagtgaagcaaaagctcagaattgcaaagtatt
oDO26	ataactttgcaattctgagctttgctcaacttcga
oDO27	caaaagaccagaattgcgctgttattctaccaatgg
oDO28	ccattggtgagaataacagcgcgaattctggtctttg
oDO29	atggcaacaagcaccggctggctgtgctccgagcct
oDO30	aggctcggcagcaagccaagccggtgctgtttgcat
oDO31	ttattctaccaatggcgtcaagcaccggattggct
oDO32	agccaatccgggtgcttgagcgcattggtgagaataa
oDO41	atatggctgcgccagctgcttttgctctattacgac
oDO42	gtcgtaatagacgcaaaaagcagctggcgagccatat
oDO43	taagcgtgcgccagctgcttttgctctattacgac
oDO44	gtcgtaatagacgcaaaaagcagctggcgagcgtta
oDO45	gctgtgcgccagctagaagctgctctattacgacgg
oDO46	ccgtcgtaatagacgcagctttagctggcgagcgc
oGT013	catatggcgccagctagatcttgcgtctattacgacggc
oGT014	gcccgtcgtaatagacgcaagatctagctggcgccatatg
oGT015	atgtacgttagaatacggaatacagcgactattacggcc
oGT016	ggccgtaatagtcgctgtattgccgattctaacgtacat
oGT09	gtcctgcttatgtacgttggaaatcggcactacagcga
oGT10	tcgctgtagtgccgattccaacgtacataagcaggac
oCS84	ggggaccactttgtacaagaaagctgggtcCTAGTAAGGCTCGGCAGCAAGC
oTK333	tatcatatggctGCGCCAGCTAGATTTTTCGTCTAT
oTK334	tatggatccCTAGTAAGGCTCGGCAGCAAG
oTK344	AGGACCCAATCTTCAAATGCATTTTTTCGACAATTTTCATCCC
oTK345	AATGTTGAGTGGAATGATGCGGCTAGTAAGGCTCGGCAGCAAG
oTK409	ggggacaagttgtacaagaaagcaggcttaATGGCGCCAGCTAGATcTTGC
oTK439	tatcatatggctGCGCCAGCTAGATcTTGCGTCTAT
oTK472	attaCATATGCAGCGTACCAAAATTGTTGTTAAAG
oTK473	attaGGATCCtcaTTTTTTCACGCTTTTCGACAACCA

Supplemental Table 4. Plasmids

Use	Plasmid name	Construct	Plasmid backbone	Insert	Cloning method	Primers	PCR template or pENTRY	Reference	
Entry clones	pSC060	AVR-Pia	pDONOR207	AVR-Pia_20-85	-	-	-	Césari et al. 2013	
	pSC41	RGAA cDNA	pDONOR207	RGAA	-	-	-	Césari et al. 2013	
	pSC42	RGAA cDNA	pDONOR207	RGAA	-	-	-	Césari et al. 2013	
	pSC120	PWL2	pDONOR207	PWL2	-	-	-	Césari et al. 2013	
	pSC129	RGAA _{C-ter}	pDONOR207	RGAA_882-1116	-	-	-	Césari et al. 2014	
	pSC207	RGAA _{RATX1}	pDONOR207	RGAA_997-1072	-	-	-	Césari et al. 2014	
	pSC210	RGAA _{RATX1}	pDONOR207	RGAA_1-996	-	-	-	Césari et al. 2014	
	pDO01	AVR-Pia R23A	pDONOR207	AVR-Pia R23A	Quick change	oDO01/oDO02	pSC060	this study	
	pDO02	AVR-Pia M40A	pDONOR207	AVR-Pia M40A	Quick change	oDO03/oDO04	pSC060	this study	
	pDO03	AVR-Pia Y41A	pDONOR207	AVR-Pia Y41A	Quick change	oDO05/oDO06	pSC060	this study	
	pDO04	AVR-Pia Y85A	pDONOR207	AVR-Pia Y85A	Quick change	oDO07/oDO08	pSC060	this study	
	pDO09	AVR-Pia R43G	pDONOR207	AVR-Pia R43G	Quick change	oGT09/oGT10	pSC060	this study	
	pDO52	AVR-Pia D29A	pDONOR207	AVR-Pia D29A	Quick change	oDO17/oDO18	pSC060	this study	
	pDO53	AVR-Pia R36A	pDONOR207	AVR-Pia R36A	Quick change	oDO19/oDO20	pSC060	this study	
	pDO54	AVR-Pia E56A	pDONOR207	AVR-Pia E56A	Quick change	oDO21/oDO22	pSC060	this study	
	pDO55	AVR-Pia E58A	pDONOR207	AVR-Pia E58A	Quick change	oDO23/oDO24	pSC060	this study	
	pDO56	AVR-Pia D63A	pDONOR207	AVR-Pia D63A	Quick change	oDO25/oDO26	pSC060	this study	
	pDO57	AVR-Pia K67A	pDONOR207	AVR-Pia K67A	Quick change	oDO27/oDO28	pSC060	this study	
	pDO58	AVR-Pia D78A	pDONOR207	AVR-Pia D78A	Quick change	oDO29/oDO30	pSC060	this study	
	pDO59	AVR-Pia K74A	pDONOR207	AVR-Pia K74A	Quick change	oDO31/oDO32	pSC060	this study	
	pCV84	AVR-Pia F24S	pDONOR207	AVR-Pia F24S	Gateway BP	oCS84/oTK409	pSC060	this study	
	Yeast two hybrid	pDO50	AD-empty	pGADT7-GW	no insert	-	-	-	Bernoux et al., 2011
		pDO51	BD-empty	pGBKT7-GW	no insert	-	-	-	Bernoux et al., 2011
		pSC003	BD-AVR-Pia	pGBKT7	AVR-Pia_20-85	-	-	-	Césari et al. 2013
		pDO49	BD-RGA5 _{C-ter}	pGBKT7-GW	RGAA_882-1116	Gateway LR	-	pSC129	this study
		pDO38	AD-RGA5 _{C-ter}	pGADT7-GW	RGAA_882-1116	Gateway LR	-	pSC129	this study
		pDO48	BD-AVR-Pia wt	pGBKT7-GW	AVR-Pia wt	Gateway LR	-	pSC060	this study
		pDO39	BD-AVR-Pia R23A	pGBKT7-GW	AVR-Pia R23A	Gateway LR	-	pDO01	this study
pDO40		BD-AVR-Pia M40A	pGBKT7-GW	AVR-Pia M40A	Gateway LR	-	pDO02	this study	
pDO41		BD-AVR-Pia Y41A	pGBKT7-GW	AVR-Pia Y41A	Gateway LR	-	pDO03	this study	
pDO42		BD-AVR-Pia Y85A	pGBKT7-GW	AVR-Pia Y85A	Gateway LR	-	pDO04	this study	
pDO47		BD-AVR-Pia R43G	pGBKT7-GW	AVR-Pia R43G	Gateway LR	-	pDO09	this study	
pDO69		BD-AVR-Pia D29A	pGBKT7-GW	AVR-Pia D29A	Gateway LR	-	pDO52	this study	
pDO70		BD-AVR-Pia R36A	pGBKT7-GW	AVR-Pia R36A	Gateway LR	-	pDO53	this study	
pDO71		BD-AVR-Pia E56A	pGBKT7-GW	AVR-Pia E56A	Gateway LR	-	pDO54	this study	
pDO72		BD-AVR-Pia E58A	pGBKT7-GW	AVR-Pia E58A	Gateway LR	-	pDO55	this study	
pDO73		BD-AVR-Pia D63A	pGBKT7-GW	AVR-Pia D63A	Gateway LR	-	pDO56	this study	
pDO74		BD-AVR-Pia K67A	pGBKT7-GW	AVR-Pia K67A	Gateway LR	-	pDO57	this study	
pDO75		BD-AVR-Pia D78A	pGBKT7-GW	AVR-Pia D78A	Gateway LR	-	pDO58	this study	
pDO76		BD-AVR-Pia K74A	pGBKT7-GW	AVR-Pia K74A	Gateway LR	-	pDO59	this study	
pGT22		BD-AVR-Pia F24S	pGBKT7	AVR-Pia F24S	Quick change	oGT013/oGT14	pSC003	this study	
pGT23	BD-AVR-Pia T46N	pGBKT7	AVR-Pia T46N	Quick change	oGT015/oGT16	pSC003	this study		
<i>N. benthamiana</i> cell & colP	pCV129	HA-RGA5	pBIN19 3XHA-GTW	RGAA5	-	-	-	Césari et al. 2014	
	pDO120	HA-RGA5 _{RATX1}	pBIN19 3XHA-GTW	RGAA5_997-1072	Gateway LR	-	pSC207	this study	
	pSC144	HA-RGA5 _{C-ter}	pBIN19 3XHA-GTW	RGAA5_882-1116	Gateway LR	-	pSC129	this study	
	pDO121	HA-RGA5 _{RATX1}	pBIN19 3XHA-GTW	RGAA5_1-996	Gateway LR	-	pSC210	this study	
<i>N. benthamiana</i> cell	pSC61	RGAA-3XHA	pBIN19 GTW-3XHA	RGAA4	-	-	-	Césari et al. 2014	
	pSC95	AVR-Pia -stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia	-	-	-	Césari et al. 2014	
	pCV91	AVR-Pia F24S-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia F24S	Gateway LR	-	pCV84	this study	
	pDO10	AVR-Pia R23A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia R23A	Gateway LR	-	pDO01	this study	
	pDO11	AVR-Pia M40A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia M40A	Gateway LR	-	pDO02	this study	
	pDO12	AVR-Pia Y41A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia Y41A	Gateway LR	-	pDO03	this study	
	pDO13	AVR-Pia Y85A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia Y85A	Gateway LR	-	pDO04	this study	
	pDO18	AVR-Pia R43G-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia R43G	Gateway LR	-	pDO09	this study	
	pDO103	AVR-Pia D29A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia D29A	Gateway LR	-	pDO52	this study	
	pDO104	AVR-Pia R36A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia R36A	Gateway LR	-	pDO53	this study	
	pDO105	AVR-Pia E56A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia E56A	Gateway LR	-	pDO54	this study	
	pDO106	AVR-Pia E58A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia E58A	Gateway LR	-	pDO55	this study	
	pDO107	AVR-Pia D63A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia D63A	Gateway LR	-	pDO56	this study	
	pDO108	AVR-Pia K67A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia K67A	Gateway LR	-	pDO57	this study	
pDO109	AVR-Pia D78A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia D78A	Gateway LR	-	pDO58	this study		
pDO110	AVR-Pia K74A-stop-3XHA	pBIN19 GTW-3XHA	AVR-Pia K74A	Gateway LR	-	pDO59	this study		
<i>N. benthamiana</i> colP	pDO119	YFP-Pwl2	pBIN19-YFP-GTW	Pwl2	Gateway LR	-	pSC120	this study	
	pSC310	YFP	pBIN19-YFP-GTW	YFP	-	-	-	Césari et al. 2014	
	pSC80	YFP-AVR-Pia	pBIN19 YFP-GTW	YFP-AVR-Pia	Gateway LR	-	pSC060	this study	
	pDO19	YFP-AVR-Pia R23A	pBIN19 YFP-GTW	AVR-Pia R23A	Gateway LR	-	pDO01	this study	
	pDO20	YFP-AVR-Pia M40A	pBIN19 YFP-GTW	AVR-Pia M40A	Gateway LR	-	pDO02	this study	
	pDO21	YFP-AVR-Pia Y41A	pBIN19 YFP-GTW	AVR-Pia Y41A	Gateway LR	-	pDO03	this study	
	pDO22	YFP-AVR-Pia Y85A	pBIN19 YFP-GTW	AVR-Pia Y85A	Gateway LR	-	pDO04	this study	
	pDO27	YFP-AVR-Pia R43G	pBIN19 YFP-GTW	AVR-Pia R43G	Gateway LR	-	pDO09	this study	
	pDO111	YFP-AVR-Pia D29A	pBIN19-YFP-GTW	AVR-Pia D29A	Gateway LR	-	pDO52	this study	
	pDO112	YFP-AVR-Pia R36A	pBIN19-YFP-GTW	AVR-Pia R36A	Gateway LR	-	pDO53	this study	
	pDO113	YFP-AVR-Pia E56A	pBIN19-YFP-GTW	AVR-Pia E56A	Gateway LR	-	pDO54	this study	
	pDO114	YFP-AVR-Pia E58A	pBIN19-YFP-GTW	AVR-Pia E58A	Gateway LR	-	pDO55	this study	
	pDO115	YFP-AVR-Pia D63A	pBIN19-YFP-GTW	AVR-Pia D63A	Gateway LR	-	pDO56	this study	
	pDO116	YFP-AVR-Pia K67A	pBIN19-YFP-GTW	AVR-Pia K67A	Gateway LR	-	pDO57	this study	
pDO117	YFP-AVR-Pia D78A	pBIN19-YFP-GTW	AVR-Pia D78A	Gateway LR	-	pDO58	this study		
pDO118	YFP-AVR-Pia K74A	pBIN19-YFP-GTW	AVR-Pia K74A	Gateway LR	-	pDO59	this study		
<i>E. coli</i> protein express	pCV64	AVR-Pia	pET15b	AVR-Pia	-	-	-	de Guillen et al., 2016	
	pCV184	RGAA _{RATX1}	pET15b	RGAA5_995-1069	restriction NdeI-BamHI	oTK472/oTK473	pTK207	this study	
	pCV148	AVR-Pia_H3	pET15b	AVR-Pia_H3	restriction NdeI-BamHI	oTK439/oTK334	pGT5	this study	
	pCV147	AVR-Pia F24S	pET15b	AVR-Pia F24S	restriction NdeI-BamHI	oTK439/oTK334	pSC60	this study	
	pCV150	AVR-Pia T46N	pET15b	AVR-Pia T46N	restriction NdeI-BamHI	oTK333/oTK334	pGT5	this study	
	pDO83	AVR-Pia R23A	pET15b	AVR-Pia R23A	Quick change	oDO41/oDO42	pCV64	this study	
	pDO84	AVR-Pia R43G	pET15b	AVR-Pia R43G	Quick change	oGT09/oGT10	pCV64	this study	
	pDO85	AVR-Pia D63A	pET15b	AVR-Pia D63A	Quick change	oDO25/oDO26	pCV64	this study	
	pDO86	AVR-Pia D29A	pET15b	AVR-Pia D29A	Quick change	oDO17/oDO18	pCV64	this study	
	pDO87	AVR-Pia R36A	pET15b	AVR-Pia R36A	Quick change	oDO19/oDO20	pCV64	this study	
pDO88	AVR-Pia E58A	pET15b	AVR-Pia E58A	Quick change	oDO23/oDO24	pCV64	this study		
<i>M. oryzae</i> transgenic	pCV76	AVR-Pia	pDL02	AVR-Pia	Yeast gap repair cloning	oTK344/oTK345	pSC60	this study	
	pCR17	mRFP	pDL02	mRFP	-	-	-	Ribot et al., 2013	
	pDO89	AVR-Pia R23A	pDL02	AVR-Pia R23A	Quick change	oDO43/oDO44	pCV76	this study	
	pDO90	AVR-Pia R43G	pDL02	AVR-Pia R43G	Quick change	oGT09/oGT10	pCV76	this study	
	pDO91	AVR-Pia D63A	pDL02	AVR-Pia D63A	Quick change	oDO25/oDO26	pCV76	this study	
	pDO92	AVR-Pia D29A	pDL02	AVR-Pia D29A	Quick change	oDO17/oDO18	pCV76	this study	
	pDO93	AVR-Pia R36A	pDL02	AVR-Pia R36A	Quick change	oDO19/oDO20	pCV76	this study	
	pDO94	AVR-Pia E58A	pDL02	AVR-Pia E58A	Quick change	oDO23/oDO24	pCV76	this study	
	pDO95	AVR-Pia F24	pDL02	AVR-Pia F24S	Quick change	oDO45/oDO46	pCV76	this study	

Supplemental Methods

Production and purification of recombinant proteins for ITC and NMR

The expression and purification of recombinant AVR-Pia mutant proteins were carried out as previously described for the AVR-Pia wild type (de Guillen et al., 2015). Protein expression was performed in *E.coli* BL21 (DE3) in autoinducing minimal media C-750501 (Studier, 2005) supplemented with $15\text{NH}_4\text{Cl}$, 13C_3 -glycerol and 13C_6 -glucose for NMR experiments. Transformed cells grew at 37°C for 8h then at 30°C for 16h. A standard purification for Histagged protein was used with these following purification buffers (A): 50 mM TrisHCl, pH 7.0, 300 mM NaCl, 1mM DTT, 0.1 mM Benzamidine; (B): A supplemented with 500 mM imidazole) then a size exclusion chromatography was performed in buffer A. Fractions containing the protein were identified by SDS-PAGE, pooled and concentrated in a Centricon®. The RATX1 domain does not contain tryptophan and a bicinchoninic acid (BCA) assay (Pierce®) was used for protein quantification.

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were carried out on a VP-ITC isothermal titration calorimeter (Microcal, Northampton, USA) at 25°C . The protein samples were all buffer exchanged using dialysis at 4°C into the ITC buffer (20 mM potassium-sodium phosphate, pH 5.4 and 150 mM NaCl) to minimize undesirable buffer-related effects. The dialysis buffer was used in all preliminary equilibration and washing steps. Protein concentrations were measured using a NanoDrop 2000spectrometer and BCA Kit assays (Pierce®). The same RATX1 sample was used in all the binding reactions. Titration of RATX1 (80uM) in the cell (2mL) was performed by sequential addition of AVR-Pia (WT and H3 at 1 mM; 22 injections of 10uL). Data were analysed by a single site model using the Origin Software. A 1H-1D-NMR control spectrum was recorded for the AVR-Pia-H3 sample. All protein samples were checked by SDS-PAGE and by gel exclusion chromatography applied to the protein mixtures at the end of the ITC experiments.

NMR samples

The NMR samples for assignments and structure calculations were prepared with 1mM purified protein at 10% D₂O and 0.5 mM DSS as a reference. The purification buffer was exchanged with phosphate buffer (20 mM potassiumsodium phosphate, pH 5.4 and 150 mM NaCl), by filtrating with Centricon®. For the D₂O experiments, a ^{15}N labeled sample was lyophilized and dissolved in D₂O.

^{15}N backbone amide NMR Relaxation data

Relaxation data for AVR-Pia-H3 were acquired at 305K on a Bruker Avance 500 MHz spectrometer using R1, R2 and $^{15}\text{N}\{^1\text{H}\}$ heteronuclear NOE pulse sequences (TOPSPIN library, v 2.1) with a protein sample at 1 mM and the experimental settings already given in (de Guillen et al., 2015). Relaxation parameters, R1, R2 and NOEs were determined from the analysis module of CCPN (Vranken et al., 2005). Lipari-Szabo analysis was performed using the DYNAMOF software (Barthe et al., 2006).

Structure Calculation

The programs CYANA (Güntert, 2004), 2004) and CNS (Brunger, 2007) were used for automatic NOE assignments and structure calculations. The NH, Ha, ^{15}N , ^{13}Ca and ^{13}Cb chemical shifts were converted into F/Y dihedral angle constraints using TALOS+ (v. 1.2) (Shen et al., 2009). Final structure calculations were performed with CYANA (v. 2.1) using 1366 distance restraints and 100 F/Y dihedral angle constraints. The 20 conformers with lowest target function starting from 200 initial structures, were refined by CNS (v. 1.2) using the refinement in water of RECOORD (Nederveen et al., 2005). These are the structures discussed herein and deposited (PDBs, 5JHJ). The final 20 structures contained no NOE violations greater than 0.4 Å and no dihedral angle constraint violations greater than 2°. Structures were validated using PROCHECK(Laskowski et al., 1993) (Table S2).

AVR-Pia / RATX1 Docking Procedure

Three models of RATX1 have been obtained from the I-tasser server (<http://zhanglab.ccmb.med.umich.edu/ITASSER>), providing the chain B of the 5A6W PDB structure as an explicit homologous template. The obtained Itasser models were docked onto the AVR-Pia NMR structure (PDB identifier 2MYW) using the protein-protein docking program RosettaDock from the Rosetta modeling suite [2015.12.57698]. Each docking simulation was initialized by randomizing the relative positions and orientations of both partners and was performed with default parameters. The following AVR-Pia residues were restrained to lie closer than 12 Å from the RATX1 residues using Rosetta site constraints: R23, F24, V37, L39, M40, Y41, V42, R43, T46, T47, A48, T49, T51, T56 and E58. These residues correspond to the AVR-Pia central surface displaying the highest chemical shift differences from the NMR titration with RATX1. For each of the 3 RATX1 models, 3000 complexes were built with the same protocol. The 2% best scored complexes (180/9000) were clustered according to the orientation of AVR-Pia relatively to RATX1. All complex interfaces of the resulting clusters were evaluated using Pisa from the CCP4 suite. Some conclusions arising from these simulations could be stressed:

- 1) In all but one clusters, the central residue on AVR-Pia interface is M40, which is also central in the NMR mapped surface. The last cluster has a slightly displaced interface whose central residue is Y41.
- 2) The complex clusters display various interface orientations on RATX1 and none of them can be clearly distinguished neither from the number of poses, the Rosetta scores nor from the Pisa ΔGs .
- 3) None of the best clusters has an interface oriented as in the AVR-Pik / HMA complex. Other docking simulations done using the Zdock server (<http://zdock.umassmed.edu>) instead of Rosetta also failed to reproduce an interface oriented as seen in the AVR-Pik / HMA complex.

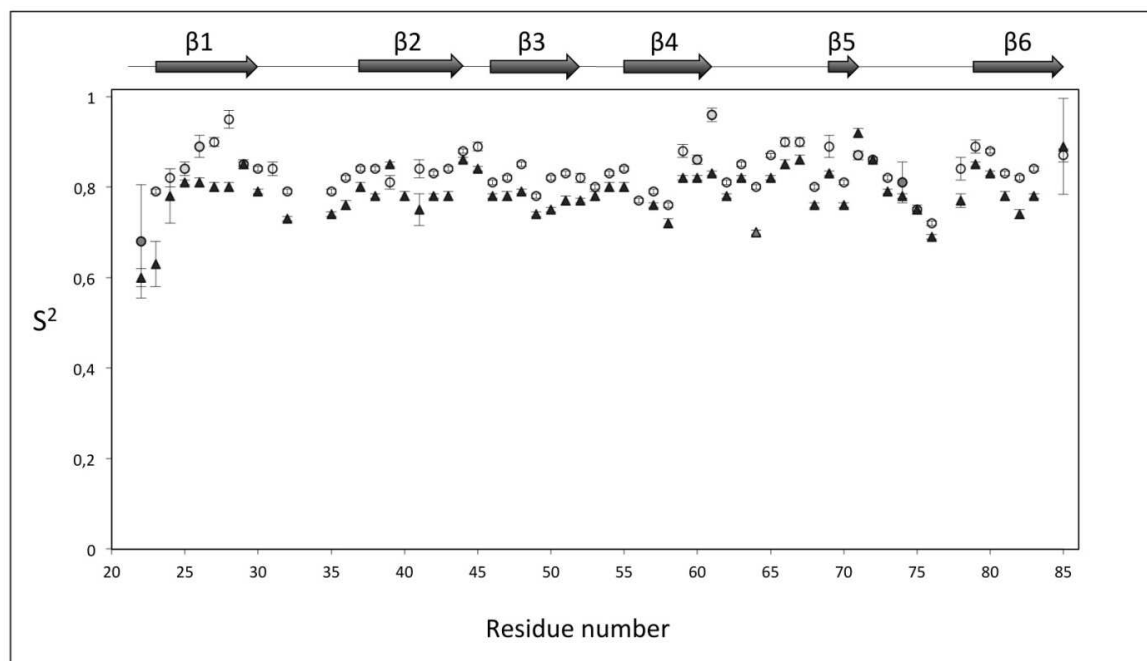
Protein extraction and Western blotting

Yeast proteins were extracted with a post-alkaline extraction method (Kushnirov, 2000), resuspended in Laemmli buffer, boiled 10 min at 90°C and separated in 12% Tris-tricine SDS-PAGE gels (GE Healthcare). After transfer to nitrocellulose membrane (Millipore), protein detection was performed using α -AVR-Pia antibodies (1000x dilution) raised against purified recombinant AVR-Pia protein in rabbits (Eurogentec).

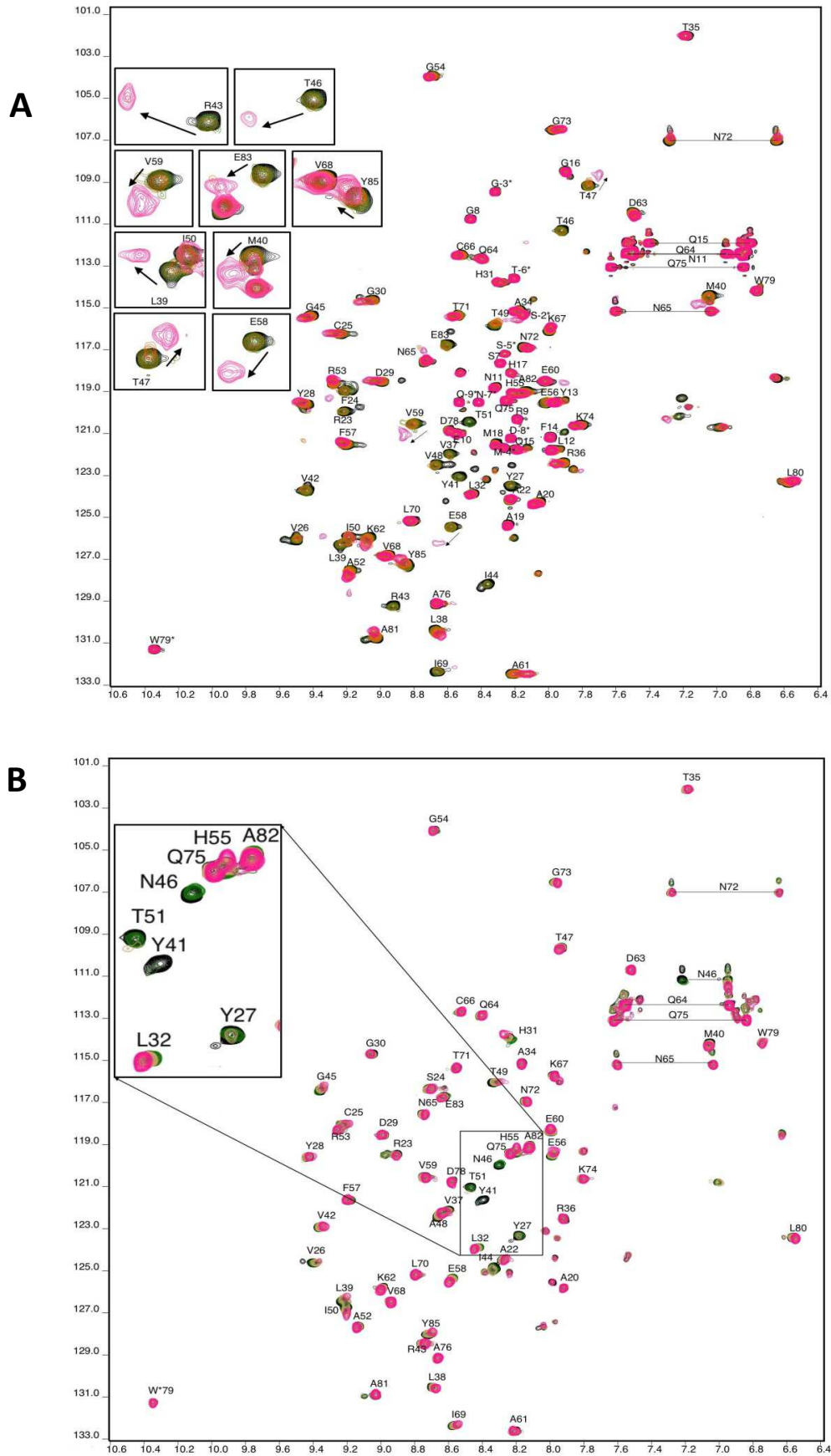
Proteins from fungal culture filtrates were dialyzed against water (cut-off 1kDa, 1 part of medium by 5 parts of water), lyophilized, resuspended in 10 mM Tris-HCL pH 7.5, separated on 16% Tris-tricine NuPAGE gels (GE Healthcare) and, after blotting, analyzed using a-AVR-Pia antibodies.

Supplemental References

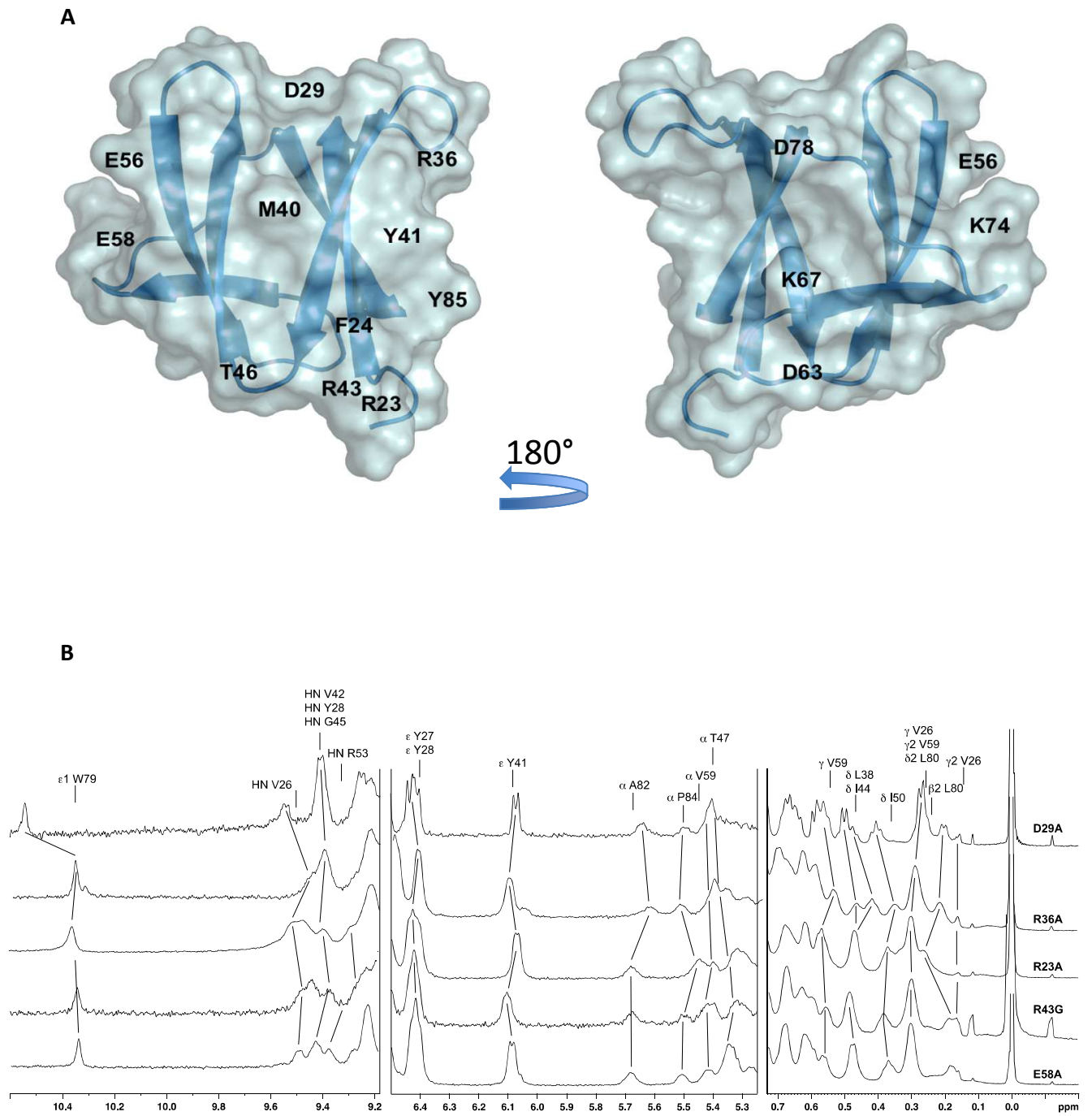
- Barthe, P., Ropars, V., and Roumestand, C.** (2006). DYNAMOF: a program for the dynamics analysis of relaxation data obtained at multiple magnetic fields. *Comptes Rendus Chim.* **9**: 503–513.
- Brunger, A.T.** (2007). Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2**: 2728–2733.
- Güntert, P.** (2004). Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**: 353–378.
- Kushnirov, V. V.** (2000). Rapid and reliable protein extraction from yeast. *Yeast* **16**: 857–860.
- Laskowski, R.A., Moss, D.S., and Thornton, J.M.** (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**: 1049–1067.
- Nederveen, A.J. et al.** (2005). RECOORD: A recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins Struct. Funct. Bioinforma.* **59**: 662–672.
- Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A.** (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**: 213–223.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D.** (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Bioinforma.* **59**: 687–696.



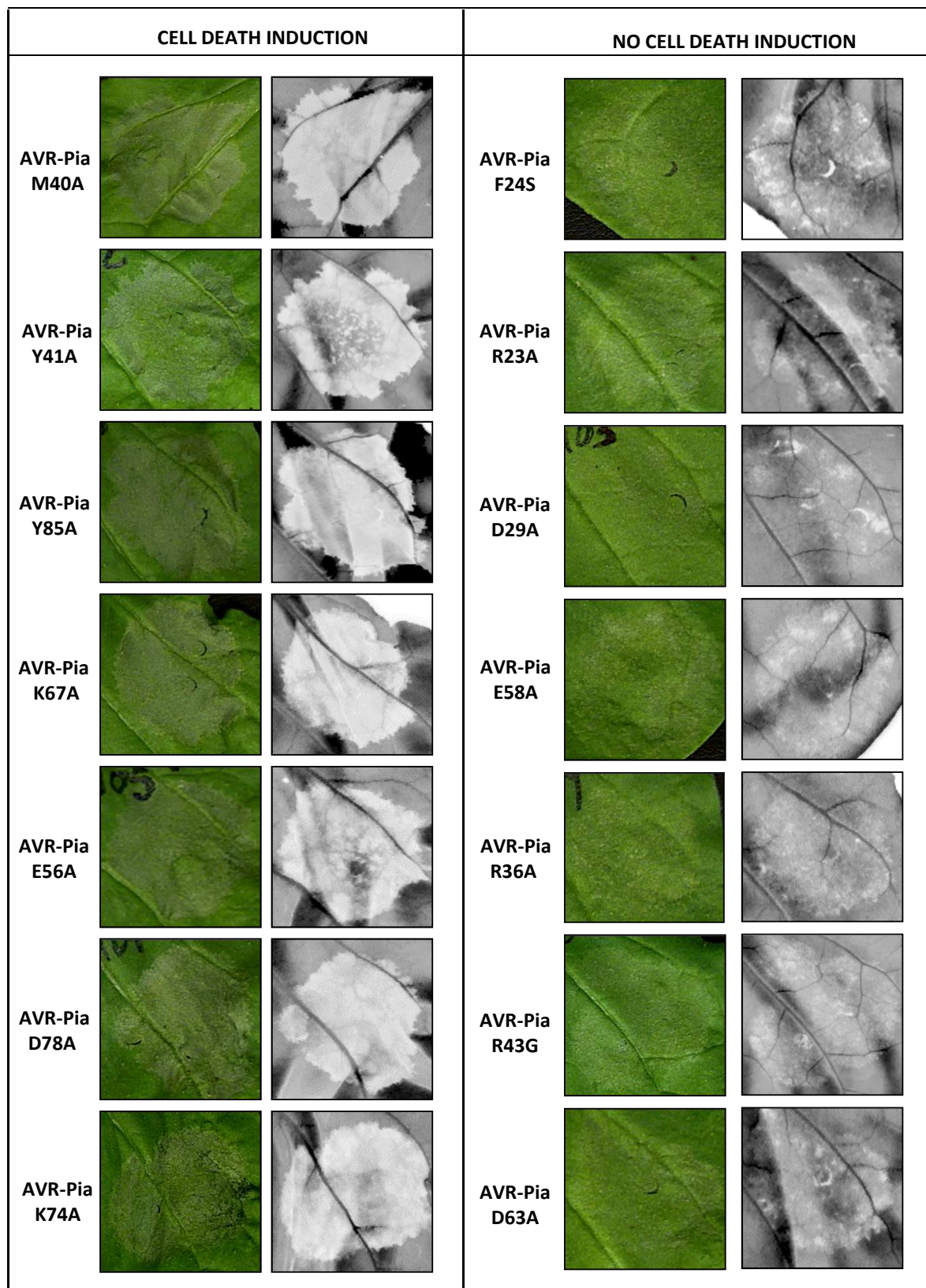
Supplemental Figure 1. Comparison of NMR relaxation of AVR-Pia and AVR-Pia-H3.



Supplemental Figure 2. HSQC spectra of AVR-Pia and AVR-Pia-H3 recorded upon titration with the RGA5_{RATX1} domain

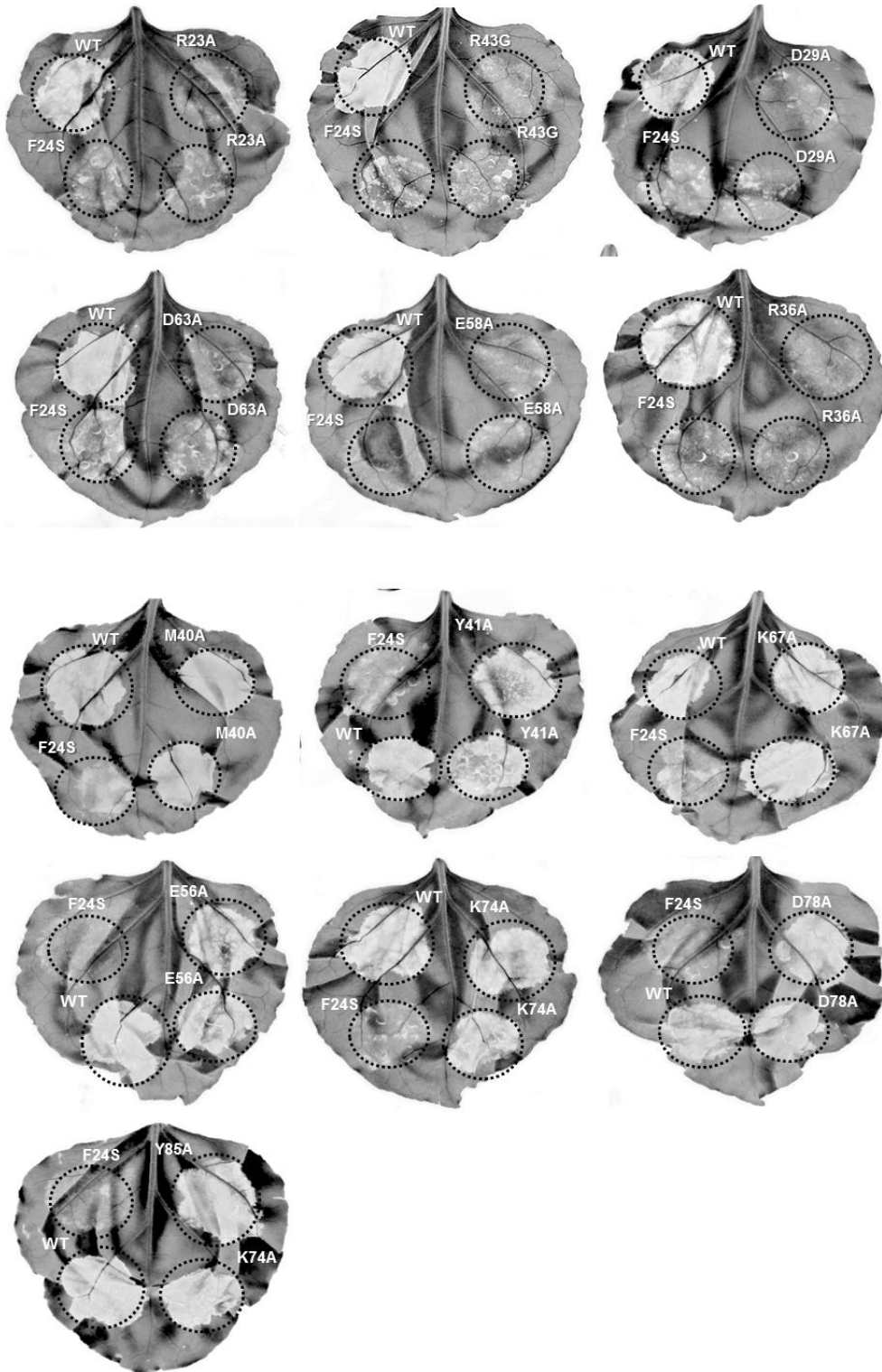


A

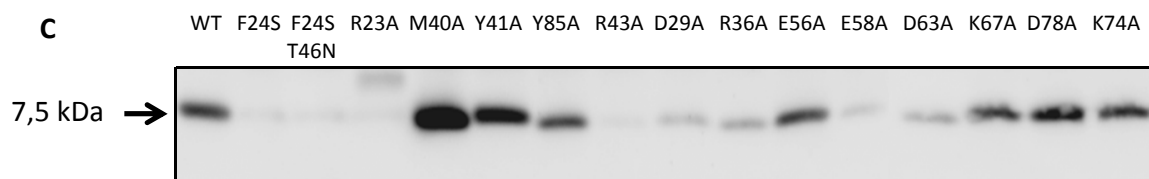


Supplemental Figure 4. AVR-Pia mutants not-affected in RGA5_{RATX1}-binding trigger HR in *N. benthamiana*

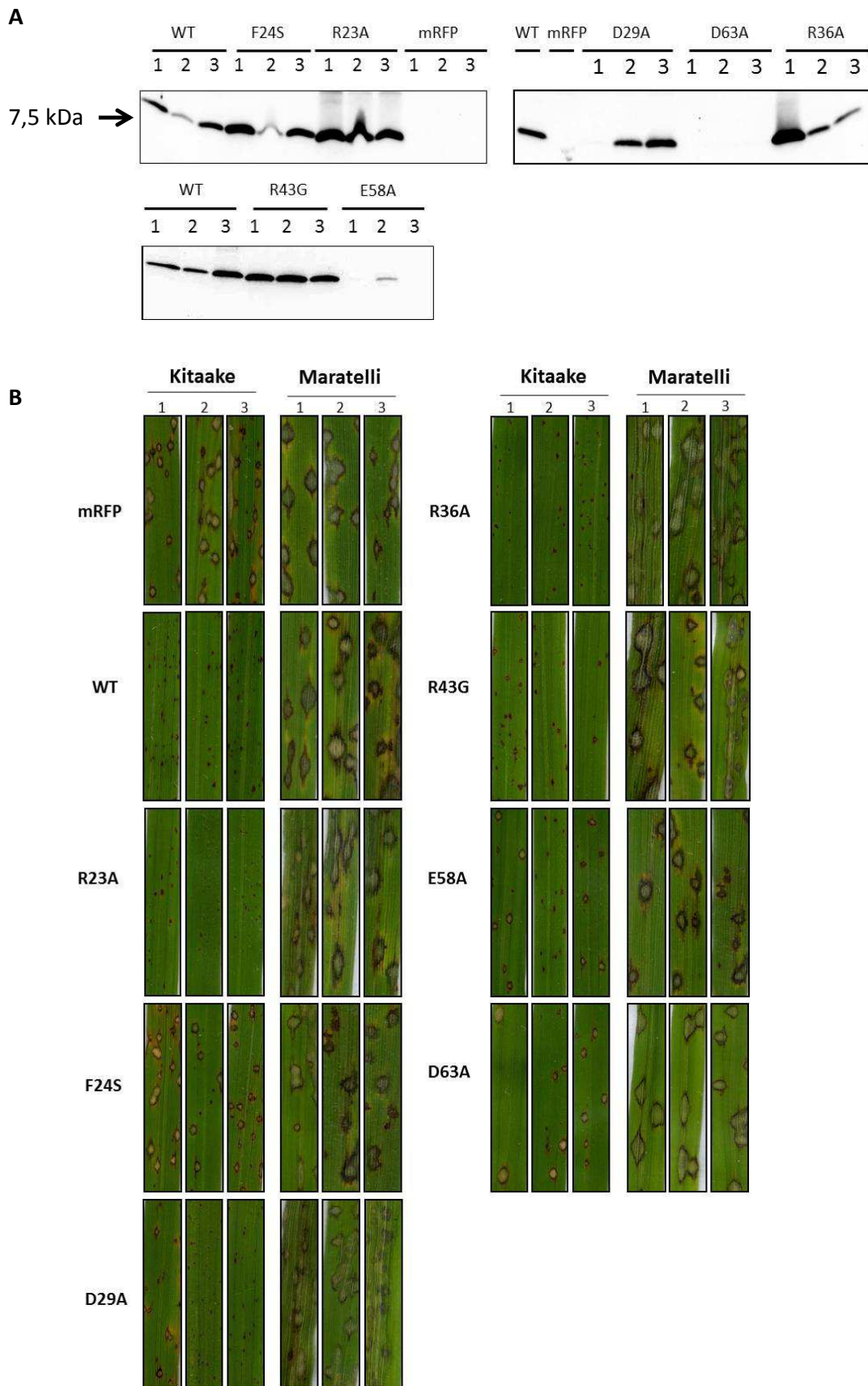
B



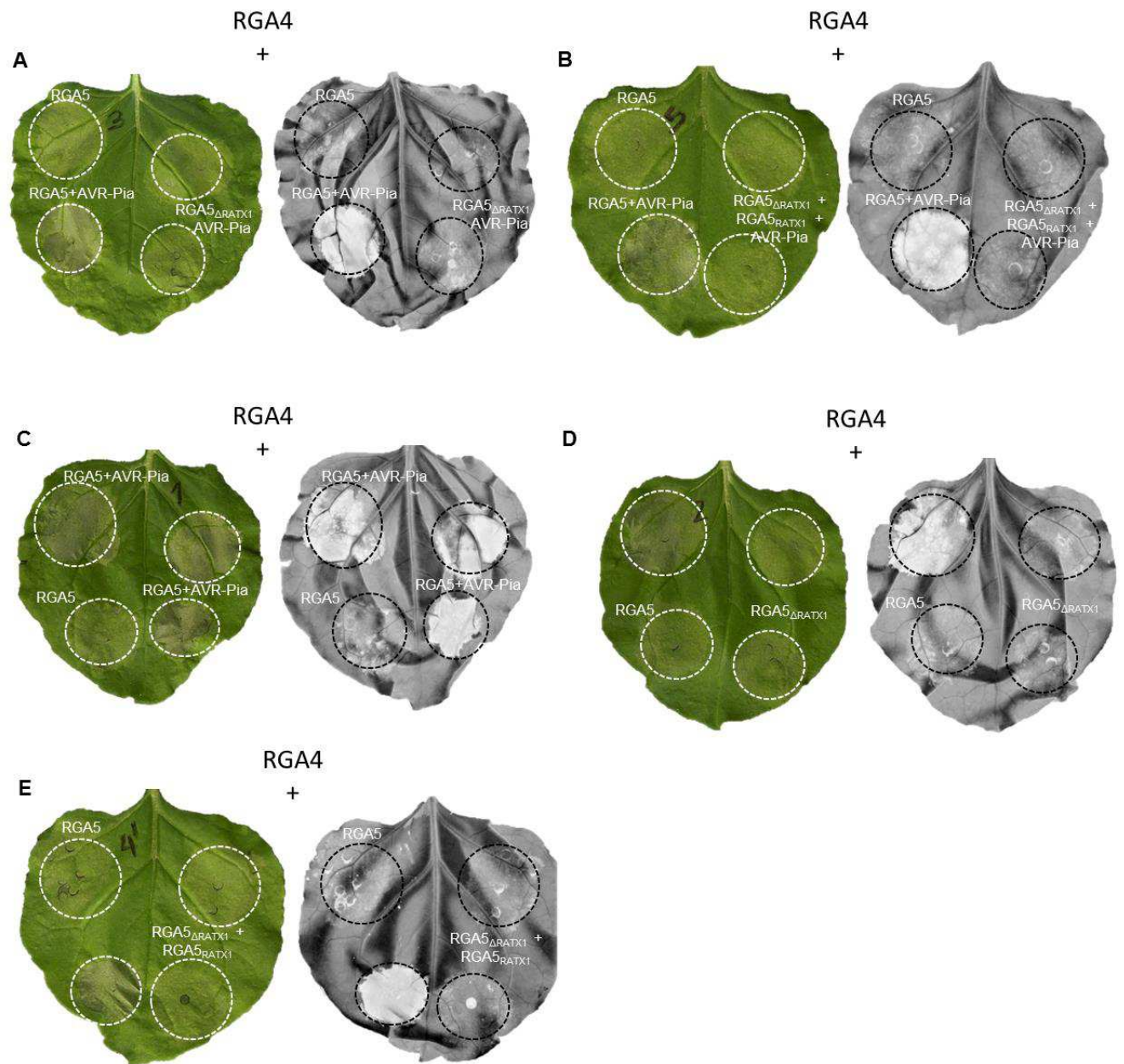
C



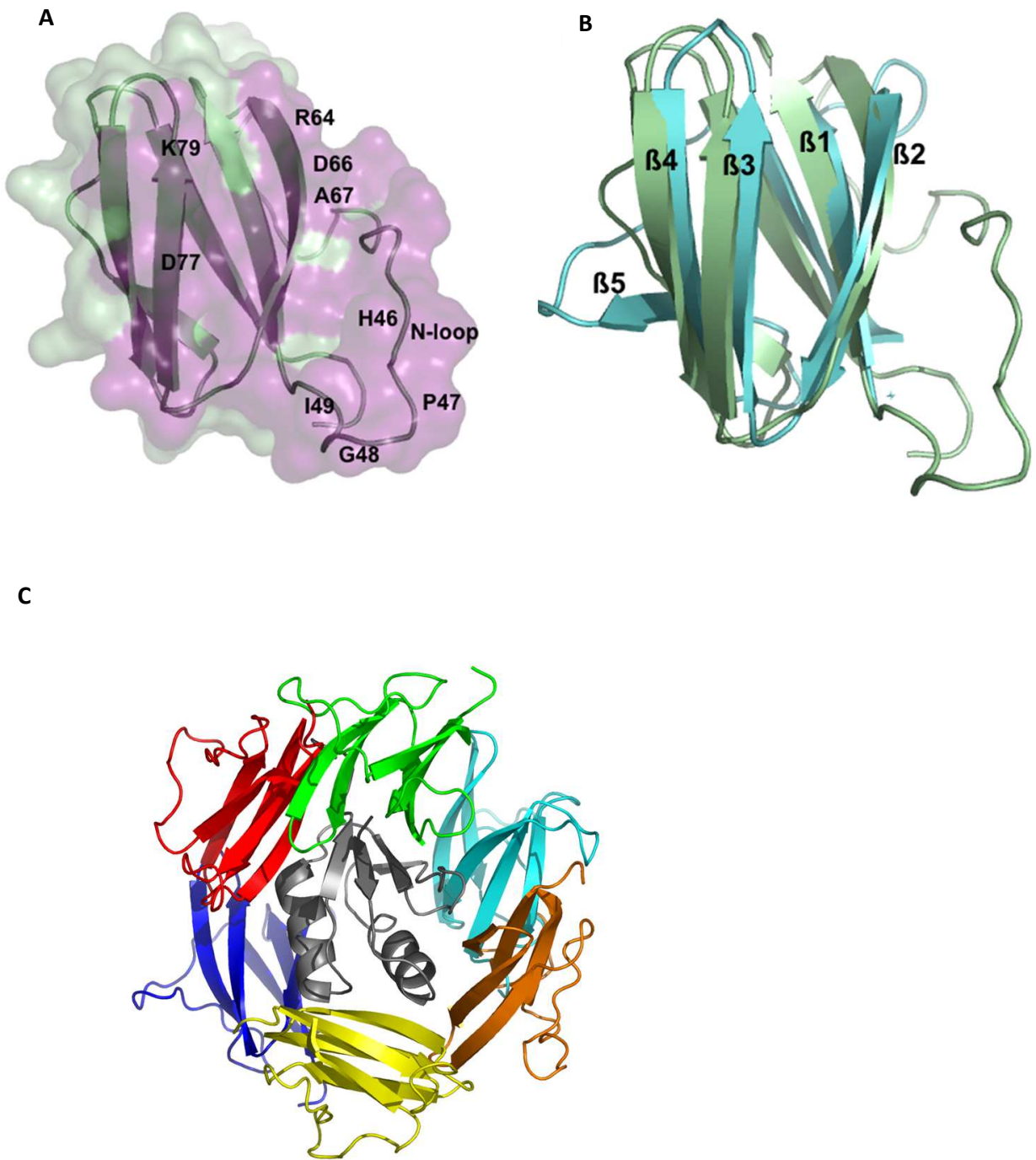
Supplemental Figure 4. AVR-Pia mutants not-affected in RGAS₅_{RATX1}-binding trigger HR in *N. benthamiana*



Supplemental Figure 5. Characterization of transgenic *M. oryzae* isolates carrying *AVR-Pia* mutant constructs.



Supplemental Figure 6. RGA5 Δ RATX1 represses RGA4-mediated cell death but does not recognize AVR-Pia.



Supplemental Figure 7. Comparison of the AVR-Pia and AVR-PikD structures and their complexes with RAX1/HMA domains.

Supplemental Figure Legends

Supplemental Figure 1. Comparison of NMR relaxation of AVR-Pia and AVR-Pia-H3.

Generalized order parameter (S_2) obtained from ^{15}N relaxation data at 500 MHz for AVR-Pia (circles) and AVR-Pia-H3 (triangles) determined from Lipari-Szabo formalism with the program DYNAMOF (Barthe et al., 2006). The different colors indicate the use of the “simple” Lipari-Szabo formalism by black triangles for AVR-Pia-H3, and by white circles for AVR-Pia, with possible additional exchange contributions shown by the symbols in grey color.

Supplemental Figure 2. HSQC spectra of AVR-Pia and AVR-Pia-H3 recorded upon titration with RGA5RATX1.

HSQC spectra are shown for 50 μM AVR-Pia (A) and AVR-Pia-H3 (B) without RGA5RATX1 (black) and various concentration of RGA5RATX1 at 12.5 μM ($R=0.25$ in grey), 25 μM ($R=0.5$ in green), 50 μM ($R=1$ in orange) and 100 μM ($R=2$ in pink). Residues belonging to the N-terminal tag are indicated by negative numbering and (*). For AVR-Pia (A) some of the largest chemical shift differences between the free ($R=0$) and bound ($R=2$) form are shown in the zoom panels at the upper left corner. With AVR-Pia-H3 (B), most of the HSQC peaks remain unperturbed and the enlarged plot shows the peaks present at the beginning of the titration for Y27, Y41, T51, N46 (in black or green for $R=0$ or $R=0.5$) that are missing at the end of the titration (in pink, $R=2$).

Supplemental Figure 3. AVR-Pia mutants affected in RGA5RATX1-binding are well-structured.

(A) AVR-Pia NMR structure (Guillen et al., 2015) showing the amino acids that were replaced in the AVR-Pia variants carrying mutations in the candidate interaction surface (left) and, for control purposes, in others surfaces (right). (B) ^1H 1D-NMR spectra for AVR-Pia point mutants E58A, R43G, R23A, R36A and D29A. Chemical shift assignments of wild type AVR-Pia are shown by vertical bars above the spectra.

Supplemental Figure 4. AVR-Pia mutants not-affected in RGA5RATX1-binding trigger HR in *N. benthamiana*

HA:RGA5, RGA4:HA and untagged AVR-Pia variants were transiently expressed in *N. benthamiana* leaves by *A. tumefaciens* infiltration and cell death responses (A and B) and AVR-Pia protein levels were determined (C). (A) The HR response was evaluated after 4 days by visual inspection (left) and a lack of chlorophyll fluorescence (right). (B) The activity of AVR-Pia mutants was determined in comparison to the positive control AVR-Piawt and the negative control AVR-PiaF24S. (C) Proteins were extracted 48 h after infiltration, immunoprecipitation of AVR-Pia variants was performed with α -AVR-Pia antibodies and anti- protein A/G agarose beads (α -protein A/G) due to low abundance of certain AVR-Pia variants and the immunoprecipitate was analyzed by immunoblotting with α -AVR-Pia antibodies.

Supplemental Figure 5. Characterization of transgenic *M. oryzae* isolates carrying AVR-Pia mutant constructs.

(A) Culture filtrates of 3 different transgenic *M. oryzae* isolates per construct grown for 5 days in a liquid complete medium was analyzed by immunoblotting with a-AVR-Pia antibodies for expression and secretion of AVR-Pia mutant proteins. (B) For each AVR-Pia mutant construct, as well as the controls AVR-Piawt and mRFP, three independent transgenic *M. oryzae* isolates were sprayed on resistant Kitaake (+Pia) and susceptible Maratelli (-Pia) rice plants. Symptoms were recorded 7 days after inoculation.

Supplemental Figure 6. RGA5 Δ RATX1 represses RGA4-mediated cell death but does not recognize AVR-Pia.

AVR-Pia and RGA4:HA were co-expressed with HA:RGA5 Δ RATX1 (A) or HA:RGA5 Δ RATX1 and HA:RGA5RATX1 (B) to evaluate the importance of RGA5RATX1-binding for AVR-Pia- recognition. As controls, the de-repression of RGA4 by AVR-Pia in the presence of HA:RGA5 (C) and the repression of RGA4 -mediated cell death by HA:RGA5 Δ RATX1 (D) or HA:RGA5 Δ RATX1 and HA:RGA5RATX1 (E) are shown. On all leaves, cell death induced by RGA4:HA and repression of cell death upon co-expression of RGA4:HA and HA:RGA5 were recorded as positive and negative controls. Cell death was evaluated after 4 days by visual inspection (left) and lack of chlorophyll fluorescence (right).

Supplemental Figure 7. Comparison of the AVR-Pia and AVR-PikD structures and their complexes with RAX1/HMA domains.

(A) AVR-PikD structure with the Pikp-1HMA interaction surface in pink and labeling of important residues from the N-terminal loop, β 2 and β 3 (based on PDB structure 5A6W, Maqbool *et al.*, 2015). (B) Superposition of AVR-Pik and AVR-Pia structures using DALI software. The rmsd over 50 residues is 2.4 Å. (C) AVR-Pia / RATX1 docking simulations using the Rosetta suite. One copy of RATX1 is displayed as a grey cartoon in the middle, and was used as a reference frame for the superposition. The top green model corresponds to the position of AVR-PikD in the AVR-PikD/ Pikp-1HMA complex (5A6W), while the other colored models are AVR-Pia from the docking by RosettaDock. Each colored AVR-Pia model is representative of a cluster of models sharing the same orientation relative to RATX1. Ordered with respect to their relative size, from the biggest to the smallest cluster, they are red, orange, yellow, cyan and blue.

DISCUSSION

&

PERSPECTIVES

1. Discussion

L'étude des interactions hôte-pathogène est d'une importance capitale pour mieux comprendre les mécanismes de virulence du pathogène et les mécanismes de reconnaissance conduisant à une réponse immunitaire efficace pour l'hôte. L'amélioration de ces connaissances permettrait d'envisager de nouvelles solutions de lutte contre les agents pathogènes.

Le but de mon projet de recherche était d'étendre et d'améliorer les connaissances sur les protéines effecteurs de *M. oryzae* et plus particulièrement les protéines AVR1-CO39 et AVR-Pia. Ainsi la résolution des structures de ces effecteurs réalisée au cours de cette thèse a permis d'aborder plusieurs questions :

- Ces deux protéines si différentes en séquence partagent-elles le même repliement ?
- Un repliement est-il conservé pour les effecteurs d'autres champignons phytopathogènes?

Ce travail de thèse a également permis d'investiguer une partie des mécanismes moléculaires mis en œuvre dans la reconnaissance entre les effecteurs et les protéines de résistance du riz en abordant la question essentielle de leur interaction et plus précisément:

- Comment l'effeteur AVR-Pia interagit-il avec le domaine RATX1 issu de la protéine de résistance RGA5 ?

Ce travail de thèse a permis la rédaction de deux articles présentés dans ce manuscrit.

1.1 La famille d'effecteurs MAX

Aucune structure d'effeteur de *M. oryzae* n'étant caractérisée dans la PDB, c'est dans le but de caractériser les structures tridimensionnelles d'AVR1-CO39 et d'AVR-Pia que c'est inscrit ce travail de thèse.

J'ai résolu les structures de ces deux effecteurs par RMN améliorant ainsi les lacunes dans ce domaine. De façon surprenante, bien que ces deux protéines n'aient que peu d'homologie de séquence primaire, elles possèdent une structure tridimensionnelle similaire en sandwich β . Ce cas

de figure est retrouvé chez les effecteurs d'oomycètes qui bien que possédant des séquences primaires différentes partagent des motifs structuraux communs. Dans le même temps, la première structure d'effecteur de *M. oryzae*, celle d' AvrPiz-t, a été publiée (Zhang et al., 2013) dont la cible de virulence APIP6 avait été identifiée récemment (Park et al., 2012). De manière intéressante le repliement de l'effecteur AvrPiz-t est également similaire à celui d'AVR1-CO39 et d'AVR-Pia. Les récents travaux sur la toxine ToxB du pathogène du blé *Pyrenophora tritici-repentis* ont mis en évidence le même type de repliement trouvé chez AvrPiz-t (Nyarko et al., 2014). Ainsi la superposition des structures des ces quatre protéines indique un repliement similaire. Cette caractéristique nous a permis de mettre en évidence une famille d'effecteurs partageant la même structure mais possédant des séquences primaires différentes que nous avons nommée MAX-effecteurs (*M. oryzae* Avrs and ToxB). Basées sur ces informations de structures nous avons mené des recherches dans des bases de donnée de génome entier de champignon qui révèlent que la famille des MAX effecteurs est majoritairement présente chez les phytopathogènes ascomycètes. En effet ce type d'effecteurs a subi une forte expansion spécifique à la lignée de champignon *Pyriculariae* dont fait partie *M. oryzae*. Nous avons réalisé une analyse du profil d'expression des gènes candidats codant des protéines appartenant à la famille des MAX-effecteurs. Cette analyse montre que ces gènes candidats sont majoritairement exprimés pendant la phase biotrophe de colonisation, pouvant ainsi être considérés comme des effecteurs potentiels de virulence.

D'autre part l'identification des cibles de virulence d'AVR1-CO39, d'AVR-Pia, ainsi que de ToxB est une première étape afin d'obtenir des détails moléculaires liés à leur fonction dans le processus de virulence et l'approche structurale fournira des informations importantes pour mieux comprendre quelle est la relation entre leur structure et leur fonction.

A l'heure actuelle seule la cible de virulence d'AvrPiz-t a été identifiée. En effet AvrPiz-t a été décrite comme ciblant le système ubiquitine du protéasome de l'hôte en se liant et en inactivant la protéine ubiquitine ligase RING E3 nommée APIP6 (Park et al., 2012). Ainsi la fonction de l'effecteur AvrPiz-t est d'inhiber la réponse induite par les PAMP (PTI) pour promouvoir la virulence du champignon. L'effecteur AvrPiz-t est reconnu dans la plante par le NLR Piz-t (Ashikawa et al., 2008), cependant les détails du mécanisme de reconnaissance restent inconnus.

Notre étude a permis l'identification d'une famille de MAX effecteurs basée sur une similarité de repliement. L'autre question que l'on peut poser est : est-ce que la fonction de ces effecteurs dans le déclenchement de la réponse immune est aussi conservée ? Le fait qu'ils soient majoritairement sécrétés pendant la phase précoce de l'infection indique qu'ils favorisent le développement de la maladie, mais sont-ils également ciblés par le même type de protéines de résistance dans la plante et cela implique-t-il les mêmes mécanismes de reconnaissance ?

Pour les effecteurs AVR-Pia et AVR1-CO39, les bases de la reconnaissance moléculaire par leur récepteur immunitaire NLR commencent à être mieux comprises. Ces deux effecteurs sont reconnus par la même paire de NLR RGA4/RGA5 et plus spécifiquement par le domaine RATX1 (ou HMA pour

Heavy Metal Activated domain) porté par la protéine RGA5 (Cesari et al., 2013). Ce domaine est considéré comme un « leurre intégré » dans la protéine de résistance en mimant la cible réelle de virulence (Cesari et al., 2014a). Ainsi, il semble que l'effecteur AvrPiz-t d'un côté et les effecteurs AVR1-CO39 et AVR-Pia de l'autre possèdent des activités moléculaires différentes et des cibles de virulence différentes au sein de la plante. On observe entre la structure d'AvrPiz-t et des deux autres effecteurs une certaine divergence dans leur forme, notamment au niveau des boucles, ainsi que dans leurs propriétés de surface, ce qui renforce l'idée qu'ils sont impliqués dans des voies et des mécanismes de régulation différents.

Les effecteurs AVR1-CO39 et AVR-Pia partagent un patch hydrophobe au niveau de la surface des brins $\beta 1$ $\beta 2$ $\beta 6$. Ce type de surface pourrait être impliqué dans les interactions protéine-protéine. Étant donné qu'ils interagissent également avec le même récepteur immunitaire en se liant avec le même domaine senseur, l'identification des acides aminés clés impliqués dans cette interaction pour chacun des effecteurs est d'une importance capitale pour une meilleure compréhension du système.

Récemment la structure d'un autre effecteur de *Magnaporthe oryzae*, AVR-Pik, a été publiée (Maqbool et al., 2015). AVR-Pik appartient à la famille des MAX effecteurs, cet effecteur présente une homologie structurale au niveau de son repliement en sandwich β mais possède une extension N-terminale qui lui est propre. De façon intéressante, la paire de NLR, Pikp-1/Pikp-2 impliquée dans l'interaction avec AVR-Pik (Ashikawa et al., 2008) lie l'effecteur via un domaine HMA, mais ce domaine est intégré entre les régions CC et NB-ARC de Pikp-1 (Maqbool et al., 2015). Les domaines HMA intégrés chez Pikp-1 et RGA5 ne partagent pas les mêmes spécificités de reconnaissance, en effet chacun des domaines est spécifique à son effecteur, ou à ses deux effecteurs pour le domaine HMA de RGA5 (Cesari et al., 2013). La structure d'AVR-Pik a été obtenue en complexe avec le domaine HMA de Pikp-1. Trois sites d'interaction apparaissent au niveau du complexe impliquant au niveau d'AVR-Pik les résidus du brin $\beta 3$, les résidus Arg64 et His46, ainsi que la partie N-terminale de Asn42 à Ile49. Ces sites d'interactions sont positionnés de manière différente de la surface hydrophobe observée dans le cas d'AVR1-CO39 et d'AVR-Pia. L'étude du complexe entre les partenaires permettrait d'avoir une meilleure compréhension du système et de comparer les sites d'interaction.

L'une des données d'interaction disponible est celle du mutant d'AVR-Pia nommé AVR-Pia-H3. Ce mutant est un double mutant (F24S, T46N) codé par un allèle naturel de *M. oryzae*, l'allèle H3, qui a perdu la capacité de se lier au domaine RATX1 de RGA5 et qui n'active pas la résistance. Le résidu phénylalanine 24 est situé au niveau de la surface hydrophobe et la thréonine 46 en bordure. Ces informations données par la structure vont permettre de guider des études fonctionnelles pour savoir si les acides aminés situés en bordure ou dans le patch hydrophobe sont impliqués dans l'interaction avec le domaine RATX1.

De plus le MAX-effecteur ToxB possède également un mutant non fonctionnel, ayant perdu la capacité à induire la mort cellulaire de la cellule. Ce mutant, *tox*_b, possède 13 acides aminés

différents par rapport à la forme active. L'analyse des structures de ces deux protéines montre qu'elles conservent la même structure. Cependant des analyses de relaxation par RMN, permettant d'accéder à la dynamique des protéines, montrent que la protéine *toxb* est plus flexible que ToxB. Ainsi ToxB semble appartenir à cette classe de protéines dont l'activité peut être modulée par des changements de dynamique plutôt que de structure. L'augmentation de la flexibilité, comme c'est le cas pour *toxb*, implique une pénalité entropique conformationnelle plus importante, ce qui peut conduire à une interaction plus faible ou abolie avec sa cible et ainsi induire une perte de l'activité (Nyarko et al., 2014). Il sera intéressant de vérifier une telle hypothèse dans le cas d'AVR-Pia-H3.

Ainsi dans la deuxième partie de mon travail de thèse je me suis principalement intéressée à l'étude de l'interaction entre AVR-Pia et le domaine RATX1 de RGA5.

1.2 Cartographie AVR Pia-RATX1

Dans ce travail nous nous sommes intéressés aux mécanismes moléculaires impliqués dans la reconnaissance de l'effecteur et dans l'activation de la résistance dans l'interaction entre AVR-Pia et le récepteur immunitaire NLR RGA5.

Dans cette étude l'interaction a été caractérisée dans un premier temps *in vitro*, par l'étude spécifique des protéines recombinantes AVR-Pia et AVR-Pia H3 et RATX1.

L'analyse de la structure du mutant AVR-Pia H3 montre qu'elle est similaire à celle de la protéine sauvage. De plus, les études de dynamique par relaxation ^{15}N montrent que ce mutant est tout aussi rigide que la protéine sauvage, ce qui permet d'éliminer l'hypothèse que l'interaction avec le domaine RATX1 soit abolie du fait de la flexibilité de la molécule.

L'étude de l'interaction réalisée par ITC montre que l'affinité entre AVR-Pia et RATX1_{RGA5} n'est pas très forte, de l'ordre de quelques micromolaire, et que celle du mutant est au moins dix fois plus faible. D'une manière intéressante l'affinité dans les cas d'AVR-Pik et du domaine HMA de Pik est bien plus forte, de l'ordre d'une centaine de nanomolaire. L'effecteur AVR-Pik possède différents allèles (A, C, D, et E) qui ont été testés dans les expériences d'interaction par résonance plasmonique de surface. La meilleure affinité obtenue avec HMA_{PikD} est pour AVR-PikD avec un Kd de 31nM (Maqbool et al., 2015). Ainsi l'interaction entre AVR-PikD et HMA_{PikD} est forte alors que l'interaction entre AVR-Pia et RATX1_{RGA5} est plutôt intermédiaire. Cela suggère que même si les partenaires composant ces complexes protéiques se ressemblent, les mécanismes de reconnaissance et d'activation sont différents.

Lors de cette étude, la réalisation d'un titrage par RMN nous a permis d'avoir accès aux acides aminés clés impliqués dans l'interaction. Ainsi grâce à cette expérience nous avons pu cartographier

la surface de reconnaissance sur la structure d'AVR-Pia en liaison avec le domaine RATX1. La surface de liaison est constituée des acides amines formant les brins β 1, 2 et 3, de deux glutamates du brin β 4 (E56 et E58) et de la tyrosine C-terminale Y85. Cette région se superpose parfaitement avec le patch hydrophobe que nous avons déterminé précédemment. Cependant il s'agit d'une vision partielle de l'interaction. En effet, la cartographie de l'interaction a été réalisée sur la protéine AVR-Pia et sur le mutant AVR-Pia-H3. Les problèmes expérimentaux rencontrés au cours de l'obtention de la protéine RATX1 ne m'ont pas permis d'obtenir suffisamment d'échantillon pour envisager une étude de sa structure par RMN ou par cristallographie.

L'un des points important à explorer lors de l'étude d'une interaction entre deux partenaires est de ne pas oublier le contexte biologique. Ainsi les résultats obtenus sur les protéines recombinantes *in vitro* doivent être complétées par des données biologiques. A partir des informations obtenues par la structure, des mutants d'AVR-Pia ont été générés, et leur interaction testée avec RGA5 dans le système hétérologue *N. benthamiana*, chez la levure et également chez le riz. Ces résultats confirment que la surface d'interaction identifiée *in vitro* est bien celle impliquée dans la reconnaissance *in vivo*. Cependant les études *in vivo* n'impliquent pas uniquement le domaine RATX1 mais aussi la participation d'autres domaines de RGA5. Ainsi, même si le domaine RATX1 n'est pas uniquement requis pour l'interaction avec AVR-Pia, il est essentiel pour l'activation complète de la protéine de résistance RGA5 et donc dans l'activation de la résistance. Il sera intéressant de déterminer à l'avenir l'affinité de ces domaines de « leurre intégré » pour leurs effecteurs par des études comparatives. Dès à présent, des études d'interaction par la protéine de reconnaissance Pik-1 avec ses effecteurs montre une diminution de la force de l'interaction. Des mutants d'AVR-PikD (AVR-PikD^{A67D} et AVR-PikD^{P47A G48D}) possédant une interaction plus faible *in vitro* et en double hybride sont cependant toujours parfaitement reconnus par Pikp-1/Pikp-2 dans des tests de co-expression transitoire et dans les expériences d'infection avec des isolats transgéniques de *M. oryzae* chez le riz. Des futures recherches devraient expliquer si cela est dû à des sites de liaisons additionnels pour AVR-Pik sur Pikp-1.

D'autre part, alors que dans la paire de NLR RGA4/RGA5, RGA4 active de façon constitutive la mort cellulaire chez *N. benthamiana* (Cesari et al., 2014a), pour Pikp-2 ce n'est pas la cas. La co-expression de Pikp-1, Pikp-2 et AVR-PikD est nécessaire pour obtenir une réponse hypersensible (Maqbool et al., 2015). L'assemblage de ce complexe semble être nécessaire pour l'activation de la signalisation. Il est intéressant de noter que la structure du domaine HMA_{Pikp} a été déterminée sous forme de dimère. Ce dimère pourrait indiquer une homodimérisation de Pikp-1 à ce niveau, cependant il n'existe pas de données indiquant que la protéine entière Pikp-1 est un dimère. Ainsi même si le domaine HMA_{Pikp} dimérise cela ne veut pas forcément dire que c'est le cas dans son contexte biologique. Dans le cas du domaine RATX1 de RGA5 il n'est pas observé de dimère. De plus il a été démontré que RGA4 et RGA5 sont capables de former des homo ou des hétérodimères via leur domaine « coiled coil » (Cesari et al., 2014b). L'interaction entre les différents partenaires et l'assemblage de ces différents domaines fait partie des futures perspectives à explorer pour mieux comprendre l'activation de la signalisation conduisant à la réponse hypersensible.

2. Perspectives

Dans la première partie de cette thèse, nous nous sommes particulièrement intéressés aux protéines du champignon phytopathogène *M. oryzae*. AVR1-CO39 et AVR-Pia. Ces effecteurs sont sécrétés dans l'apoplasme, espace situé entre le champignon et la membrane plasmique de la cellule de plante. Ces effecteurs sont transloqués dans le cytoplasme par un mécanisme encore inconnu. Des études ont été réalisées sur des protoplastes de riz capables de sécréter AVR1-CO39 par le réticulum endoplasmique. L'analyse de ce phénomène par microscopie de fluorescence a permis de suivre la protéine AVR1-CO39 couplée à la mRFP. Une fois la protéine sécrétée celle-ci est capable de rentrer à nouveau dans le cytoplasme de l'hôte indiquant que la translocation est indépendante du champignon (Ribot et al., 2013). L'une des questions qui reste sans réponse est par quel phénomène les MAX effecteurs sont-ils transloqués dans le cytoplasme ? La topologie conservée de ces effecteurs en est-elle responsable et quels sont le ou les motifs cruciaux impliqués dans la translocation ? Dans la littérature, différents motifs sont décrits comme étant impliqués de la translocation comme le motif RGD de ToxA (Manning et al., 2008), le motif hydrophobe de AVR-L567 (Rafiqi et al., 2010) ou le motif RxLR des effecteurs de cette famille (Whisson et al., 2007). Rien de tel chez les effecteurs MAX n'a été jusqu'à présent décelé. L'hypothèse de la conservation topologique liée à la fonction de translocation pourrait s'appuyer sur des études de la structure sur laquelle des variations seraient introduites. Il serait possible dans un premier temps de réaliser une étude structure-fonction en ciblant certains résidus clé comme les cystéines impliquées dans les ponts disulfures. Généralement les ponts disulfures stabilisent la structure. C'est le cas pour les différents MAX effecteurs dont la structure a été préalablement déterminée, même si pour AvrPiz-t le pont disulfure conservé dans les autres n'est pas présent, malgré la présence de deux cystéines à la bonne position. La structure d'Avr-Piz-t est cependant stabilisée par un autre pont disulfure (Cys62-Cys75). Ainsi, de tels mutants pourraient ensuite être testés par microscopie de fluorescence pour contrôler si la fonction de translocation a été modifiée, par des techniques qui caractérisent cette fonction (Ribot et al., 2013). L'analyse de la structure des mutants ayant perdu l'activité pourrait éventuellement mettre en évidence des motifs structuraux conservés chez les MAX-effecteurs.

D'autre part, il serait intéressant de mieux caractériser les effecteurs n'appartenant pas à la famille des MAX effecteurs et d'identifier ainsi d'autres familles. En effet nos analyses ont montré qu'environ 10-12% des protéines sécrétées par *M. oryzae* sont prédites par le profil HMM comme appartenant à la famille des MAX effecteurs. Récemment l'étude du génome et du transcriptome d'autres isolats de *M. oryzae* ont mis en évidence la présence de nouveaux effecteurs (Dong et al., 2015b). Dans cette étude 134 potentiels nouveaux effecteurs ont été mis en évidence par le profil d'expression de leur gène, sur les 1732 protéines potentiellement sécrétées par le champignon. Il serait donc intéressant d'appliquer une stratégie de recherche d'homologues basée sur la structure,

en se basant sur les petites protéines sécrétées, entre 200 et 250 résidus, pour une étude par RMN. Dans un premier temps, il faudrait pouvoir identifier des homologues de chacune de ces séquences dans une base de données de sécrétome d'ascomycètes. Si c'est le cas, la sélection dans une de ces familles de quelques membres représentatifs pourrait être faite. Il faudrait ensuite exprimer, purifier et déterminer la structure de ces membres. L'alignement structural permettrait de créer un profil HMM pour une recherche plus exhaustive et ainsi d'étendre à des membres plus éloignés permettant de détecter une diversification. Dès lors, ce type d'étude pourrait éventuellement mettre en évidence d'autres sous-classes d'effecteurs chez les ascomycètes ayant subi aussi une diversification.

Les NLR de plante sont impliqués dans la reconnaissance hautement spécifique des facteurs de virulence du pathogène. Cette reconnaissance induit une réponse rapide et forte conduisant à la mort de la cellule. Cette technique de la « terre brûlée » empêche le pathogène de se développer. Ainsi les NLR en l'absence de pathogène sont maintenus dans un état inactif pour éviter une activation inappropriée de la mort cellulaire. Il semble évident que de multiples interactions intramoléculaires entre les différents domaines des protéines NLR maintiennent l'assemblage global dans un état inactif. Ce travail de thèse s'est concentré sur un domaine de reconnaissance, caractérisé comme un domaine de leurre intégré, chez RGA5. Des études récentes montrent que les NLR possédant un domaine de leurre intégré sont découvertes de plus en plus fréquemment chez le riz mais également au sein d'autres espèces comme le blé (Kroj et al., 2016). Grâce aux données obtenues du côté des facteurs de virulence de *M. oryzae*, AVR-Pia et son mutant H3, et sur la protéine RGA5 de riz, nous avons pu mieux caractériser cette interaction. Cependant ces résultats soulèvent d'autres questions. D'une part au niveau structural, il serait intéressant de connaître la nature et d'avoir ainsi une description à l'échelle atomique du complexe entre RATX1 et AVR-Pia, ainsi que celle du complexe entre RATX1 et AVR1-CO39. Bien que structurellement apparentées, les protéines AVR-Pia et AVR1-CO39 possèdent des variations importantes de leurs surfaces. Comment un seul domaine RATX1 peut interagir avec deux AVR aussi différents ? L'étude de ces complexes permettrait de mieux comprendre l'interaction entre ces différents partenaires. Notamment le même type d'analyse que ce que j'ai réalisé pour l'AVR-Pia pourrait être réalisé dans le cas d'AVR1-CO39 pour répondre à la question de savoir s'il existe une similitude de la surface reconnue par le RATX1. Ainsi, ces données compléteraient les connaissances concernant le mode de reconnaissance de l'effecteur par les protéines de résistance.

L'étude de l'interaction entre ces domaines est un premier pas vers une meilleure compréhension de la reconnaissance entre les effecteurs et les NLR. Il est important de resituer la mécanique du système dans son ensemble, et de s'intéresser aux NLR RGA4/RGA5 dans leur intégralité. Pour l'instant, la stoechiométrie du système en veille n'est pas connue, la paire de NLR pourrait constituer un hétérodimère (RGA4/RGA5) ou bien par un tétramère (2xRGA4/2xRGA5). Lors de la reconnaissance, la détection de l'effecteur pourrait engendrer des changements conformationnels modifiant ainsi le complexe formé par les NLR. Ainsi RGA4 est activé (sous forme de monomère ou

de dimère) et peut induire la réponse hypersensible. L'interaction de RGA5 en complexe avec l'effecteur, empêcherait la reformation de l'état de veille. A l'heure actuelle, ceci n'est que spéculation, d'où la nécessité d'approfondir les connaissances nous permettant de connaître les différentes étapes de ce mécanisme. Dans un premier temps, il nous faudrait déterminer comment le domaine RATX1 est intégré dans le récepteur RGA5, puis quels sont les changements conformationnels induits au niveau de RGA5 lors de la reconnaissance de l'effecteur et enfin comment cette suite d'évènements conduit à la dé-répression du RGA4.

Pour le premier point soulevé, l'identification des autres domaines de RGA5 impliqués dans la reconnaissance de l'effecteur nous permettrait de mieux comprendre son fonctionnement. En effet RGA5 Δ RATX1 est capable de réprimer l'action de RGA4, par contre le système devient insensible à la présence de l'effecteur. Cependant des effecteurs possédant des résidus mutés au niveau de la surface d'interaction avec RATX1, et qui ont une liaison réduite, voir abolie avec RATX1 (en Y2H et co-IP par exemple) peuvent toujours interagir avec RGA5 Δ RATX1 (en co-IP) sauf dans le cas du mutant Avr-Pia F24S. Nous avons donc émis l'hypothèse qu'il existe au moins un autre site d'interaction avec l'AVR. Ainsi pour identifier ce ou ces autres sites d'interaction, une approche basée sur des tests biologiques pourrait être utilisée (double hybride ou co-immuno-précipitation) pour identifier ces zones. Il faudrait ensuite produire et purifier les protéines recombinantes correspondantes à ces zones d'intérêt permettant de déterminer une surface additionnelle d'interaction avec l'effecteur. L'un des problèmes auquel nous pourrions être confrontés est l'obtention de ces protéines. De plus l'étude de l'interaction ne sera possible uniquement si ce nouveau domaine correspond à un domaine de liaison autonome, c'est à dire que les autres parties de la protéine RGA5 ne soient pas indispensables à l'interaction.

Dans un deuxième temps il serait intéressant d'étudier les changements conformationnels de RGA5 lors de son interaction avec l'effecteur. Cependant, cette analyse nécessite de disposer du récepteur RGA5 entier. Cela peut être envisagé en utilisant un système d'expression dans des cellules d'insectes infectées par du baculovirus. L'étude des changements de conformation pourrait être analysé par RMN ou par cristallographie. Compte tenu de la taille de la protéine RGA5, plus de 1000 résidus, une étude par RMN nécessiterait l'utilisation de protéine deutérée et des marquages sélectifs (méthyles ou acides aminés ciblés), ce qui constitue un réel challenge. L'étude par cristallographie paraît donc plus judicieuse. Cependant, il ne faut pas oublier qu'il doit exister une dynamique intrinsèque de ces domaines, et également une dynamique des zones de jonction, qui sont autant de facteurs qui posent problème dans une étude structurale.

Parmi les évènements qui conduisent à la dé-répression, nous avons démontré que le domaine RATX1 est nécessaire pour la dé-répression mais pas pour la répression. Outre l'interaction de l'effecteur avec différents domaines de RGA5, il semble que la variation dans l'état d'oligomérisation des récepteurs RGA4 et RGA5 par le domaine « coiled-coil » (CC) positionné à l'extrémité N-terminale joue un rôle prépondérant dans l'activation de la signalisation conduisant à la réponse hypersensible (Kroj et al., 2016). A l'heure actuelle, des données obtenues sur les domaines « coiled-

coil » indiquent que les protéines RGA4 et RGA5 forment des homo- et des hétéro-complexes dans la plante (Cesari et al., 2014b). Ainsi même si le mécanisme de régulation n'est pas connu, une des hypothèses pourrait être que dans le complexe inactif RGA4-RGA5, le domaine LRR de RGA4 soit en liaison avec le NB de RGA5. Les domaines CC de RGA4 et RGA5 ne participeraient que de façon secondaire. L'interaction entre le domaine RATX1 et l'effecteur provoquerait un changement de conformation dans lequel le NB de RGA5 s'auto-associerait, avec le domaine RATX1 toujours lié à l'effecteur et LRR, formant un homodimère, incluant aussi le CC de RGA5 (phase de dé-répression). Le RGA4 serait alors libre de s'auto-associer via son CC (phase d'activation) pour induire une signalisation conduisant à la réponse hypersensible. Une étude structurale de ces domaines « coiled-coil » en complexe homo ou hétéromérique donnerait des premières informations liées à l'oligomérisation et potentiellement à la régulation de ces protéines NLR.

L'étude de ces protéines de résistance de plante, en grande partie responsable des mécanismes de défense en réponse au pathogène, est donc d'une importance capitale tant au niveau moléculaire qu'à l'échelle du champ agricole en lui même. Depuis 2009, une directive européenne incite les états membres à prendre toutes les mesures nécessaires pour éviter l'utilisation de pesticides pour protéger les cultures. L'amélioration des connaissances sur les interactions hôtes-pathogènes permet de mieux comprendre le fonctionnement des récepteurs immunitaires des plantes. Ces informations sont cruciales pour accélérer les processus de sélection de nouvelles souches résistantes contre les principales maladies et ont un rôle décisif à jouer dans la protection durable des cultures. Les gènes de résistances codant des récepteurs immunitaires peuvent être sélectionnés pour produire des variétés « élites », capables de détecter des agents pathogènes émergents. L'utilisation de ces variétés de cultures résistantes représente la solution économique et écologique la plus avantageuse pour le contrôle phytosanitaire.

MATÉRIELS

&

MÉT'HODES

A. Production et purification des effecteurs recombinants

1. Matériel biologique

1.1 Souches et plasmides

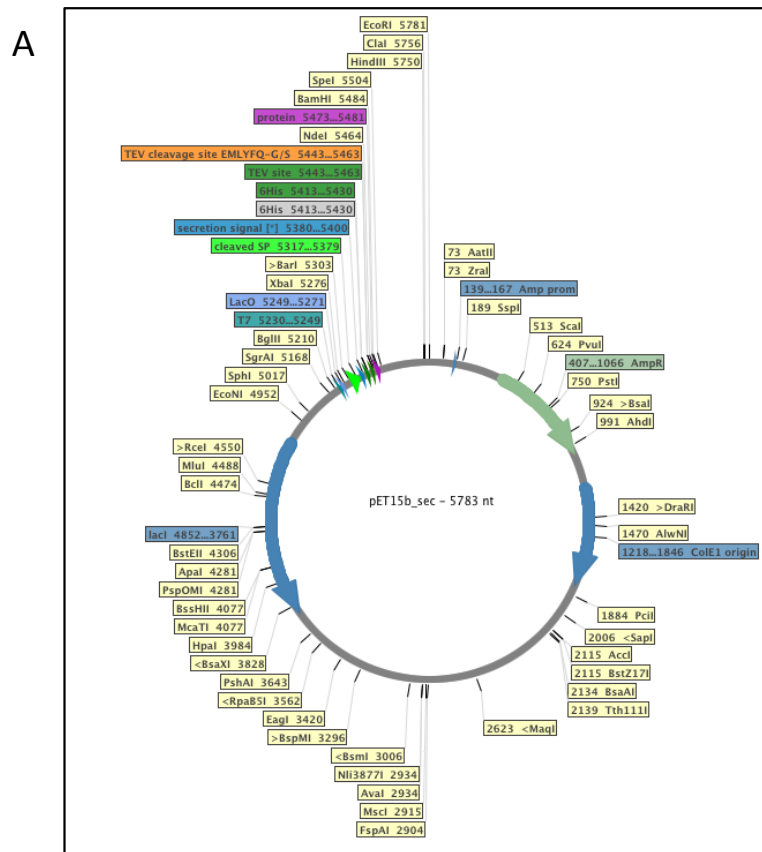
Les gènes des effecteurs AVR1-CO39, AVR-Pia (WT et mutants) de *Magnaporthe oryzae* sont clonés dans un plasmide d'expression dérivé de pET15b (Merck-Millipore) nommé pEt15b_sec (Fig M1). Ce dernier permet d'exprimer un signal de sécrétion au periplasme, une étiquette 6-histidines et un site de coupure à la TEV protéase en N-terminal. Le gène d'intérêt est cloné entre Nde1 et BamH1 sous le contrôle du promoteur T7, inductible à l'IPTG.

Des souches bactériennes compétentes BL21 (DE3) d'*Escherichia. coli* ont été transformées avec les différents plasmides.

Les bactéries sont rendues compétentes par la méthode du choc au calcium. Pour cela des bactéries BL21 (DE3) (invitrogen ®) sont mises en culture dans du milieu LB (20ml) pendant 3H à 37°C jusqu'à l'obtention d'une DO_{600nm} de 0.3-0.4 correspondant à la phase exponentielle de croissance. A partir de ce stade, les cellules sont maintenues à 4°C. Elles sont centrifugées 10 min à 5000 rpm à 4°C dans des tubes froids. Le surnageant est éliminé et le culot resuspendu dans 1/2 volume (10ml) de culture initial de solution froide de $CaCl_2$ 50 mM et incubé 20 min sur glace. Après une centrifugation de 10 min à 5000 rpm à 4°C ; les culots sont repris dans 1/20^{ème} de volume (1ml) de solution de $CaCl_2$ 50 mM. Des aliquotes de 200µl avec 15% de glycérol en final sont réalisés, et les cellules congelées à -80°C.

Les cellules BL21 (DE3) compétentes sont transformées avec le vecteur pEt15b_sec_AVR par choc thermique. Quelques microlitres de plasmide sont ajoutés à un aliquote de 200µl de cellules compétentes et incubés sur glace pendant 20 min. Un choc thermique à 42°C pendant 90 sec est réalisé. Les cellules sont ensuite replacées au froid pendant 2 à 3 min puis incubées dans 0,8ml de LB à 37°C pendant 1 heure. La culture est centrifugée 5 min à 5000 rpm, le surnageant jeté et le culot repris dans 20-30µl de LB puis étalé sur une boîte de pétri contenant l'antibiotique approprié et est ensuite placée toute la nuit dans l'étuve à 37°C.

Des mini-tests d'induction sont réalisés pour vérifier la production des protéines et pour garder en stocks glycérol les colonies les plus efficaces. Le protocole est décrit dans la figure M2.



B

Protéine	Vecteur	Souche/ AB	PM (Da)	pl	ϵ ($M^{-1}.cm^{-1}$)
AVR1-CO39	pET15b_sec_AVR1-CO39	BI21(DE3) <i>Ampicilline ou carbenicilline</i>	11228,1	5,12	14440
<p><u>MKKTAIAIAVALAGFATVAQA</u> / <u>APQDNTSMGSSHHHHHHSSGRENLYFQ</u> / <u>GHMAWKDCIIQRYKGDVNNIYTANRNEEITIEEYKVFVNEACHPYPVILPDRSVLSGDFTSAYADDDDESC</u></p>					
AVR PIA	pET15_sec_AVRPia	BI21(DE3) <i>Ampicilline ou carbenicilline</i>	10860,1	6,82	12950
<p><u>MKKTAIAIAVALAGFATVAQA</u> / <u>APQDNTSMGSSHHHHHHSSGRENLYFQ</u> / <u>GHMAAPARFCVYYDGHLPATRVLLMYVRIGTTATITARGHEFEVEAKDQNKVILTNGKQAPDWLAAEPY</u></p>					
<p>AVR PIA H3 (double mutant F24S T46N)</p>					
<p><u>MKKTAIAIAVALAGFATVAQA</u> / <u>APQDNTSMGSSHHHHHHSSGRENLYFQ</u> / <u>GHMAAPARSCVYYDGHLPATRVLLMYVRIGNTATITARGHEFEVEAKDQNKVILTNGKQAPDWLAAEPY</u></p>					

Fig M1. Plasmide et séquences des effecteurs recombinants :

A : Plasmide pET15b_sec

B: Séquence et caractéristiques des protéines effecteurs clonées. La séquence de sécrétion au périplasme clivée et non-clivée est respectivement soulignée en vert et en bleu; l'étiquette 6-histidines est soulignée en noir; le site de clivage à la protéase TEV est souligné en orange. La séquence de la protéine est en gris. Pour AVR-PIA les simples mutants suivant sont réalisés: R23A, F24S, D29A, R36A, M40A, Y41A, R43A, T46N, E56A, E58A, D63A, K67A, K74A, D78A, Y85A. Le poids moléculaire, le point isoélectrique et le coefficient d'extinction molaire sont déterminés par l'outil ProtParam d'ExPaSy.

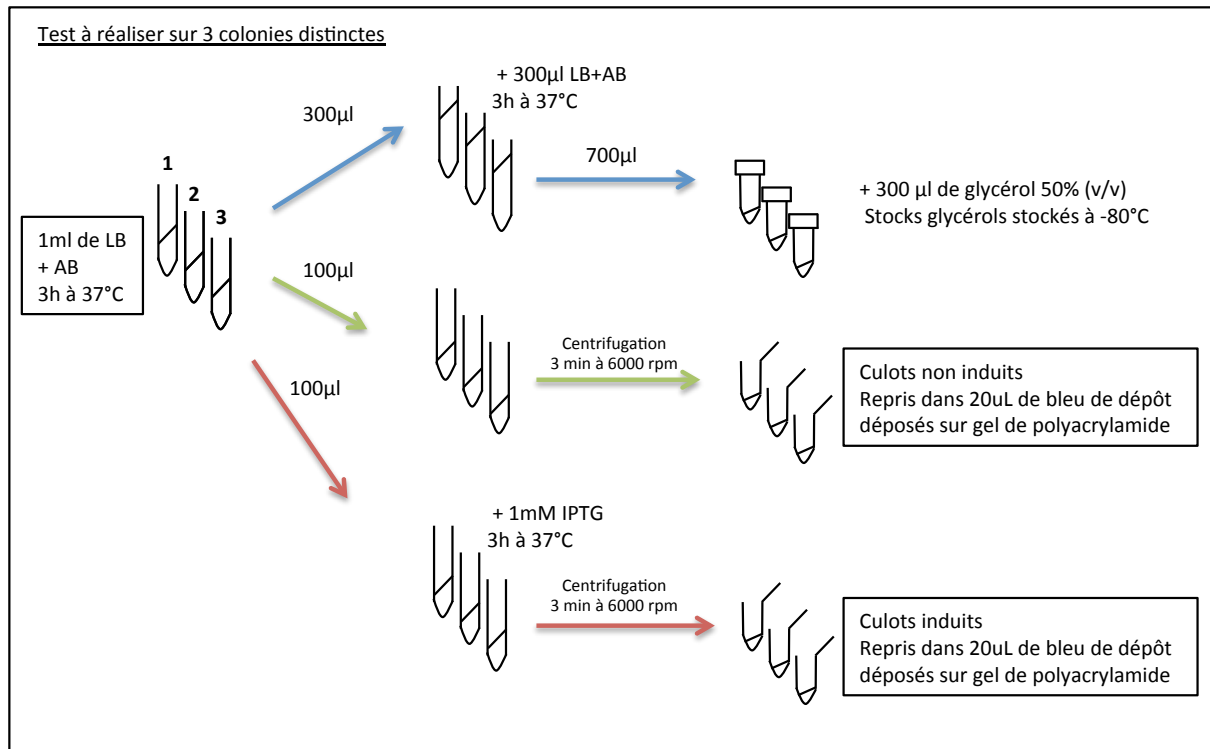


Fig M2. Protocole pour mini tests d'induction et stocks glycérols

2. Techniques de Biochimie

Les différentes techniques de biochimie décrites ci-après sont réalisées pour AVR1-CO39 et AVR-Pia (WT et Mutants) de façon semblable.

2.1 Production

Une pré-culture en milieu riche LB est réalisée à partir du stock glycérol des bactéries transformées, suivi d'une pré-culture en M9. Le milieu est complété avec de la carbenicilline (25µg/ml) de manière à maintenir le plasmide pEt15b_sec_AVR dans la bactérie. Une culture d' 1l de milieu C750501 auto-induit par le lactose (Studier, 2005) est lancée à partir de la pré-culture en M9. La culture est réalisée à 37°C sous agitation à 220 rpm pendant 7h puis passée à 30 °C toute une nuit. Lorsque la DO_{600nm} mesurée est supérieure à 6, la culture est centrifugée à 6500 rpm pendant 20min, et les culots gardés à froid (4°) ou congelés à -40°C pour une purification ultérieure. Les milieux M9 et C750501 (table M1) utilisés pour la culture sont soit non-marqués soit simplement marqués ^{15}N soit doublement marqués $^{15}N/^{13}C$. Le $^{13}C_6$ -Glucose, le $^{13}C_3$ -Glycérol et le $^{15}NH_4Cl$ sont respectivement les seules sources de carbone et d'azote.

M9	
Pour 500mL de milieu (qsp 1l H₂O)	
Na ₂ HPO ₄	6 g
KH ₂ PO ₄	3 g
NaCl	0.5g
Autoclaver 15min puis stocker à RT	
Ajouter stérilement pour 100 ml	
MgSO ₄ 1M	200µl
Ca ₂ Cl ₂ 1M	10µl
NH ₄ Cl 20% (m/v) ¹⁴ N ou ¹⁵ N	660µl
Glucose 50% (m/v) ¹² C ou ¹³ C	600µl
Thiamine 2.5% (m/v)	100µl
Biotine 0.2% (m/v)	500µl
C750501	
Pour 500mL de milieu (qsp 500ml H₂O)	
Na ₂ HPO ₄ 50mM (MW :178) :	4.45 g
KH ₂ PO ₄ 50mM (MW :136) :	3.48 g
Autoclaver 15min puis stocker à RT	
Ajouter stérilement:	
Na ₂ SO ₄ 0.5M	5ml
MgSO ₄ 1M	1ml
NH ₄ Cl 20% (m/v) ¹⁴ N ou ¹⁵ N	6.5ml
Lactose 2% (m/v)	2.5ml
Glycerol 50% (m/v) ¹² C ou ¹³ C	7.5ml
Glucose 5% (m/v) ¹² C ou ¹³ C	5ml
Metal mix	100µl
Mélange de métaux (Metal Mix)	
Solutions stocks Ss à 0.1M qsp 10ml H₂O filtrés 0,22µm	
FeCl ₃ :	m = 162.00 mg ds 0.12 M HCl
MnCl ₂ :	m = 197.91 mg
ZnSO ₄ :	m = 287.54 mg
CoCl ₂ :	m = 129.84 mg
CuCl ₂ :	m = 134.45 mg
NiCl ₂ :	m = 237.71 mg
Na ₂ MoO ₂ :	m = 241.95 mg
Na ₂ SeO ₃ :	m = 172.94 mg
H ₃ BO ₃ :	m = 61.84 mg
Solution stock à 1M qsp 10ml H₂O	
CaCl ₂ :	m = 1.47 g
Mix 1X qsp 5ml H₂O filtrés 0,22µm	
FeCl ₃	50mM Ss : 2.5ml
MnCl ₂	10mM Ss : 500µl
ZnSO ₄	10mM Ss : 500µl
CoCl ₂	2mM Ss : 100µl
CuCl ₂	2mM Ss : 100µl
NiCl ₂	2mM Ss : 100µl
Na ₂ MoO ₂	2mM Ss : 100µl
Na ₂ SeO ₃	2mM Ss : 100µl
H ₃ BO ₃	2mM Ss : 100µl
CaCl ₂	20mM Ss : 100µl
200µl de mix 1X pour 1L de milieu	

Table M1. Milieux de culture

2.2 Extraction de la fraction periplasmique

Pour cette étape toutes les manipulations sont faites à froid, sur de la glace, les tampons étant eux même préalablement mis à 4°C. Les culots cellulaires sont repris dans 120ml de TS-Buffer (200mM TrisHCl pH8, 500mM sucrose) auquel sont ajouté 40ml d'EDTA à 5mM pH8, puis 40ml de lysozyme à 0.1 mg/ml dans 50mM de Tris-HCl pH8. Une fois le mélange homogène, 200ml de TE-Buffer (200mM Tris-HCl pH8 ; 0.5mM EDTA) sont ajoutés et homogénéisés doucement. Le mélange est incubé sans agitation sur glace pendant 30 min. Après centrifugation pendant 30 min à 8000 rpm le surnageant est filtré à 0.45µm et gardé sur glace avant purification.

2.3 Purification des protéines

2.3.1 *Chromatographie par affinité*

L'extrait periplasmique est purifié par un système FPLC AKTA sur une colonne His-trap 5ml (GE Healthcare) équilibré dans du Tampon A (50 mM TrisHCl, pH 8.0, 300 mM NaCl, 1 mM DTT, 0.1 mM Benzamidine) . L'échantillon auquel est ajouté 2% de Tampon B (Tampon A + 500 mM imidazole) est injecté à 5ml/min par une pompe de chargement. La colonne est ensuite lavée en Tampon A + 2% B pendant 20 volumes de colonne. L'élution est effectuée avec un gradient d'imidazole (20 volumes de colonne pour arriver à 100% de Tampon B). La sortie de la protéine est suivie par absorption UV-vis à 260 et 280 nm. Les fractions contenant la protéine sont identifiées par SDS-PAGE (Cf section 3.1) et rassemblées.

2.3.2 *Chromatographie par filtration sur gel*

Les fractions rassemblées correspondant à un volume inférieur à 10ml seront directement injectées sur une colonne de gel filtration de type Superdex S75 26/60 (GE Healthcare) préalablement équilibrée en tampon A. Cette colonne sépare les protéines en fonction de leur taille. L'élution est réalisée à un débit de 2,5ml/min et des fractions de 2,5 ml sont collectées. Le profil d'élution indique un pic et les fractions contenant la protéine sont rassemblées, concentrées par filtration sur Centricon® jusqu'à environ 1mM et les échantillons sont utilisés tels quels ou congelés à -20°C. Le dosage des protéines est réalisé par Nanodrop 2000. (Cf section 3.2)

2.3.3 *Coupure à la protéase TEV*

Les fractions correspondant à l'éluion de la protéine sont incubées avec la protéase TEV (à 1mg/ml) toute la nuit à 20°C. Après centrifugation (15min à 5500rpm) l'échantillon est purifié par filtration sur gel comme décrit dans la section 2.3.2. Les fractions sont contrôlées sur gel SDS-PAGE et regroupées.

Etant donné le faible rendement obtenu lors des différents essais de coupure (environ 30%), il a été décidé de préparer les protéines sans clivage à la TEV.

3. Analyses Biochimiques

3.1 Electrophorèse par gels SDS-PAGE

Tampon d'électrophorèse : Tris-Glycine 5X (1litre)		
Tris-HCl	15,1g	
Glycine	72g	
SDS 10%	50 ml	
	H ₂ O qsp	
Bleu de dépôt 2X pour 15 ml		
SDS 20%	3 ml	
Glycérol 50%	6 ml	
DTT 1M	3 ml	
Tris 1M pH 6,8	1,5 ml	
Bleu de Bromophénol	0,03g	
	H ₂ O qsp	
Solution de Coloration		
Ethanol 40% (v/v)		
Acide Acétique 10% (v/v)		
Bleu de coomassie R 0,275% (m/v)		
Solution de Décoloration		
Ethanol 30% (v/v)		
Acide Acétique 10% (v/v)		
Pour 4 gels de 0,75 mm	Gel de séparation (17%)	Gel de concentration (4%)
H ₂ O	2,6 ml	6,15 ml
1,5M Tris-HCl pH 8,8	3,75 ml	/
0,5M Tris-HCl pH 6,8	/	2,5 ml
20% (m/v) SDS	75 µl	50 µl
Acrylamide/Bis-acrylamide 30% (m/v)	8,5 ml	1,34 ml
Ammonium persulfate (APS)	40 µl	25 µl
TEMED	20 µl	20 µl
<i>Le TEMED sera ajouté juste avant de couler les gels car ils vont polymériser assez vite.</i>		
<i>(ref. Laemmli U. K 1970)</i>		

Table M2. Protocole de préparation des solutions et des gels de polyacrylamide pour l'électrophorèse

L'électrophorèse sur gel polyacrylamide en conditions dénaturantes, sépare les protéines en fonction de leur masse moléculaire. Elle permet de contrôler, la présence de la protéine lors de test d'induction, la masse moléculaire et d'estimer la pureté de l'échantillon. Compte tenu du faible poids moléculaires des protéines étudiées, des gels à 17% d'acrylamide sont nécessaires pour avoir une bonne séparation des bandes.

Le protocole pour la préparation des solutions et des gels est détaillé dans la table M2.

Les échantillons sont dilués une fois dans du bleu de dépôt, chauffés 5 min à 95°C, centrifugés 1 min puis déposés sur le gel. La migration est effectuée sous une tension de 180V, à 30 mA pendant 60min dans le tampon de migration.

Après migration, le gel est coloré pendant une heure dans une solution de coloration. Puis rincé à l'eau et décoloré après incubation dans une solution de décoloration pendant 30min. Pour une détection des petits poids moléculaires le marqueur BenchMark™ est utilisé (marqueur de poids moléculaires de 180kDa à 6kDa).

3.2 Dosage standard

Les protéines sont dosées par la mesure d'absorbance à 280 nm avec un Nanodrop 2000.

D'après la loi de Beer-Lambert : $A = \epsilon \cdot l \cdot C$

A correspond à l'absorbance mesurée à 280nm, ϵ le coefficient d'extinction molaire en $M^{-1}.cm^{-1}$ caractéristique de la protéine par sa séquence d'acides aminés, l le trajet optique en cm^{-1} et C la concentration en Moles.

Les coefficients d'extinction molaires des protéines sont obtenus par l'outil ProtParam du site ExpASy.

3.3 Détermination des ponts dis-sulfures

Le dosage du réactif d'Ellman ou DTNB (5,5'-dithiobis-(2-nitrobenzoïque acide)) permet la quantification des groupements thiol libres à partir de courbe standard. Celles ci sont établies par la mesure d'absorbance à 412 nm d'une solution de cystéine (Sigma®), à différentes concentrations (de 12.5 μ M à 75 μ M) en condition native (Tampon 1 : Phosphate de sodium pH8 100mM, EDTA 0.1mM)

et dénaturante (Tampon 2 : Phosphate de sodium pH8 100mM, 6M de Chlorure de Guanidinium) après réaction avec le DTNB.

Le protocole est décrit ci dessous :

A partir d'une solution initiale de DTNB à 10mM (Tampon 1) deux solutions de DTNB à 0.1mM sont réalisées dans le Tampon 1 (Sol. DTNB_1) et dans le Tampon 2 (Sol. DTNB_2).

Les solutions de cystéines sont réalisées selon le tableau ci dessous:

Standard	#1	#2	#3	#4	#5	#6	#7 (REF)
Tampon (1 ou 2)	1 ml	50 µl	100 µl	150 µl	200 µl	250 µl	300 µl
Cystéine	2.634mg	250 µl de standard #1	200 µl de standard #1	150 µl de standard #1	100 µl de standard #1	50 µl de standard #1	-
Concentration en cystéine	1.5 mM	1.25mM	1.0mM	0.75mM	0.5mM	0.25mM	0mM

Pour réaliser la mesure, 5 µl de Standard sont dilués dans 95 µl de réactif (Sol. DTNB_1 ou Sol. DTNB_2), incubés 10min à 25°C. L'absorbance est mesuré à 412nm sur un Nanodrop 2000.

Le calcul de la concentration de thiol libre est obtenu par la loi de Beer Lambert (Cf 3.2).

$$\text{Soit : } C (\text{SH libre}) = \frac{A}{\epsilon \cdot l} \cdot d$$

où d est le facteur de dilution ici égal à 20.

ϵ le coefficient d'extinction molaire de la cystéine en condition native (14150 M⁻¹.cm⁻¹)
ou en condition dénaturante (13700 M⁻¹.cm⁻¹).

B. Production et purification des protéines recombinantes R dérivées de RGA-5

1. Matériel biologique

1.1 Souches et plasmides

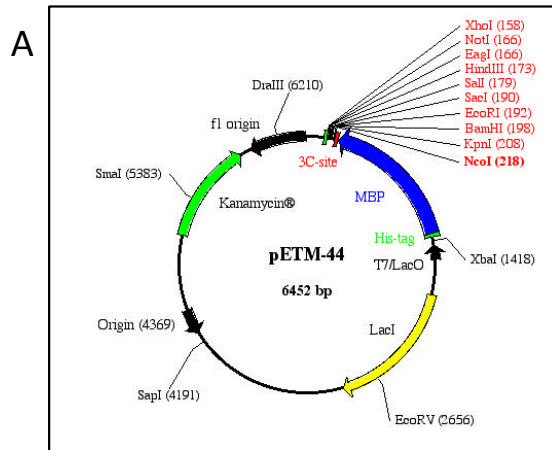
Les différents découpages de la protéine RGA-5 sont basés sur les études d'interaction in vivo réalisés par nos collaborateurs du BGPI (Cesari et al., 2013) Quatre constructions sont réalisées : RGA5 _long, RGA5_short, RGA5_deletion Cter et MBP-RGA5. La dernière construction correspondant à un contrôle négatif car elle est constituée d'une partie de séquence du plasmide de départ et d'un fragment de RGA5. Les séquences sont insérées dans un plasmide pETM-44 permettant d'exprimer une étiquette 6-histidines, la Maltose Binding Protein (MBP) et un site de clivage à la protéase 3C en N-terminal du gène codant pour la protéine d'intérêt. Ce plasmide est sous le contrôle du promoteur T7, inductible à l'IPTG.

Les plasmides sont transformés dans des bactéries compétentes BL21 (DE3) (Cf section A 1.1) Des mini test d'induction pour vérifier la production des protéines sont réalisés comme précédemment décrit.

2. Techniques de Biochimie

2.1 Production

A partir de 10ul de stock glycérol une pré-culture de 20 ml de LB est lancée toute la nuit à 37°C sous agitation (220 rpm). Le lendemain matin une culture de 700ml de LB estensemencée avec la pré-culture pour repartir avec une DO_{600nm} de 0,15. La culture est laissée à 37°C et 220 rpm jusqu'à ce que la DO_{600nm} atteigne 0,6, la température est alors abaissée à 30°C et la culture induite avec 0,2mM d'IPTG pendant 4h. Les cellules sont centrifugées à 6500 rpm pendant 20min, et les culots gardés à froid ou congelés à -40°C pour une purification ultérieure.



B

Protéine	Vecteur	Souche/ AB	PM (kDa)	pl	ϵ ($M^{-1}.cm^{-1}$)
RGA5c_long	pTK163.3 (pETM44)	BL21(DE3) <i>Kanamicine</i>	69,838	5,57	85260
<p>MK<u>HHHHHH</u>PMKIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVHEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPKAFQDKL YPFTWDAVRYNGKLIAYPIAVEALSILYKNDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGYDIKDVGVNDAGAK AGLTFLVDLIKKNHMNADTDYSIAEAFNKGEMTAMTINGPWAWNSNIDTSKVNNGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENLLTDEG LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT<u>PGSLEVL</u>FQ / GPMEYYPYDVPDYAHMEHVLSLQDIATRIGVDTSTGESDRRAEASFKEVTNKHPRCLRSSLQWVWVSTEEESHPLEKQHHKREKGGSSAGHGVLKESV EDSEKNTDRVQTLSPQLSNMESVVESALTGQRTKIVVKVPCGKSRKAMALAASVNGVNSVEITGEDKQDLVVVCGGIDAVRLAALLRKKYGLAEILV VESVKKKTKLAGEKASVPDYNKHPTYNAPELDYVEYYPLQMTEDPKGCSIM</p>					
RGA5c_short	pTK166.1(pETM44)	BL21(DE3) <i>Kanamicine</i>	58,769	5,47	79760
<p>MK<u>HHHHHH</u>PMKIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVHEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPKAFQDKL YPFTWDAVRYNGKLIAYPIAVEALSILYKNDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGYDIKDVGVNDAGAK AGLTFLVDLIKKNHMNADTDYSIAEAFNKGEMTAMTINGPWAWNSNIDTSKVNNGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENLLTDEG LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT<u>PGSLEVL</u>FQ/ GPMEYYPYDVPDYAHMLSNMESVVESALTGQRTKIVVKVPCGKSRKAMALAASVNGVNSVEITGEDKQDLVVVCGGIDAVRLAALLRKKYGLAEILV SVKKEKTKLAGEKASVPDYNKHPTYNAPELDYVEYYPLQMTEDPKGCSIM</p>					
RGA5c deletion C-ter	pTK167.3(pETM44)	BL21(DE3) <i>Kanamicine</i>	59,511	5,73	76320
<p>MK<u>HHHHHH</u>PMKIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVHEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPKAFQDKL YPFTWDAVRYNGKLIAYPIAVEALSILYKNDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGYDIKDVGVNDAGAK AGLTFLVDLIKKNHMNADTDYSIAEAFNKGEMTAMTINGPWAWNSNIDTSKVNNGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENLLTDEG LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT<u>PGSLEVL</u>FQ/ GPMEYYPYDVPDYAHMEHVLSLQDIATRIGVDTSTGESDRRAEASFKEVTNKHPRCLRSSLQWVWVSTEEESHPLEKQHHKREKGGSSAGHGVLKESV EDSEKNTDRVQTLSPQLSNMESVVESALTGQRTKIVVKVPCGKSRKAMALAASVNG</p>					
MBP-RGA5c	pTK168(pETM44)	BL21(DE3) <i>Kanamicine</i>	50,838	5,49	81820
<p>MK<u>HHHHHH</u>PMKIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVHEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPKAFQDKL YPFTWDAVRYNGKLIAYPIAVEALSILYKNDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGYDIKDVGVNDAGAK AGLTFLVDLIKKNHMNADTDYSIAEAFNKGEMTAMTINGPWAWNSNIDTSKVNNGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENLLTDEG LEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQT<u>PGSLEVL</u>FQ/ GPMEYYPYDVPDYAHMAMEASEFHPSSGINAEWPLWPLSNMESVVESALTGQRTKIVVKVPCGKSRKAMALAASVNG</p>					

Fig M3. Plasmide et séquences des différents construits de RGA5 recombinants :

A : Vecteur pETM-44

B: Séquence et caractéristiques des protéines RGA5 clonées. La séquence codante pour la MBP est en bleu; l'étiquette 6-histidines est souligné en noir; le site de clivage à la protéase 3C est souligné en orange, le tag-HA est souligné et en gras. La séquence de la protéine est en gris. Pour la protéine MBP-RGA5 la partie en italique provient du plasmide de départ, la partie grise est un fragment de RGA5. Le poids moléculaire, le point isoélectrique et le coefficient d'extinction molaire sont déterminés par l'outil ProtParam d'ExPaSy

2.2 Extraction de la fraction soluble

Le culot est repris dans 50ml (pour 700ml de culture initiale) de tampon de casse (Tris pH8 50mM, NaCl 150mM) puis incubé sur glace pendant 30min avec 100µg/ml de lysozyme et 100µg/ml de DNase. La suspension bactérienne est lysée par un sonicateur à ultrasons (Bioblock Scientific VibraCell™) à une amplitude de 40 Watt, avec des impulsions de 3 secondes toutes les 3 secondes pendant 3min à 4°C) puis centrifugée pendant 20min à 18000 rpm à 6°C pour éliminer les débris cellulaires. De la benzamidine à 1mM en concentration finale est ajoutée au surnageant pour éviter l'action des protéases. Le surnageant est filtré sur 0,45µm puis gardé sur glace avant purification.

2.3 Purification des protéines

2.3.1 *Chromatographie d'affinité*

L'extrait de protéines solubles, supplémenté de 2% de tampon B, est purifié par FPLC (AKTA) sur une colonne His-trap 1ml (GE Healthcare) équilibrée dans du Tampon A (50 mM TrisHCl, pH 8.0, 300 mM NaCl, 1 mM DTT, 0.1 mM Benzamidine) . L'échantillon est injecté à 1ml/min par une pompe de chargement. La colonne est ensuite lavée en Tampon A + 2% B (A + 500mM Imidazole) pendant 20 volumes de colonne. L'élution est effectuée avec un gradient d'imidazole (20 volumes de colonne pour arriver à 100% de Tampon B). La sortie de la protéine est suivie par absorption UV-vis à 260 et 280 nm. Les fractions contenant la protéine sont identifiées par SDS-PAGE (Cf sectionA 3.1) et rassemblées.

La protéine est dialysée avec un boudin de dialyse snakeskin™ (cut-off 3,5 kDa, 3,7ml/cm) plongée dans du tampon A (2 x 1h dans 1l à 4°C sous agitation puis toute la nuit dans 2l à 4°C). L'échantillon est ensuite dosé par Nanodrop 2000, aliquoté, et utilisé tel quel ou congelé à -20°C.

2.3.2 *Chromatographie par filtration sur gel*

Pour augmenter la pureté une étape de purification par gel filtration est effectuée sur la protéine en sortie de colonne d'affinité. Une colonne de gel filtration S200 26/60 (GE Healthcare) est équilibrée en tampon A à 2,5ml/min sur la FPLC. Cette colonne sépare les protéines en fonction de leur taille. Les fractions de 2,5ml sont récupérées, contrôlées par SDS-PAGE (Cf section A 3.1) et rassemblées. les échantillons dosés et congelés à -20°C ou utilisés tels quels.

2.4 Coupure à la protéase 3C

La protéine purifiée est toujours liée à la MBP. Un site de clivage à la protéase 3C est situé entre la MBP et la protéine d'intérêt. La coupure est réalisée toute la nuit à 4°C sur environ 1ml de protéine entière dialysée (en Tampon A sans benzamidine) à environ 10mg/ml et 50µg de protéase 3C comprenant une étiquette GST (Glutathion-S-transférase).

L'échantillon est ensuite déposé sur une colonne MBP de 1ml suivie d'une colonne de GST de 1ml (GE Healthcare) préalablement équilibrées en Tampon A, à un débit de 1ml/min par FPLC. Des fractions de 1ml sont récupérées. La sortie de la protéine est suivie par absorption UV-vis à 260 et 280 nm. L'élution de la MBP est réalisée par un gradient sur 5 volumes de colonne de Tampon M1 (Tampon A + 10 mM de Maltose). L'élution de la GST est réalisée par l'ajout de 5ml de Tampon G1 (Tampon A + 10M de Glutathione réduite). Le pic observé après injection et lavage des colonnes en Tampon A contient la protéine clivée.

Malheureusement dans ces conditions toutes les protéines clivées se sont dégradées très rapidement. Les études suivantes sont donc réalisées avec les protéines entières.

3. Étude d'interaction par Anisotropie de Fluorescence

3.1 Anisotropie de fluorescence

L'utilisation d'une lumière polarisée permet l'excitation d'un fluorophore. L'émission de fluorescence ou intensité (I) est mesurée selon deux angles de polarisation (parallèle \parallel et perpendiculaire \perp par rapport à la lumière d'excitation).

L'effet d'anisotropie de fluorescence A est définie par :

$$A = \frac{(I \parallel) - (I \perp)}{(I \parallel) + 2(I \perp)}$$

Dans notre cas l'anisotropie de fluorescence va être utilisée pour mesurer une constante d'affinité par un changement de temps de rotation de la molécule. Le fluorophore est lié à une petite protéine, sa vitesse de rotation va diminuer lors de son interaction avec une grosse protéine. La mesure correspondant à une différence de polarisation entre les états liés et non liés. L'affinité est calculée en titrant les deux partenaires de liaison l'un avec l'autre.

3.1.1 Marquage à l'Alexa 488 Succinyl-Ester

Dans notre cas, les protéines effecteurs (Avr1_CO39, Avr-Pia, Avr-Pia-H3) sont marquées par la sonde fluorescente.

Le marquage est réalisé avec le fluorophore Alexa 488 Succinyl-Ester. Le fluorophore va réagir avec l'amine libre de la partie N-terminale de la protéine. Le pH du tampon doit être supérieur à 7 pour éviter le marquage des lysines (NH_3^+) et favoriser le marquage du NH_2 terminal.

Protocole de marquage :

Un aliquote lyophilisé de fluorophore est resuspendu dans 4 μL de DMSO à une concentration finale de 50nM.

La protéine (25 μM dans 200 μL de tampon phosphate : NaP 50mM pH7,5 NaCl 200mM) est incubée avec la sonde pendant 3H à 25°C à l'abri de la lumière.

L'arrêt de la réaction est effectué par l'ajout de 5 μl de Tris 1M pH 7,5 pour saturer les amines libres.

La protéine ainsi marquée est récupérée après centrifugation sur une colonne de dessalage de 2ml (Zeba™ Desalt Spin Columns ; Thermo scientific)

La concentration de protéine et la concentration en Alexa 488 sont évaluées par dosage au Nanodrop 2000.

3.1.2 Tests d'interaction par anisotropie de fluorescence

Les protéines à titrer (dans notre cas les différents construits de RGA5) sont diluées dans une gamme de concentration allant de 100 μM à 0,01 μM .

Une plaque de titration Corning NBS de 384 puits est réalisée, en triplicate, avec la protéine marquée à une concentration de 5 μM finale.

Les mesures sont réalisées sur un appareil Zapire 2 Tecan. Pour l'Alexa 488 la longueur d'onde d'excitation est de 496nm, celle d'émission à 519 nm. Les intensités et l'anisotropie de fluorescence sont mesurées pour chaque point du titrage. Aucun facteur de correction n'est appliqué. Les courbes obtenues sont présentées dans le chapitre 2, 2.1 de cette thèse.

C. Production et purification du domaine RATX1_RGA5

1. Matériel biologique

1.1 Souches et plasmides

Le gène codant pour RGA5 _long possède des codons rares pour l'expression chez *E.coli* (rare codon Calculator RACC, NIH MBI Laboratory for Structural Genomics and Proteomics). Un gène synthétique (GeneArt) pour lequel les codons sont optimisés pour la production dans *E.coli* est inséré dans un plasmide de type pET15b (Merck-Millipore) possédant une étiquette 6-histidines et un site de coupure à la TEV protéase en N-terminal et contrôlé par le promoteur T7, inductible à l'IPTG (Fig M4.A.)

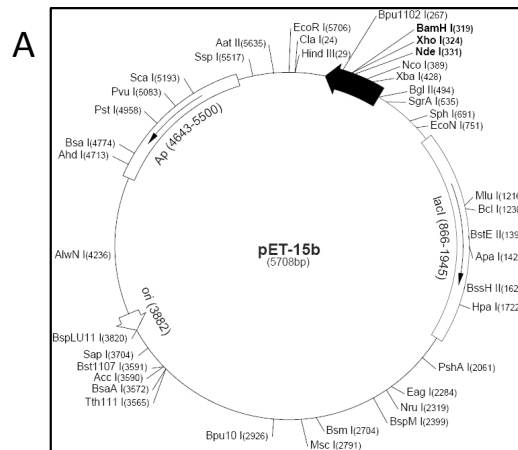
Les constructions correspondant aux protéines recombinantes de RGA5_new et du domaine Ratx1 sont détaillées dans la figure M4.B

Les plasmides sont transformés dans des bactéries compétentes BL21 (DE3) (Cf section A 1.1) Des mini test d'inductions pour vérifier la production est protéines sont réalisés comme précédemment décrit (Cf Fig M2).

2. Techniques de Biochimie

2.1 Production

Une pré-culture en milieu riche LB est réalisée à partir du stock glycérol des bactéries transformées, suivi d'une pré-culture en M9. Le milieu est complété avec de la carbenicilline (25µg/ml). Une culture d' 1l de milieu C750501 auto-induit est lancée à partir de la pré-culture en M9. La culture est réalisée à 37°C sous agitation à 220 rpm pendant 7h puis passée à 30 °C toute une nuit. Lorsque la DO_{600nm} est supérieure à 6, la culture est centrifugée à 6500 rpm pendant 20min, et les culots gardés à froid ou congelés à -40°C pour une purification ultérieure. La préparation des milieux de culture est réalisée selon le protocole précédemment décrit (Cf section A. 2.1 table1).



B

Protéine	Vecteur	Souche/ AB	PM (kDa)	pl	ϵ ($M^{-1}.cm^{-1}$)
domaine RATX1 (allèle CO39)	pCV158(pET15b)	BL21(DE3) <i>Ampicilline ou carbenicilline</i>	10,499	9,64	3105
<p>MGSS<u>HHHHHH</u>SSGRENLYFQ / <u>GH</u>MQRTKIVVKVPCGKSRAMALAASVNGVNSVEITGEDKDQLVVVGGIDAVRLAALLRKKYGLAEILVVESVKK</p>					

Fig M4. Plasmide et séquence du RATX1 recombinant :

A : Plasmide pET-15b

B: Séquence et caractéristiques des protéines RGA5 clonées. L'étiquette 6-histidines est soulignée en noir; le site de clivage à la protéase TEV est souligné en orange. La séquence de la protéine est en gris. Le poids moléculaire, le point isoélectrique et le coefficient d'extinction molaire sont déterminés par l'outil ProtParam d'ExPaSy

2.2 Extraction de la fraction soluble

La fraction soluble est extraite selon le protocole précédemment décrit dans la section B. 2.2.

Le culot est repris dans 50ml (pour 500ml de culture initiale) de tampon de casse (Tris pH7 50mM, NaCl 150mM). Après la lyse (chimique et physique) et centrifugation, le surnageant filtré est gardé sur glace avant purification.

2.3 Purification des protéines

2.3.1 Chromatographie par affinité

La purification par chromatographie d'affinité (colonne His-Trap 1ml) est réalisée comme décrit précédemment (Cf section B. 2.3.1). Le pH du Tampon A (50 mM TrisHCl, pH 7.0, 300 mM NaCl, 1 mM DTT, 0.1 mM Benzamidine) et B (A + 500mM Imidazole) utilisé est de 7. La protéine Rax1 ne possédant pas de tryptophane, l'élution est suivie par absorption UV-vis à 280 et 215 nm (longueur d'onde correspondant à l'absorbance des liaisons peptidiques). Toutes les fractions sont caractérisées par SDS_PAGE, celles contenant la protéine sont rassemblées.

2.3.2 *Chromatographie par filtration sur gel*

La purification par chromatographie par filtration sur gel (colonne S75 26/60) est réalisée comme décrit précédemment (Cf section B. 2.3.1) en Tampon A à pH7. Les fractions de 2,5ml, identifiées par absorption UV-vis à 280 et 215 nm sont récupérées, contrôlées par SDS-PAGE (Cf section A 3.1), rassemblées, l'échantillon est concentré par filtration à l'aide d'un Centricon® (cut-off: 2 kDa) et congelé à -20°C.

L'estimation de la concentration en protéine Rax1 est réalisé par dosage au BCA comme décrit ci dessous.

3. Analyses Biochimiques

3.1 Dosage BCA

Le dosage BCA est basé sur la détection par colorimétrie des cations cuivreux Cu^{1+} par l'acide bichichiniquique (BCA). Tout d'abord il faut procéder à une réaction de Biuret ou la protéine chélate le cuivre (Cu^{2+}) en milieu alcalin pour former un complexe bleu clair. Dans une seconde étape, l'acide bicinchiniquique réagit avec le cation cuivreux réduit conduisant à un changement de couleur vers le pourpre. Le complexe de BCA / cuivre est soluble dans l'eau et présente une forte absorbance à 562 nm qui est linéaire avec des concentrations croissantes de protéine. Une courbe standard est utilisée pour déterminer la concentration de chaque échantillon.

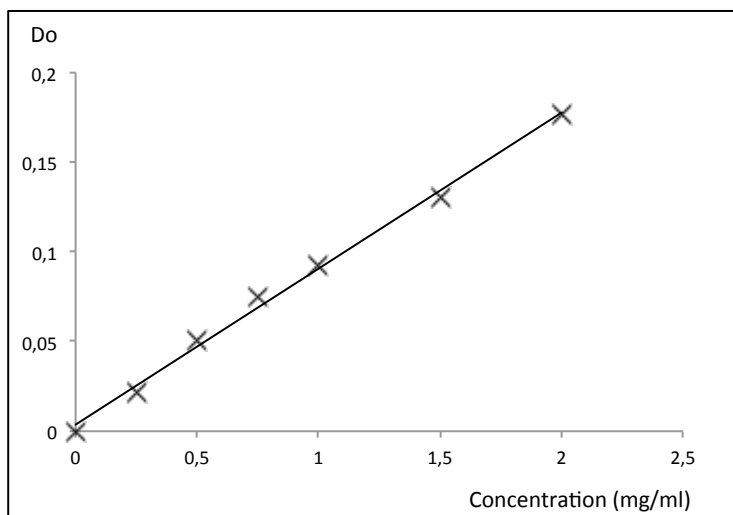


Fig M5. Courbe de calibration BCA permettant de déterminer la concentration de l'échantillon à partir de la valeur d'absorbance obtenue à 562nm.

Le kit Pierce™ Thermo SCIENTIFIC BCA protein Assay est utilisé pour réaliser ce dosage.

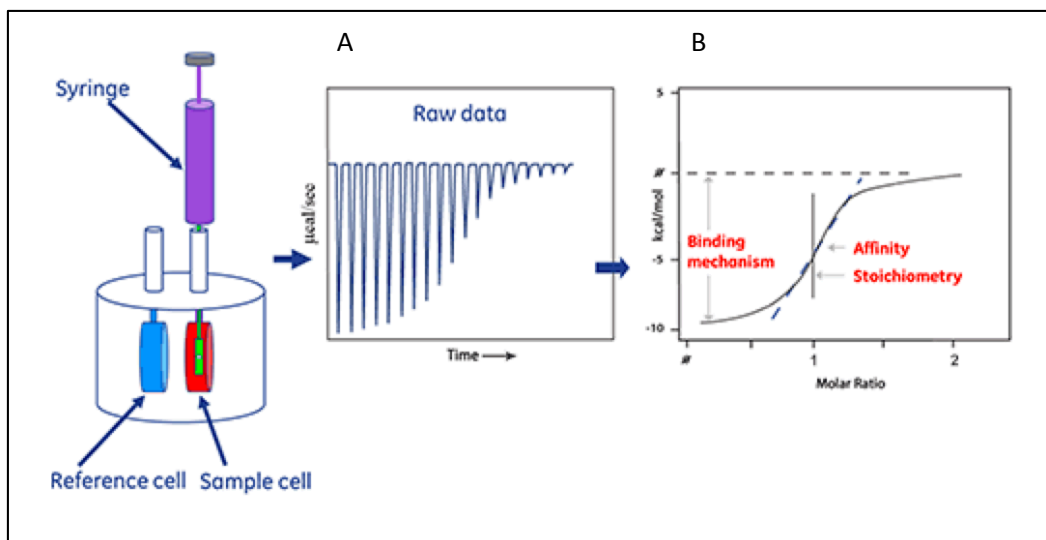
La courbe standard est réalisée avec de la BSA (Bovine Serum Albumine) de 5 à 250 µg/ml. La réaction se fait lors du mélange des échantillons avec la solution A (contenant du carbonate de sodium, bicarbonate de sodium, l'acide bicinchinique et du tartrate de sodium dans 0.1M d'hydroxyde de sodium) et la solution B (avec 4% de sulfate de cuivre) selon le protocole décrit dans le kit. Les échantillons sont incubés 30 min à 60°C puis refroidis à température ambiante. La mesure d'absorbance à 562nm est réalisée sur un Nanodrop 2000, et la concentration est déterminée par rapport à la courbe standard.

4. Étude d'interaction par Titrage Calorimétrique Isotherme

4.1 Titrage Calorimétrique Isotherme (ITC)

La titrage calorimétrique isotherme (ITC Isothermal titration calorimetry en anglais) est une technique de mesure de paramètres thermodynamiques de l'interaction entre deux partenaires en solution, en biologie généralement des macromolécules. C'est une mesure directe de chaleur générée ou absorbée lors de la liaison entre les molécules. Cette méthode permet une détermination précise de la constante d'association (K_A), de la stœchiométrie (n), de l'enthalpie (ΔH) de la réaction.

4.1.1 Principe de la mesure

**Fig M6. Schéma du calorimètre :**

A. Pics mesurés au cours des injections.

B. Courbe isotherme résultant de l'expérience.

(Ref. www.malvern.com)

L'utilisation d'un calorimètre permet de mesurer la variation de chaleur générée par la liaison d'un ligand et de son partenaire. Le microcalorimètre comporte deux cellules, l'une contient le tampon et sert de cellule de référence, l'autre contient l'échantillon (cuve centrale). Le microcalorimètre doit maintenir la cellule centrale à la température de l'expérience. La cuve latérale de référence est maintenue avec un écart de 1°C par rapport à la cuve centrale. Le ligand est chargé dans une seringue et un volume (volume d'injection) est injecté dans la cellule d'échantillon contenant la protéine étudiée. L'expérience de titrage consiste en une série d'injections d'une solution de ligand, réalisées dans la cuve centrale contenant la solution de protéine. Pour chaque injection, après une variation initiale de chaleur, un retour progressif à la valeur de base (valeur zéro) est observé. Pour chaque pic ainsi obtenu son intégrale représente la quantité de chaleur libérée ou absorbée lors de l'interaction des molécules et est proportionnelle à la variation d'enthalpie (ΔH) de la réaction pondérée par la fraction de saturation (Θ). Ainsi la chaleur totale (Q) contenue dans le volume utile de la réaction (V_0) peut être déterminée par la relation suivante :

$$Q = n \cdot \Theta \cdot M_t \cdot \Delta H \cdot V_0$$

Avec n : nombre de site Θ : fraction de saturation ou fraction de site occupé par le ligand M_t : concentration totale de protéine ΔH : enthalpie de réaction ou chaleur molaire de liaison du ligand

La constante d'affinité K_A est décrite par la relation suivante :

$$K_A = \frac{\Theta}{(1 - \Theta)[X]}$$

Avec $[X]$: concentration de ligand libre en solution

La concentration totale de ligand (X_t) est donnée par l'équation qui suit:

$$X_t = [X] + n\Theta M_t$$

Alors :

$$\Theta^2 - \Theta \left[1 + \frac{X_t}{nM_t} + \frac{1}{nK_A M_t} \right] + \frac{X_t}{nM_t} = 0$$

Ainsi, la résolution de l'équation du second degré pour Θ permet d'obtenir la valeur de Q selon la relation suivante :

$$Q = \frac{n \cdot M_t \cdot \Delta H \cdot V_0}{2} \left[1 + \frac{X_t}{nM_t} + \frac{1}{nK_A M_t} - \sqrt{\left(1 + \frac{X_t}{nM_t} + \frac{1}{nK_A M_t} \right)^2 - \frac{4X_t}{nM_t}} \right]$$

Donc, pour chaque injection, la valeur de Q mesurée peut être corrélée à une valeur de n , de K_A et de ΔH .

4.1.2 Préparation des échantillons

Les échantillons sont dialysés toute la nuit à 4°C dans le tampon d'ITC (NaP 20mM pH5,4, NaCl 150mM). Le tampon (20ml) et les échantillons sont dégazés avec une pompe à vide pendant 10min.

L'échantillon dans la cuve centrale est à une concentration de 80 μ M de RATX1, pour un volume de 1.78 mL. Pour remplir la cuve il est nécessaire de disposer de 2 mL. Le volume d'échantillon dans la seringue est quand à lui de 300 μ L à 1mM (500 μ L pour pouvoir remplir la seringue). Le volume de chaque injection est de 10 μ L (temps d'injection de 4 sec.), avec une injection initiale (injection de préparation) de 2 μ L et un délai entre les injections de 5 min avec une vitesse de rotation de la pale d'agitation de 300 rpm.

L'expérience réalisée à 25°C, a consisté à enregistrer 22 injections, à remplacer après lavage l'échantillon de la cuve centrale par du tampon seul, et à enregistrer une série de 10 injections pour pouvoir quantifier l'effet du tampon seul sur l'expérience d'ITC. Cette dernière série de 10 injections est utilisée pour calculer par régression linéaire l'effet du tampon lors du titrage, qui sera alors soustrait de l'expérience initiale lors de son analyse.

4.1.3 Mesure et interprétation

Dans notre cas la réaction est exothermique et les pics générés sont orientés vers le bas. Le rapport molaire entre le ligand et la protéine augmente progressivement lors des injections successives de ligand (Fig. M6). La protéine est progressivement saturée par le ligand, et la variation d'intensité des pics commence à décroître. La réaction atteint la saturation, lorsque la fraction des sites occupés (θ) ne varie plus.

Les données sont traitées avec le logiciel Origin®, V7.0. qui permet de manipuler les données expérimentales pour l'intégration et la correction de ligne de base. La courbe isotherme qui en résulte peut être ajustée par un modèle de liaison à un site, permettant de déduire l'affinité (K_A) et d'extraire le ΔH , et le nombre de site ($n=1$)

BIBLIOGRAPHIE

Abramovitch, R.B., Kim, Y.-J., Chen, S., Dickman, M.B., and Martin, G.B. (2003). *Pseudomonas* type III effector AvrPtoB induces plant disease susceptibility by inhibition of host programmed cell death. *EMBO J.* 22, 60–69.

Ade, J., DeYoung, B.J., Golstein, C., and Innes, R.W. (2007). Indirect activation of a plant nucleotide binding site–leucine-rich repeat protein by a bacterial protease. *Proc. Natl. Acad. Sci.* 104, 2531–2536.

Akamatsu, A., Wong, H.L., Fujiwara, M., Okuda, J., Nishide, K., Uno, K., Imai, K., Umemura, K., Kawasaki, T., Kawano, Y., et al. (2013). An OsCEBiP/OsCERK1-OsRacGEF1-OsRac1 Module Is an Essential Early Component of Chitin-Induced Rice Immunity. *Cell Host Microbe* 13, 465–476.

Ashikawa, I., Hayashi, N., Yamane, H., Kanamori, H., Wu, J., Matsumoto, T., Ono, K., and Yano, M. (2008). Two Adjacent Nucleotide-Binding Site-Leucine-Rich Repeat Class Genes Are Required to Confer Pikm-Specific Rice Blast Resistance. *Genetics* 180, 2267–2276.

Ausubel, F.M. (2005). Are innate immune signaling pathways in plants and animals conserved? *Nat. Immunol.* 6, 973–979.

Azizi, P., Rafii, M.Y., Abdullah, S.N.A., Nejat, N., Maziah, M., Hanafi, M.M., Latif, M.A., and Sahebi, M. (2014). Toward understanding of rice innate immunity against *Magnaporthe oryzae*. *Crit. Rev. Biotechnol.* 1–10.

Barthe, P., Ropars, V., and Roumestand, C. (2006). DYNAMOF: a program for the dynamics analysis of relaxation data obtained at multiple magnetic fields. *Comptes Rendus Chim.* 9, 503–513.

Belkhadir, Y., Nimchuk, Z., Hubert, D.A., Mackey, D., and Dangl, J.L. (2004). Arabidopsis RIN4 Negatively Regulates Disease Resistance Mediated by RPS2 and RPM1 Downstream or Independent of the NDR1 Signal Modulator and Is Not Required for the Virulence Functions of Bacterial Type III Effectors AvrRpt2 or AvrRpm1. *Plant Cell Online* 16, 2822–2835.

Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., Zhang, X., Ellis, J.G., Kobe, B., and Dodds, P.N. (2011a). Structural and Functional Analysis of

a Plant Resistance Protein TIR Domain Reveals Interfaces for Self-Association, Signaling, and Autoregulation. *Cell Host Microbe* 9, 200–211.

Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., Zhang, X., Ellis, J.G., Kobe, B., and Dodds, P.N. (2011b). Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell Host Microbe* 9, 200–211.

Bernoux, M., Burdett, H., Williams, S.J., Zhang, X., Chen, C., Newell, K., Lawrence, G., Kobe, B., Ellis, J.G., Anderson, P., et al. (2016). Comparative analysis of the flax immune receptors L6 and L7 suggests an equilibrium-based switch activation model. *Plant Cell* TPC2015-00303-RA.

Berruyer, R., Adreit, H., Milazzo, J., Gaillard, S., Berger, a., Dioh, W., Lebrun, M.H., and Tharreau, D. (2003). Identification and fine mapping of Pi33, the rice resistance gene corresponding to the Magnaporthe grisea avirulence gene ACE1. *Theor. Appl. Genet.* 107, 1139–1147.

Böhm, H., Albert, I., Fan, L., Reinhard, A., and Nürnberger, T. (2014a). Immune receptor complexes at the plant cell surface. *Curr. Opin. Plant Biol.* 20, 47–54.

Böhm, H., Albert, I., Fan, L., Reinhard, A., and Nürnberger, T. (2014b). Immune receptor complexes at the plant cell surface. *Curr. Opin. Plant Biol.* 20C, 47–54.

Bohnert, H.U. (2004). A Putative Polyketide Synthase/Peptide Synthetase from Magnaporthe grisea Signals Pathogen Attack to Resistant Rice. *PLANT CELL ONLINE* 16, 2499–2513.

Boller, T., and Felix, G. (2009). A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. *Annu. Rev. Plant Biol.* 60, 379–406.

Bonardi, V., Tang, S., Stallmann, A., Roberts, M., Cherkis, K., and Dangl, J.L. (2011). Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16463–16468.

Boudsocq, M., Willmann, M.R., McCormack, M., Lee, H., Shan, L., He, P., Bush, J., Cheng, S.-H., and Sheen, J. (2010). Differential innate immune signalling via Ca²⁺ sensor protein kinases. *Nature* *464*, 418–422.

Bourett, T.M., Sweigard, J. a, Czymmek, K.J., Carroll, A., and Howard, R.J. (2002). Reef coral fluorescent proteins for visualizing fungal pathogens. *Fungal Genet. Biol. FG B* *37*, 211–220.

Boutemy, L.S., King, S.R.F., Win, J., Hughes, R.K., Clarke, T.A., Blumenschein, T.M.A., Kamoun, S., and Banfield, M.J. (2011). Structures of Phytophthora RXLR Effector Proteins: A CONSERVED BUT ADAPTABLE FOLD UNDERPINS FUNCTIONAL DIVERSITY. *J. Biol. Chem.* *286*, 35834–35842.

Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* *2*, 2728–2733.

Bull, P.C., and Cox, D.W. (1994). Wilson disease and Menkes disease: new handles on heavy-metal transport. *Trends Genet.* *10*, 246–252.

Caplan, J., Padmanabhan, M., and Dinesh-Kumar, S.P. (2008). Plant NB-LRR Immune Receptors: From Recognition to Transcriptional Reprogramming. *Cell Host Microbe* *3*, 126–135.

Caracuel-Rios, Z., and Talbot, N.J. (2007). Cellular differentiation and host invasion by the rice blast fungus *Magnaporthe grisea*. *Curr. Opin. Microbiol.* *10*, 339–345.

Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., Rivas, S., Alaux, L., Kanzaki, H., Okuyama, Y., et al. (2013a). The Rice Resistance Protein Pair RGA4/RGA5 Recognizes the *Magnaporthe oryzae* Effectors AVR-Pia and AVR1-CO39 by Direct Binding. *Plant Cell Online*.

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

Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T., and Dodds, P.N. (2014a). A novel conserved mechanism for plant NLR protein pairs: the “integrated decoy” hypothesis. *Front. Plant Sci.* 5.

Césari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., Shimamoto, K., Dodds, P., Terauchi, R., and Kroj, T. (2014a). The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *EMBO J.* 33, 1941–1959.

Césari, S., Bernoux, M., Moncuquet, P., Kroj, T., and Dodds, P.N. (2014b). A novel conserved mechanism for plant NLR protein pairs: the “integrated decoy” hypothesis. *Front. Plant Sci.* 5, 606.

Césari, S., Bernoux, M., Moncuquet, P., Kroj, T., and Dodds, P.N. (2014c). A novel conserved mechanism for plant NLR protein pairs : the “ integrated decoy ” hypothesis. *Front. Plant Sci.* 5, 577.

Cesari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., Shimamoto, K., Dodds, P., Terauchi, R., and Kroj, T. (2014b). The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *EMBO J.* 33, 1941–1959.

Chauhan, R., Farman, M., Zhang, H.-B., and Leong, S. (2002). Genetic and physical mapping of a rice blast resistance locus, Pi-CO39(t) , that corresponds to the avirulence gene AVR1-CO39 of *Magnaporthe grisea*. *Mol. Genet. Genomics* 267, 603–612.

Chen, C., Chen, M., Hu, J., Zhang, W., and Zhong, Z. (2014). Sequence Variation and Recognition Specificity of the Avirulence Gene AvrPiz-T in *Magnaporthe Oryzae* Field Populations. *Fungal Genom Biol* 4, 2.

Chen, Q., Han, Z., Jiang, H., Tian, D., and Yang, S. (2010). Strong Positive Selection Drives Rapid Diversification of R-Genes in Arabidopsis Relatives. *J. Mol. Evol.* 70, 137–148.

Chen, S., Tao, L., Zeng, L., Vega-Sanchez, M.E., Umemura, K., and Wang, G.-L. (2006). A highly efficient transient protoplast system for analyzing defence gene

expression and protein-protein interactions in rice. *Mol. Plant Pathol.* 7, 417–427.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J.D.G., Felix, G., and Boller, T. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* 448, 497–500.

Chisholm, S.T., Coaker, G., Day, B., and Staskawicz, B.J. (2006). Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response. *Cell* 124, 803–814.

Chou, S., Krasileva, K.V., Holton, J.M., Steinbrenner, A.D., Alber, T., and Staskawicz, B.J. (2011a). Hyaloperonospora arabidopsidis ATR1 effector is a repeat protein with distributed recognition surfaces. *Proc. Natl. Acad. Sci.* 108, 13323–13328.

Chou, S., Krasileva, K. V, Holton, J.M., Steinbrenner, A.D., Alber, T., and Staskawicz, B.J. (2011b). Hyaloperonospora arabidopsidis ATR1 effector is a repeat protein with distributed recognition surfaces. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13323–13328.

Chung, E.-H., da Cunha, L., Wu, A.-J., Gao, Z., Cherkis, K., Afzal, A.J., Mackey, D., and Dangl, J.L. (2011). Specific Threonine Phosphorylation of a Host Target by Two Unrelated Type III Effectors Activates a Host Innate Immune Receptor in Plants. *Cell Host Microbe* 9, 125–136.

Collier, S.M., and Moffett, P. (2009a). NB-LRRs work a “bait and switch” on pathogens. *Trends Plant Sci.* 14, 521–529.

Collier, S.M., and Moffett, P. (2009b). NB-LRRs work a “bait and switch” on pathogens. *Trends Plant Sci.* 14, 521–529.

Cook, D.E., Mesarich, C.H., and Thomma, B.P.H.J. (2014). Understanding Plant Immunity as a Surveillance System to Detect Invasion. *Annu. Rev. Phytopathol.* 53, 150605182533006.

Couch, B.C., and Kohn, L.M. (2002). A multilocus gene genealogy concordant with host preference indicates segregation of a new species, *Magnaporthe oryzae*,

from *M. grisea*. *Mycologia* **94**, 683–693.

Cui, H., Tsuda, K., and Parker, J.E. (2015a). Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. *Annu. Rev. Plant Biol.* **66**, 487–511.

Cui, H., Tsuda, K., and Parker, J.E. (2015b). Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. *Annu. Rev. Plant Biol.* **66**, 487–511.

Danot, O., Marquenet, E., Vidal-Ingigliardi, D., and Richet, E. (2009). Wheel of Life, Wheel of Death: A Mechanistic Insight into Signaling by STAND Proteins. *Structure* **17**, 172–182.

Dean, R., Van Kan, J.A.L., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., Ellis, J., et al. (2012). The Top 10 fungal pathogens in molecular plant pathology: Top 10 fungal pathogens. *Mol. Plant Pathol.* **13**, 414–430.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.-R., and Pan, H. (2005a). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980–986.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.-R., and Pan, H. (2005b). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980–986.

Deslandes, L., Olivier, J., Peeters, N., Feng, D.X., Khounlotham, M., Boucher, C., Somssich, I., Genin, S., and Marco, Y. (2003a). Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc. Natl. Acad. Sci.* **100**, 8024–8029.

Deslandes, L., Olivier, J., Peeters, N., Feng, D.X., Khounlotham, M., Boucher, C., Somssich, I., Genin, S., and Marco, Y. (2003b). Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8024–8029.

Di Matteo, A., Federici, L., Mattei, B., Salvi, G., Johnson, K.A., Savino, C., De Lorenzo, G., Tsernoglou, D., and Cervone, F. (2003). The crystal structure of

polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proc. Natl. Acad. Sci.* *100*, 10124–10128.

Dodds, P.N., and Rathjen, J.P. (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* *11*, 539–548.

Dodds, P.N., Lawrence, G.J., and Ellis, J.G. (2001). Six amino acid changes confined to the leucine-rich repeat β -strand/ β -turn motif determine the difference between the P and P2 rust resistance specificities in flax. *Plant Cell* *13*, 163–178.

Dong, S., Raffaele, S., and Kamoun, S. (2015a). The two-speed genomes of filamentous pathogens: waltz with plants. *Curr. Opin. Genet. Dev.* *35*, 57–65.

Dong, Y., Li, Y., Zhao, M., Jing, M., Liu, X., Liu, M., Guo, X., Zhang, X., Chen, Y., Liu, Y., et al. (2015b). Global Genome and Transcriptome Analyses of *Magnaporthe oryzae* Epidemic Isolate 98-06 Uncover Novel Effectors and Pathogenicity-Related Genes, Revealing Gene Gain and Lose Dynamics in Genome Evolution. *PLOS Pathog.* *11*, e1004801.

Ellis, J.G., Lawrence, G.J., Luck, J.E., and Dodds, P.N. (1999). Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell* *11*, 495–506.

Ellis, J.G., Dodds, P.N., and Lawrence, G.J. (2007). Flax Rust Resistance Gene Specificity is Based on Direct Resistance-Avirulence Protein Interactions. *Annu. Rev. Phytopathol.* *45*, 289–306.

Faivre-Rampant, O., Thomas, J., Allègre, M., Morel, J.-B., Tharreau, D., Nottèghem, J.-L., Lebrun, M.-H., Schaffrath, U., and Piffanelli, P. (2008). Characterization of the model system rice-*Magnaporthe* for the study of nonhost resistance in cereals. *New Phytol.* *180*, 899–910.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* *39*, W29–W37.

Flor, H.. (1971). Current status of the gene-for-gene concept. *Annu Rev Phytopathol* 275–296.

Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., Hayashi, N., Takahashi, A., Hirochika, H., Okuno, K., et al. (2009a). Loss of Function of a Proline-Containing Protein Confers Durable Disease Resistance in Rice. *Science* 325, 998–1001.

Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., Hayashi, N., Takahashi, A., Hirochika, H., Okuno, K., et al. (2009b). Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science* 325, 998–1001.

Fuller, D.Q. (2011). Pathways to Asian Civilizations: Tracing the Origins and Spread of Rice and Rice Cultures. *Rice* 4, 78–92.

Gabriëls, S.H.E.J., Vossen, J.H., Ekengren, S.K., van Ooijen, G., Abd-El-Halim, A.M., van den Berg, G.C.M., Rainey, D.Y., Martin, G.B., Takken, F.L.W., de Wit, P.J.G.M., et al. (2007). An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins. *Plant J. Cell Mol. Biol.* 50, 14–28.

Gao, Q.-M., Zhu, S., Kachroo, P., and Kachroo, A. (2015). Signal regulators of systemic acquired resistance. *Front. Plant Sci.* 6.

Ge, S., Sang, T., Lu, B.-R., and Hong, D.-Y. (1999). Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc. Natl. Acad. Sci.* 96, 14400–14405.

Giraldo, M.C., and Valent, B. (2013a). Filamentous plant pathogen effectors in action. *Nat. Rev. Microbiol.* 11, 800–814.

Giraldo, M.C., and Valent, B. (2013b). Filamentous plant pathogen effectors in action. *Nat. Rev. Microbiol.* 11, 800–814.

Giraldo, M.C., Dagdas, Y.F., Gupta, Y.K., Mentlak, T.A., Yi, M., Martinez-Rocha, A.L., Saitoh, H., Terauchi, R., Talbot, N.J., and Valent, B. (2013a). Two distinct secretion systems facilitate tissue invasion by the rice blast fungus *Magnaporthe oryzae*. *Nat. Commun.* 4.

Giraldo, M.C., Dagdas, Y.F., Gupta, Y.K., Mentlak, T.A., Yi, M., Martinez-Rocha,

A.L., Saitoh, H., Terauchi, R., Talbot, N.J., and Valent, B. (2013b). Two distinct secretion systems facilitate tissue invasion by the rice blast fungus *Magnaporthe oryzae*. *Nat. Commun.* 4.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., et al. (2002a). A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., et al. (2002b). A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100.

Guillen, K., Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., and Padilla, A. (2015). Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. *PLoS Pathog.* *accepted f.*

de Guillen, K., Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., and Padilla, A. (2015). Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. *PLOS Pathog* 11, e1005228.

Güntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol. Biol. Clifton NJ* 278, 353–378.

Hao, W., Collier, S.M., Moffett, P., and Chai, J. (2013). Structural basis for the interaction between the potato virus X resistance protein (Rx) and its cofactor Ran GTPase-activating protein 2 (RanGAP2). *J. Biol. Chem.* 288, 35868–35876.

Hauck, P., Thilmony, R., and He, S.Y. (2003). A *Pseudomonas syringae* type III effector suppresses cell wall-based extracellular defense in susceptible *Arabidopsis* plants. *Proc. Natl. Acad. Sci.* 100, 8577–8582.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545–W549.

van der Hoorn, R.A.L., and Kamoun, S. (2008a). From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors. *PLANT CELL ONLINE* 20, 2009–2017.

van der Hoorn, R.A.L., and Kamoun, S. (2008b). From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors. *Plant Cell* 20, 2009–2017.

Huang, J., Si, W., Deng, Q., Li, P., and Yang, S. (2014). Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. *BMC Genet.* 15, 45.

Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., Li, W., Zhan, Q., Cheng, B., Xia, J., et al. (2015). Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.* 6, 6258.

Inohara, N., Chamaillard, M., McDonald, C., and Nuñez, G. (2005). NOD-LRR PROTEINS: Role in Host-Microbial Interactions and Inflammatory Disease. *Annu. Rev. Biochem.* 74, 355–383.

Iriti, M., and Faoro, F. (2007). Review of innate and specific immunity in plants and animals. *Mycopathologia* 164, 57–64.

Jacob, F., Vernaldi, S., and Maekawa, T. (2013a). Evolution and Conservation of Plant NLR Functions. *Front. Immunol.* 4.

Jacob, F., Vernaldi, S., and Maekawa, T. (2013b). Evolution and Conservation of Plant NLR Functions. *Front. Immunol.* 4, 297.

Jayamohan, N.S., and Kumudini, B.S. (2011). Host Pathogen Interaction at the Plant Cell Wall. *Int. Res. J. Pharm. Pharmacol.* 1, 242–249.

Jia, Y., McAdams, S., Bryan, G.T., Hershey, H.P., and Valent, B. (2000). Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J.* 19, 4004–4014.

Jones, D.T., and Swindells, M.B. (2002). Getting the most from PSI-BLAST. *Trends Biochem. Sci.* 27, 161–164.

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

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

Kamoun, S. (2007). Groovy times: filamentous pathogen effectors revealed. *Curr. Opin. Plant Biol.* *10*, 358–365.

Kanzaki, H., Yoshida, K., Saitoh, H., Fujisaki, K., Hirabuchi, A., Alaux, L., Fournier, E., Tharreau, D., and Terauchi, R. (2012a). Arms race co-evolution of *Magnaporthe oryzae* AVR-Pik and rice *Pik* genes driven by their physical interactions: *Co-evolution of fungal and rice genes*. *Plant J.* n/a-n/a.

Kanzaki, H., Yoshida, K., Saitoh, H., Fujisaki, K., Hirabuchi, A., Alaux, L., Fournier, E., Tharreau, D., Terauchi, R., Allaux, L., et al. (2012b). Arms race co-evolution of *Magnaporthe oryzae* AVR-Pik and rice *Pik* genes driven by their physical interactions. *Plant J. Cell Mol. Biol.* 894–907.

Kawano, Y., Chen, L., and Shimamoto, K. (2010). The Function of Rac Small GTPase and Associated Proteins in Rice Innate Immunity. *Rice* *3*, 112–121.

Khan, M., Subramaniam, R., and Desveaux, D. (2016). Of guards, decoys, baits and traps: pathogen perception in plants by type III effector sensors. *Curr. Opin. Microbiol.* *29*, 49–55.

Khang, C.H., Park, S.-Y., Lee, Y.-H., Valent, B., and Kang, S. (2008). Genome organization and evolution of the AVR-Pita avirulence gene family in the *Magnaporthe grisea* species complex. *Mol. Plant. Microbe Interact.* *21*, 658–670.

Khang, C.H., Berruyer, R., Giraldo, M.C., Kankanala, P., Park, S.-Y., Czymmek, K., Kang, S., and Valent, B. (2010a). Translocation of *Magnaporthe oryzae* Effectors into Rice Cells and Their Subsequent Cell-to-Cell Movement. *Plant Cell* *22*, 1388–1403.

Khang, C.H., Berruyer, R., Giraldo, M.C., Kankanala, P., Park, S.-Y., Czymmek, K., Kang, S., and Valent, B. (2010b). Translocation of *Magnaporthe oryzae* effectors into rice cells and their subsequent cell-to-cell movement. *Plant Cell* *22*, 1388–1403.

Kobe, B., and Deisenhofer, J. (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine rich repeats. *Nature* 751–756.

Kobe, B., and Kajava, A.V. (2001). The leucine-rich repeat as a protein

recognition motif. *Curr. Opin. Struct. Biol.* *11*, 725–732.

Koeck, M., Hardham, A.R., and Dodds, P.N. (2011). The role of effectors of biotrophic and hemibiotrophic fungi in infection: Effectors of biotrophic fungi. *Cell. Microbiol.* *13*, 1849–1857.

Krasileva, K. V, Dahlbeck, D., and Staskawicz, B.J. (2010). Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. *Plant Cell* *22*, 2444–2458.

Kroj, T., Chanclud, E., Michel-Romiti, C., Grand, X., and Morel, J.-B. (2016). Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytol.* n/a-n/a.

Laskowski, R.A., Moss, D.S., and Thornton, J.M. (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* *231*, 1049–1067.

Lee, S.-K., Song, M.-Y., Seo, Y.-S., Kim, H.-K., Ko, S., Cao, P.-J., Suh, J.-P., Yi, G., Roh, J.-H., Lee, S., et al. (2009). Rice Pi5-Mediated Resistance to *Magnaporthe oryzae* Requires the Presence of Two Coiled-Coil-Nucleotide-Binding-Leucine-Rich Repeat Genes. *Genetics* *181*, 1627–1638.

Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Trémousaygue, D., Kraut, A., Zhou, B., Levaillant, M., Adachi, H., Yoshioka, H., et al. (2015). A Receptor Pair with an Integrated Decoy Converts Pathogen Disabling of Transcription Factors to Immunity. *Cell* *161*, 1074–1088.

Lewis, J.D., Lee, A.H.-Y., Hassan, J.A., Wan, J., Hurley, B., Jhingree, J.R., Wang, P.W., Lo, T., Youn, J.-Y., Guttman, D.S., et al. (2013). The Arabidopsis ZED1 pseudokinase is required for ZAR1-mediated immunity induced by the *Pseudomonas syringae* type III effector HopZ1a. *Proc. Natl. Acad. Sci.* *110*, 18722–18727.

Li, W., Wang, B., Wu, J., Lu, G., Hu, Y., Zhang, X., Zhang, Z., Zhao, Q., Feng, Q., Zhang, H., et al. (2009a). The *Magnaporthe oryzae* avirulence gene *AvrPiz-t* encodes a predicted secreted protein that triggers the immunity in rice mediated by the blast resistance gene *Piz-t*. *Mol. Plant. Microbe Interact.* *22*, 411–420.

Li, W., Wang, B., Wu, J., Lu, G., Hu, Y., Zhang, X., Zhang, Z., Zhao, Q., Feng, Q., Zhang, H., et al. (2009b). The Magnaporthe oryzae avirulence gene AvrPiz-t encodes a predicted secreted protein that triggers the immunity in rice mediated by the blast resistance gene Piz-t. *Mol. Plant-Microbe Interact. MPMI* 22, 411–420.

Li, Y., Huang, F., Lu, Y., Shi, Y., Zhang, M., Fan, J., and Wang, W. (2013). Mechanism of plant–microbe interaction and its utilization in disease-resistance breeding for modern agriculture. *Physiol. Mol. Plant Pathol.* 83, 51–58.

Liu, P., Hu, Z., Zhou, B., Liu, S., and Chai, J. (2013). Crystal structure of an LRR protein with two solenoids. *Cell Res.* 23, 303.

Liu, W., Liu, J., Triplett, L., Leach, J.E., and Wang, G.-L. (2014a). Novel Insights into Rice Innate Immunity Against Bacterial and Fungal Pathogens. *Annu. Rev. Phytopathol.* 52, 213–241.

Liu, W., Liu, J., Triplett, L., Leach, J.E., and Wang, G.-L. (2014b). Novel Insights into Rice Innate Immunity Against Bacterial and Fungal Pathogens. *Annu. Rev. Phytopathol.* 52, 213–241.

Liu, W., Liu, J., Triplett, L., Leach, J.E., and Wang, G.-L. (2014c). Novel Insights into Rice Innate Immunity Against Bacterial and Fungal Pathogens.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., and Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. *Annu. Rev. Plant Biol.* 66, 513–545.

Lukasik, E., and Takken, F.L. (2009). STANDING strong, resistance proteins instigators of plant defence. *Curr. Opin. Plant Biol.* 12, 427–436.

Mackey, D., Holt, B.F., Wiig, A., and Dangl, J.L. (2002). RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* 108, 743–754.

Maekawa, T., Kufer, T.A., and Schulze-Lefert, P. (2011a). NLR functions in plant and animal immune systems: so far and yet so close. *Nat. Immunol.* 12, 817–826.

Maekawa, T., Cheng, W., Spiridon, L.N., Töller, A., Lukasik, E., Saijo, Y., Liu, P.,

Shen, Q.-H., Micluta, M.A., Somssich, I.E., et al. (2011b). Coiled-Coil Domain-Dependent Homodimerization of Intracellular Barley Immune Receptors Defines a Minimal Functional Module for Triggering Cell Death. *Cell Host Microbe* 9, 187–199.

Maekawa, T., Cheng, W., Spiridon, L.N., Töller, A., Lukasik, E., Saijo, Y., Liu, P., Shen, Q.-H., Micluta, M. a, Somssich, I.E., et al. (2011c). Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. *Cell Host Microbe* 9, 187–199.

Manning, V.A., Hamilton, S.M., Karplus, P.A., and Ciuffetti, L.M. (2008). The Arg-Gly-Asp-containing, solvent-exposed loop of Ptr ToxA is required for internalization. *Mol. Plant. Microbe Interact.* 21, 315–325.

Maqbool, a, Saitoh, H., Franceschetti, M., Stevenson, C., Uemura, a, Kanzaki, H., Kamoun, S., Terauchi, R., and Banfield, M. (2015a). Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife* 4, 1–24.

Maqbool, A., Saitoh, H., Franceschetti, M., Stevenson, C.E.M., Uemura, A., Kanzaki, H., Kamoun, S., Terauchi, R., and Banfield, M.J. (2015b). Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife* 4, e08709.

Marcotte, E.M., Pellegrini, M., Ng, H.-L., Rice, D.W., Yeates, T.O., and Eisenberg, D. (1999). Detecting Protein Function and Protein-Protein Interactions from Genome Sequences. *Science* 285.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-Wide Analysis of NBS-LRR–Encoding Genes in Arabidopsis. *Plant Cell Online* 15, 809–834.

Mosquera, G., Giraldo, M.C., Khang, C.H., Coughlan, S., and Valent, B. (2009). Interaction Transcriptome Analysis Identifies Magnaporthe oryzae BAS1-4 as Biotrophy-Associated Secreted Proteins in Rice Blast Disease. *PLANT CELL ONLINE* 21, 1273–1290.

Nachimuthu, V.V., Muthurajan, R., Duraijalaguraja, S., Sivakami, R., Pandian,

B.A., Ponniah, G., Gunasekaran, K., Swaminathan, M., K K, S., and Sabariappan, R. (2015). Analysis of Population Structure and Genetic Diversity in Rice Germplasm Using SSR Markers: An Initiative Towards Association Mapping of Agronomic Traits in *Oryza Sativa*. *Rice* 8.

Nederveen, A.J., Doreleijers, J.F., Vranken, W., Miller, Z., Spronk, C.A.E.M., Nabuurs, S.B., Güntert, P., Livny, M., Markley, J.L., Nilges, M., et al. (2005). RECOORD: A recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins Struct. Funct. Bioinforma.* 59, 662–672.

Nishimura, M.T., and Dangl, J.L. (2014). Plant science. Paired plant immune receptors. *Science* 344, 267–268.

Nüumberger, T., Brunner, F., Kemmerling, B., and Piater, L. (2004). Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol. Rev.* 198, 249–266.

Nyarko, A., Singarapu, K.K., Figueroa, M., Manning, V.A., Pandelova, I., Wolpert, T.J., Ciuffetti, L.M., and Barbar, E. (2014a). Solution NMR Structures of *Pyrenophora tritici-repentis* ToxB and Its Inactive Homolog Reveal Potential Determinants of Toxin Activity. *J. Biol. Chem.* 289, 25946–25956.

Nyarko, A., Singarapu, K.K., Figueroa, M., Manning, V. a., Pandelova, I., Wolpert, T.J., Ciuffetti, L.M., and Barbar, E. (2014b). Solution NMR Structures of *Pyrenophora tritici-repentis* ToxB and Its Inactive Homolog Reveal Potential Determinants of Toxin Activity. *J. Biol. Chem.* 289, 25946–25956.

Okuyama, Y., Kanzaki, H., Abe, A., Yoshida, K., Tamiru, M., Saitoh, H., Fujibe, T., Matsumura, H., Shenton, M., Galam, D.C., et al. (2011a). A multifaceted genomics approach allows the isolation of the rice Pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes: Rice Pia R gene. *Plant J.* 66, 467–479.

Okuyama, Y., Kanzaki, H., Abe, A., Yoshida, K., Tamiru, M., Saitoh, H., Fujibe, T., Matsumura, H., Shenton, M., Galam, D.C., et al. (2011b). A multifaceted genomics approach allows the isolation of the rice Pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes. *Plant J. Cell Mol. Biol.* 66, 467–479.

Oliveira-Garcia, E., and Valent, B. (2015). How eukaryotic filamentous pathogens evade plant recognition. *Curr. Opin. Microbiol.* 26, 92–101.

Ose, T., Oikawa, A., Nakamura, Y., Maenaka, K., Higuchi, Y., Satoh, Y., Fujiwara, S., Demura, M., Sone, T., and Kamiya, M. (2015). Solution structure of an avirulence protein, AVR-Pia, from *Magnaporthe oryzae*. *J. Biomol. NMR.*

Padmanabhan, M., Cournoyer, P., and Dinesh-Kumar, S.P. (2009). The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. *Cell. Microbiol.* 11, 191–198.

Park, C.-H., Chen, S., Shirsekar, G., Zhou, B., Khang, C.H., Songkumarn, P., Afzal, A.J., Ning, Y., Wang, R., Bellizzi, M., et al. (2012). The *Magnaporthe oryzae* Effector AvrPiz-t Targets the RING E3 Ubiquitin Ligase APIP6 to Suppress Pathogen-Associated Molecular Pattern-Triggered Immunity in Rice. *Plant Cell* 24, 4748–4762.

Pennisi, E. Armed and Dangerous.

Petre, B., and Kamoun, S. (2014). How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biol.* 12, e1001801.

Qi, D., and Innes, R. (2013a). Recent Advances in Plant NLR Structure, Function, Localization, and Signaling. *Front. Immunol.* 221.

Qi, D., and Innes, R.W. (2013b). Recent Advances in Plant NLR Structure, Function, Localization, and Signaling. *Front. Immunol.* 4.

Rafiqi, M., Bernoux, M., Ellis, J.G., and Dodds, P.N. (2009). In the trenches of plant pathogen recognition: Role of NB-LRR proteins. *Semin. Cell Dev. Biol.* 20, 1017–1024.

Rafiqi, M., Gan, P.H.P., Ravensdale, M., Lawrence, G.J., Ellis, J.G., Jones, D.A., Hardham, A.R., and Dodds, P.N. (2010). Internalization of flax rust avirulence proteins into flax and tobacco cells can occur in the absence of the pathogen. *Plant Cell* 22, 2017–2032.

Rairdan, G.J., and Moffett, P. (2006). Distinct Domains in the ARC Region of the

Potato Resistance Protein Rx Mediate LRR Binding and Inhibition of Activation. *PLANT CELL ONLINE* 18, 2082–2093.

Rairdan, G.J., Collier, S.M., Sacco, M.A., Baldwin, T.T., Boettrich, T., and Moffett, P. (2008). The Coiled-Coil and Nucleotide Binding Domains of the Potato Rx Disease Resistance Protein Function in Pathogen Recognition and Signaling. *PLANT CELL ONLINE* 20, 739–751.

Ribot, C., Hirsch, J., Balzergue, S., Tharreau, D., Nottéghem, J.-L., Lebrun, M.-H., and Morel, J.-B. (2008). Susceptibility of rice to the blast fungus, *Magnaporthe grisea*. *J. Plant Physiol.* 165, 114–124.

Ribot, C., Césari, S., Abidi, I., Chalvon, V., Bournaud, C., Vallet, J., Lebrun, M.-H., Morel, J.-B., and Kroj, T. (2012). The *Magnaporthe oryzae* effector AVR1-CO39 is translocated into rice cells independently of a fungal-derived machinery. *Plant J. Cell Mol. Biol.* 74, 1–12.

Ribot, C., Césari, S., Abidi, I., Chalvon, V., Bournaud, C., Vallet, J., Lebrun, M.-H., Morel, J.-B., and Kroj, T. (2013). The *Magnaporthe oryzae* effector AVR1–CO39 is translocated into rice cells independently of a fungal-derived machinery. *Plant J.* 74, 1–12.

Riedl, S.J., Li, W., Chao, Y., Choi, H.J., Schwarzenbacher, R., and Shi, Y. (2005). Structure of the apoptotic protease-activating factor 1 bound to ADP. *Nature* 434, 927–933.

Ronald, P.C., and Beutler, B. (2010). Plant and Animal Sensors of Conserved Microbial Signatures. *Science* 330, 1061–1064.

Rouxel, T., and Balesdent, M.-H. (2010). Avirulence Genes. In *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd, ed. (Chichester, UK: John Wiley & Sons, Ltd), p.

Saitoh, H., Fujisawa, S., Mitsuoka, C., Ito, A., Hirabuchi, A., Ikeda, K., Irieda, H., Yoshino, K., Yoshida, K., Matsumura, H., et al. (2012). Large-Scale Gene Disruption in *Magnaporthe oryzae* Identifies MC69, a Secreted Protein Required for Infection by Monocot and Dicot Fungal Pathogens. *PLoS Pathog.* 8, e1002711.

Sarris, P.F., Duxbury, Z., Huh, S.U., Ma, Y., Segonzac, C., Sklenar, J., Derbyshire, P., Cevik, V., Rallapalli, G., Saucet, S.B., et al. (2015). A Plant Immune Receptor Detects Pathogen Effectors that Target WRKY Transcription Factors. *Cell* 161, 1089–1100.

Schneider, D.R.S., Saraiva, A.M., Azzoni, A.R., Miranda, H.R.C.A.N., de Toledo, M.A.S., Pelloso, A.C., and Souza, A.P. (2010). Overexpression and purification of PWL2D, a mutant of the effector protein PWL2 from *Magnaporthe grisea*. *Protein Expr. Purif.* 74, 24–31.

Schornack, S., Huitema, E., Cano, L.M., Bozkurt, T.O., Oliva, R., Van Damme, M., Schwizer, S., Raffaele, S., Chaparro-Garcia, A., Farrer, R., et al. (2009). Ten things to know about oomycete effectors. *Mol. Plant Pathol.* 10, 795–803.

Schumann, F.H., Riepl, H., Maurer, T., Gronwald, W., Neidig, K.P., and Kalbitzer, H.R. (2007). Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J. Biomol. NMR* 39, 275–289.

Senthil-Kumar, M., and Mysore, K.S. (2013). Nonhost Resistance Against Bacterial Pathogens: Retrospectives and Prospects. *Annu. Rev. Phytopathol.* 51, 407–427.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44, 213–223.

Skamnioti, P., and Gurr, S.J. (2009a). Against the grain: safeguarding rice from rice blast disease. *Trends Biotechnol.* 27, 141–150.

Skamnioti, P., and Gurr, S.J. (2009b). Against the grain: safeguarding rice from rice blast disease. *Trends Biotechnol.* 27, 141–150.

Sohn, K.H., Hughes, R.K., Piquerez, S.J., Jones, J.D.G., and Banfield, M.J. (2012). Distinct regions of the *Pseudomonas syringae* coiled-coil effector AvrRps4 are required for activation of immunity. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16371–16376.

Sperschneider, J., Dodds, P.N., Gardiner, D.M., Manners, J.M., Singh, K.B., and Taylor, J.M. (2015). Advances and Challenges in Computational Prediction of Effectors from Plant Pathogenic Fungi. *PLOS Pathog.* *11*, e1004806.

Spoel, S.H., and Dong, X. (2012). How do plants achieve immunity? Defence without specialized immune cells. *Nat. Rev. Immunol.* *12*, 89–100.

Steinbrenner, A.D., Goritschnig, S., and Staskawicz, B.J. (2015). Recognition and Activation Domains Contribute to Allele-Specific Responses of an Arabidopsis NLR Receptor to an Oomycete Effector Protein. *PLOS Pathog.* *11*, e1004665.

Stergiopoulos, I., and de Wit, P.J.G.. (2009). Fungal Effector Proteins. *Annu. Rev. Phytopathol.* *47*, 233–263.

Stuart, L.M., Paquette, N., and Boyer, L. (2013). Effector-triggered versus pattern-triggered immunity: how animals sense pathogens. *Nat. Rev. Immunol.* *13*, 199–206.

Studier, F.W. (2005a). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* *41*, 207–234.

Studier, F.W. (2005b). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* *41*, 207–234.

Sweeney, M., and McCouch, S. (2007). The Complex History of the Domestication of Rice. *Ann. Bot.* *100*, 951–957.

Takken, F.L., and Goverse, A. (2012a). How to build a pathogen detector: structural basis of NB-LRR function. *Curr. Opin. Plant Biol.* *15*, 375–384.

Takken, F.L.W., and Goverse, A. (2012b). How to build a pathogen detector: structural basis of NB-LRR function. *Curr. Opin. Plant Biol.* *15*, 375–384.

Takken, F.L.W., and Tameling, W.I.L. (2009). To Nibble at Plant Resistance Proteins. *Science* *324*, 744–746.

Takken, F.L., Albrecht, M., and Tameling, W.I. (2006). Resistance proteins: molecular switches of plant defence. *Curr. Opin. Plant Biol.* *9*, 383–390.

Tameling, W.I.L. (2006). Mutations in the NB-ARC Domain of I-2 That Impair ATP

Hydrolysis Cause Autoactivation. *PLANT Physiol.* *140*, 1233–1245.

Tapping, R.I. (2009). Innate immune sensing and activation of cell surface Toll-like receptors. *Semin. Immunol.* *21*, 175–184.

Trdàj, L., Boutrot, F., Claverie, J., Brulé, D., Dorey, S., and Poinssot, B. (2015). Perception of pathogenic or beneficial bacteria and their evasion of host immunity: pattern recognition receptors in the frontline. *Front. Plant Sci.* *6*.

Tsiamis, G., Mansfield, J.W., Hockenhull, R., Jackson, R.W., Sesma, A., Athanassopoulos, E., Bennett, M.A., Stevens, C., Vivian, A., Taylor, J.D., et al. (2000). Cultivar-specific avirulence and virulence functions assigned to avrPphF in *Pseudomonas syringae* pv. *phaseolicola*, the cause of bean halo-blight disease. *EMBO* *19*, 3204–3214.

Van Der Biezen, E.A., and Jones, J.D. (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem. Sci.* *23*, 454–456.

Ve, T., Williams, S.J., Catanzariti, A.-M., Rafiqi, M., Rahman, M., Ellis, J.G., Hardham, A.R., Jones, D.A., Anderson, P.A., Dodds, P.N., et al. (2013). Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 17594–17599.

Villalba, F., Collemare, J., Landraud, P., Lambou, K., Brozek, V., Cirer, B., Morin, D., Bruel, C., Beffa, R., and Lebrun, M.-H. (2008). Improved gene targeting in *Magnaporthe grisea* by inactivation of MgKU80 required for non-homologous end joining. *Fungal Genet. Biol.* *FG B* *45*, 68–75.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Bioinforma.* *59*, 687–696.

Wang, C.-I.A., Guncar, G., Forwood, J.K., Teh, T., Catanzariti, A.-M., Lawrence, G.J., Loughlin, F.E., Mackay, J.P., Schirra, H.J., Anderson, P.A., et al. (2007a). Crystal Structures of Flax Rust Avirulence Proteins AvrL567-A and -D Reveal Details of the Structural Basis for Flax Disease Resistance Specificity. *PLANT CELL*

ONLINE 19, 2898–2912.

Wang, C.-I.A.C.-I. a, Guncar, G., Forwood, J.K., Teh, T., Catanzariti, A.-M.A.-M., Lawrence, G.J., Loughlin, F.E., Mackay, J.P., Schirra, H.J., Anderson, P.A., et al. (2007b). Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell* 19, 2898–2912.

Whisson, S.C., Boevink, P.C., Moleleki, L., Avrova, A.O., Morales, J.G., Gilroy, E.M., Armstrong, M.R., Grouffaud, S., van West, P., Chapman, S., et al. (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* 450, 115–118.

Williams, S.J., Sohn, K.H., Wan, L., Bernoux, M., Sarris, P.F., Segonzac, C., Ve, T., Ma, Y., Saucet, S.B., Ericsson, D.J., et al. (2014). Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science* 344, 299–303.

Wilson, R.A., and Talbot, N.J. (2009). Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nat. Rev. Microbiol.* 7, 185–195.

Win, J., Krasileva, K.V., Kamoun, S., Shirasu, K., Staskawicz, B.J., and Banfield, M.J. (2012). Sequence Divergent RXLR Effectors Share a Structural Fold Conserved across Plant Pathogenic Oomycete Species. *PLoS Pathog.* 8, e1002400.

Wirthmueller, L., Maqbool, A., and Banfield, M.J. (2013). On the front line: structural insights into plant–pathogen interactions. *Nat. Rev. Microbiol.* 11, 761–776.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., and Sykes, B.D. (1995). ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* 6, 135–140.

Wong, G.K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *indica*). *Science*.

Wu, C.-H., Krasileva, K.V., Banfield, M.J., Terauchi, R., and Kamoun, S. (2015a). The “sensor domains” of plant NLR proteins: more than decoys? *Front. Plant Sci.* 6.

Wu, C.-H., Belhaj, K., Bozkurt, T.O., Birk, M.S., and Kamoun, S. (2016). Helper NLR proteins NRC2a/b and NRC3 but not NRC1 are required for Pto-mediated cell death and resistance in *Nicotiana benthamiana*. *New Phytol.* doi: <http://dx.doi.org/10.1101/015479>.

Wu, J., Kou, Y., Bao, J., Li, Y., Tang, M., Zhu, X., Ponaya, A., Xiao, G., Li, J., Li, C., et al. (2015b). Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice.

Xing, W., Zou, Y., Liu, Q., Liu, J., Luo, X., Huang, Q., Chen, S., Zhu, L., Bi, R., Hao, Q., et al. (2007). The structural basis for activation of plant immunity by bacterial effector protein AvrPto. *Nature* **449**, 243–247.

Xiong, L., and Yang, Y. (2003). Disease Resistance and Abiotic Stress Tolerance in Rice Are Inversely Modulated by an Abscisic Acid-Inducible Mitogen-Activated Protein Kinase. *PLANT CELL ONLINE* **15**, 745–759.

Yaeno, T., Li, H., Chaparro-Garcia, A., Schornack, S., Koshiba, S., Watanabe, S., Kigawa, T., Kamoun, S., and Shirasu, K. (2011). Phosphatidylinositol monophosphate-binding interface in the oomycete RXLR effector AVR3a is required for its stability in host cells to modulate plant immunity. *Proc. Natl. Acad. Sci.* **108**, 14682–14687.

Yan, N., Chai, J., Lee, E.S., Gu, L., Liu, Q., He, J., Wu, J.-W., Kokel, D., Li, H., Hao, Q., et al. (2005). Structure of the CED-4–CED-9 complex provides insights into programmed cell death in *Caenorhabditis elegans*. *Nature* **437**, 831–837.

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

Zhang, S., and Xu, J.-R. (2014). Effectors and Effector Delivery in *Magnaporthe oryzae*. *PLoS Pathog.* **10**, e1003826.

Zhang, S., Wang, L., Wu, W., He, L., Yang, X., and Pan, Q. (2015a). Function and evolution of *Magnaporthe oryzae* avirulence gene AvrPib responding to the rice

blast resistance gene Pib. *Sci. Rep.* **5**, 11642.

Zhang, S., Wang, L., Wu, W., He, L., Yang, X., and Pan, Q. (2015b). Function and evolution of *Magnaporthe oryzae* avirulence gene AvrPib responding to the rice blast resistance gene Pib. *Sci. Rep.* **5**, 11642.

Zhang, Y., Su, J., Duan, S., Ao, Y., Dai, J., Liu, J., Wang, P., Li, Y., Liu, B., Feng, D., et al. (2011). A highly efficient rice green tissue protoplast system for transient gene expression and studying light/chloroplast-related processes. *Plant Methods* **7**.

Zhang, Z.-M., Zhang, X., Zhou, Z.-R., Hu, H.-Y., Liu, M., Zhou, B., and Zhou, J. (2013a). Solution structure of the *Magnaporthe oryzae* avirulence protein AvrPiz-t. *J. Biomol. NMR* **55**, 219–223.

Zhang, Z.-M., Zhang, X., Zhou, Z.-R., Hu, H.-Y., Liu, M., Zhou, B., and Zhou, J. (2013b). Solution structure of the *Magnaporthe oryzae* avirulence protein AvrPiz-t. *J. Biomol. NMR* **55**, 219–223.

Zipfel, C. (2014a). Plant pattern-recognition receptors. *Trends Immunol.* **35**, 345–351.

Zipfel, C. (2014b). Plant pattern-recognition receptors. *Trends Immunol.* **35**, 345–351.