

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

DES EFFECTEURS CANDIDATS DE ROUILLE FONGIQUE
NON HOMOLOGUES AGISSENT SUR DES VOIES APPARENTÉES

*UNRELATED FUNGAL RUST CANDIDATE EFFECTORS
ACT ON OVERLAPPING PLANT FUNCTIONS*

THÈSE PRÉSENTÉE
COMME EXIGENCE PARTIELLE DU
DOCTORAT EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE

PAR
KAREN CRISTINE GONÇALVES DOS SANTOS

AVRIL 2021

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

DOCTORAT EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE (PH. D.)

Direction de recherche :

Hugo Germain	Directeur de recherche
--------------	------------------------

Isabel Desgagné-Penix	Codirectrice de recherche
-----------------------	---------------------------

Jury d'évaluation de la thèse :

Hugo Germain	Directeur de recherche
--------------	------------------------

Isabel Desgagné-Penix	Codirectrice de recherche
-----------------------	---------------------------

Geneviève Pépin	Présidente de jury
-----------------	--------------------

Guus Bakkeren	Évaluateur externe
---------------	--------------------

Mathieu Lavallée-Adam	Évaluateur externe
-----------------------	--------------------

Thèse soutenue le 22 avril 2021.

“All interpretations made by a scientist are hypotheses, and all hypotheses are tentative. They must forever be tested and they must be revised if found to be unsatisfactory. Hence, a change of mind in a scientist, and particularly in a great scientist, is not only not a sign of weakness but rather evidence for continuing attention to the respective problem and an ability to test the hypothesis again and again.”

Ernst Mayr

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors. Hugo, you are an incredible teacher. Thank you for inviting me to be a part of your lab and for your support. Isabel, thank you for the supervision and the insightful conversations. To both of you, thank you for giving me room to grow as a researcher and for making sense of my crazy half-baked ideas during our meetings.

A huge thank you to Dr. Sébastien Duplessis for receiving me in your lab so I could take my first steps into bioinformatics and for inviting me to coauthor papers with you and your team. Thanks to Dr. Peter Solomon and all the members of his lab, specially Dr. Megan McDonald, for receiving me at ANU and making me feel so welcomed, and for teaching me so much about DNA extraction for long read sequencing and about bioinformatics.

Thanks to all my lab mates, Joëlle Rancourt, Alexandre Brisson, Zainab Fakih, Mathias Bisailon, Dr. Saifur Rahman, Teura Barff, Andrew Diamond, Ariane Garand, Fadoua Dhaouadi, Dr. Fatima Awwad, Laurence Tousignant, Marguerite Cinq-Mars, Michelle Boivin, Narimene Fradj and Serge Nouemssi. Thank you for all the “sorties du lab”, the lab Christmas parties, lab barbecues and the lunches together. Melodie B. Plourde, thanks for correcting my papers and for lending winter clothes, you basically saved two lives that winter. A special thanks to Geneviève Laperrière for all the laughs we had during my summer internship in 2015, for your warm welcome back, helping me survive my first winter and for taking care of Tom during the holidays and when I went to Australia. Another thanks to Hur Madina, you are an amazing and lovely person and a super skilled researcher, I am lucky that I can call you my friend.

To my new family, Ilse Ileana Cardenas Bates, Aracely Maribel Diaz Garza, Ingrid Berenice Sanchez Carrillo and Elisa Ines Fantino, thanks for keeping me “sane” these four years. Eli and Maribel, thank you for understanding my sense of humor and

sharing my lack of politeness. Ingrid, thanks for the constant pep talk, it is difficult to maintain a strong impostor syndrome having you around. Ilse, thank you for everything, it is still weird how we became friends simply because we did not know anybody else and months later you were already my best friend, and still is. Ilse and Eli, thanks for the driving lessons.

Thanks to my friends and family from Brazil. Specially to Jannie Francianne Guimarães, Gessyca Fernanda, Beatriz Mitidiero Stachissini Arcain and Açucena Veleh Rivas. Obrigada por todas as conversas estranhas, de apoio moral e pelos stickers de duplo sentido. À minha mãe, aos meus irmãos e cunhadas e cunhado, e aos meus sobrinhos: Luís Alberto, Heloísa, e Gabriela, espero ver vocês de volta, e conhecer quem eu ainda não conheço entre vocês. Muchas gracias a mis suegros, Carlos Leonczuk y Adriana Emilia Minetto, por hacerme sentir tan bien accepta en su familia.

Lastly, and most importantly, a huge thanks to Robert Alexis Leonczuk Minetto. You are the best partner I could hope to have. You made this stage of my life enjoyable and easier. Thank you for the incredible food, for taking care of me when I could barely move, for your support, your company. Most of all, thank you for your love. I love you.

RÉSUMÉ

Les champignons de la rouille infectent les plantes de tous les groupes des plantes, des fougères aux angiospermes, et ils menacent la production de cultures importantes, comme le blé, le soja et le peuplier. Pour mieux comprendre les interactions plantes-pathogènes et élucider ce qui rend les plantes susceptibles à ces pathogènes, nous devons étudier des effecteurs, qui sont des protéines du pathogène utilisées pour contrôler l'hôte. *Melampsora larici-populina* (*Mlp*) provoque la rouille du peuplier et code pour au moins 1 184 effecteurs candidats (CEs), mais ces protéines partagent peu d'homologie avec des protéines connues, ce qui empêche la prédiction de leur fonction. Des études antérieures ont sélectionné des CEs de *Mlp* pour leur caractérisation fonctionnelle chez *Nicotiana benthamiana* et *Arabidopsis thaliana*. Ils ont montré que ces protéines s'accumulent dans divers compartiments cellulaires et ont un impact sur la susceptibilité d'*Arabidopsis* à *Pseudomonas syringae* ou à *Hyaloperonospora arabidopsidis*, ce qui indique qu'elles conduisent à une susceptibilité déclenchée par l'effecteur.

L'objectif de ce travail était de mieux caractériser l'impact des CEs de *Mlp* sélectionnés dans les plantes. Pour cela, nous avons effectué des analyses de séquençage d'ARN et de spectrométrie de masse de plantes *Arabidopsis* transgéniques exprimant de manière constitutive des CEs marqués à la GFP. Pour s'assurer que les données étaient correctement normalisées, nous avons utilisé des gènes de référence pour l'analyse de l'expression différentielle. Cependant, les gènes « housekeeping » et les ensembles de références présélectionnés disponibles pour les semis d'*Arabidopsis* ont montré une grande variabilité d'expression dans nos échantillons. En outre, les méthodes de sélection des références sont conçues pour des analyses RT-qPCR ou nécessitent des ensembles prédéterminés de références potentielles ou de gènes cibles. Ainsi, nous avons créé un pipeline pour la sélection de gènes de contrôle interne qui dépend des nombres de lectures pour chaque gène et la taille des gènes. Ce pipeline transforme le nombre de lectures en Transcripts par Million (TPM), puis calcule la valeur TPM minimale pour que les gènes soient considérés comme exprimés dans chaque échantillon. L'expression minimale moyenne est calculée et utilisée pour exclure les gènes faiblement exprimés. Enfin, les gènes restants ayant la plus faible covariance de TPM sont sélectionnés comme références.

En utilisant cette méthode de sélection des gènes de référence pour nos données transcriptomiques, nous avons trouvé un total de 2 299 gènes dérégulés dans nos 14 lignées transgéniques exprimant des CEs. Les ensembles de gènes coexprimés ont été calculés par une analyse de réseau de corrélation pondérée et l'analyse d'enrichissement GO a montré qu'un ensemble de gènes, formé par des gènes régulés à la baisse dans au moins une lignée transgénique, était enrichi en gènes liés à la défense des plantes, ce qui indique que les CEs étudiés ont un impact sur l'immunité des plantes. L'analyse des différentes lignées transgéniques a montré que les voies KEGG « voie de signalisation MAPK » et « interaction plante-pathogène » étaient respectivement surreprésentées dans

les gènes régulés à la baisse de 6 et 5 des 14 lignées transgéniques. L'analyse métabolomique de ces plantes par spectrométrie de masse à ultra-haute résolution a permis de détecter 5 192 composés au total, mais 81,03 % d'entre eux n'ont pas été identifiés en raison du manque d'informations dans les bases de données publiques, et 11,56 % des composés détectés correspondaient à plusieurs métabolites connus, ce qui a empêché leur identification. En comparant les concentrations de métabolites dans les plantes exprimant des CEs avec celles des plantes témoins, nous avons trouvé un total de 680 métabolites dérégulés avec des composés phénoliques et hautement insaturés enrichis parmi métabolites régulés à la baisse et des peptides enrichis parmi les métabolites régulés à la hausse.

Il est intéressant de noter que les analyses transcriptomiques et métabolomiques ont montré que les CEs paralogues ne dérégulaient pas les mêmes gènes ou métabolites, alors que les CEs non-apparentés avaient des patrons de dérégulation des gènes et des métabolites corrélées. Par conséquent, nos résultats suggèrent que la similarité des séquences des effecteurs et leur appartenance à des familles ne sont pas nécessairement un bon prédicteur de leur impact sur la plante, et l'utilisation de ce paramètre peut donc fausser les efforts de caractérisation fonctionnelle des effecteurs et de sélection des CEs prioritaires pour les études futures.

Mots-clés : *Melampsora larici-populina*; Transcriptomique; Métabolomique; Protéines effectrices; Champignon de la rouille; Immunité des plantes.

ABSTRACT

Rust fungi infect plants of all groups, from angiosperms to ferns, and they threaten the production of important crops, such as wheat, soy, and poplar. To better understand plant-pathogen interactions and elucidate what makes plants susceptible to these pathogens, we need to study effector proteins, which are used by the pathogen to control their hosts. *Melampsora larici-populina* (*Mlp*) causes the poplar rust and encodes at least 1 184 candidate effectors (CEs), however these proteins share little homology to proteins from organisms outside the Pucciniales order, preventing the prediction of their function. Previous studies selected CEs from *Mlp* for functional characterization in *Nicotiana benthamiana* or *Arabidopsis thaliana*. They showed that these proteins accumulate in various cellular compartments and impact the susceptibility of *Arabidopsis* to *Pseudomonas syringae* or to *Hyaloperonospora arabidopsidis*, which indicate they led to effector-triggered susceptibility.

The objective of this work was to further characterize the impact of the selected *Mlp* CEs in plants. For this, we performed RNA sequencing and mass spectrometry analyses of transgenic *Arabidopsis* plants constitutively expressing CEs tagged with GFP. To ensure the data was correctly normalized, we used reference genes for differential expression analysis. However, housekeeping genes and available pre-selected sets of references for *Arabidopsis* seedlings showed high expression variability in our samples. In addition, methods for selection of references are designed for RT-qPCR analyses or require pre-determined sets of potential references or target genes. Thus, we created a pipeline for selection of internal control genes that relies on read counts and gene sizes. This pipeline transforms read counts into Transcripts per Million (TPM), then calculates the minimum TPM value for genes to be considered expressed in each sample. The average minimal expression is calculated and used to exclude weakly expressed genes. Finally, the remaining genes with lowest covariance of TPM are selected as references.

Using this method for selection of reference genes for our transcriptomic data, we found a total of 2 299 genes deregulated across our 14 CE-expressing transgenic lines. Sets of co-expressed genes were calculated with weighted correlation network analysis and GO enrichment analysis showed that one gene set, formed by down-regulated genes in at least one transgenic line, was enriched in genes related to plant defense, indicating the CEs studied impacted plant immunity. The analysis of individual transgenic lines showed that the KEGG pathways “MAPK signaling pathway” and “Plant-pathogen interaction” were respectively over-represented in six and five of the 14 transgenic lines. Metabolomic analysis of these plants using ultra-high-resolution mass spectrometry led to the detection of 5 192 compounds in total, however 81.03% of these were not identified because of the lack of information in public databases, and 11.56% of the compounds detected matched multiple known metabolites, preventing their identification. By comparing metabolite concentrations in CE-expressing plants against control plants, we found a total of 680 metabolites deregulated with highly unsaturated and phenolic compounds enriched in down-regulated metabolites and peptides enriched among up-regulated metabolites.

Interestingly, both transcriptomic and metabolomic analysis showed that paralogous CEs did not deregulate the same genes or metabolites, while unrelated CEs had correlated patterns of gene and metabolite deregulation. Hence, our results suggest that similarity of effectors' sequences and their belonging to families may not be a good predictor of their impact on the plant, thus using this parameter may misguide efforts for functional characterization of effectors and the selection of priority CEs for future studies.

Key words: *Melampsora larici-populina*; Transcriptomics; Metabolomics; Effector proteins; Rust fungi; Plant immunity.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iv
RÉSUMÉ.....	vi
ABSTRACT	viii
LISTE OF TABLES.....	xiii
LISTE OF FIGURES.....	xiv
LIST OF ABBREVIATIONS AND ACRONYMS	xvii
CHAPTER I	
INTRODUCTION.....	1
1.1 Importance of plant diseases.....	1
1.2 Plant defenses	1
1.2.1 Pattern-Triggered Immunity	4
1.2.2 Effector-Triggered Susceptibility	6
1.2.3 Effector-Triggered Immunity	14
1.2.4 Interaction between PTI and ETI.....	19
1.2.5 ETI and necrotrophic pathogens	20
1.3 Effectoromics.....	21
1.4 Rust fungi	29
1.4.1 Life cycle	30
1.4.2 Infection process	32
1.5 Research objectives	34
CHAPTER II	
CUSTOM SELECTED REFERENCE GENES OUTPERFORM PRE-DEFINED REFERENCE GENES IN TRANSCRIPTOMIC ANALYSIS	38
2.1 Author contributions.....	38
2.2 Résumé de l'article	38
2.3 Full article in English: Custom selected reference genes outperform pre-defined reference genes in transcriptomic analysis.....	40
Abstract.....	40
Background.....	41

Results	43
Discussion.....	49
Conclusions	50
Methods	51
Availability of data and materials.....	52
Acknowledgements.....	52
References.....	53
Supplementary material	57

CHAPTER III

DIFFERENTIAL ALTERATION OF PLANT FUNCTIONS BY HOMOLOGOUS FUNGAL CANDIDATE EFFECTORS	59
3.1 Author contributions.....	59
3.2 Résumé de l'article	60
3.3 Full article in English: Differential alteration of plant functions by homologous fungal candidate effectors	62
Abstract.....	62
Introduction.....	63
Results	66
<i>In planta</i> expression of candidate fungal effectors results in important deregulation at the transcriptome level.....	66
Hierarchical clustering based on gene expression groups effectors independently of amino acid sequence homology	68
Effectors converge on deregulating the same metabolic pathways while others display unique patterns.....	74
Discussion.....	80
Materials and Methods	84
Plant growth conditions	84
RNA extraction and transcriptome analysis	85
Metabolite extraction and metabolomics analysis	86
Sequence analysis and integration	87
Data availability	88
Acknowledgements.....	88
References.....	89
Supplementary material	97

CHAPTER IV	
CONCLUSION.....	119
4.1 Conclusion	119
4.2 Perspectives	122
4.3 Final conclusions	124
REFERENCES.....	125
ANNEX A	
ADVANCES IN UNDERSTANDING OBLIGATE BIOTROPHY IN RUST FUNGI.....	146
ANNEX B	
A RUST FUNGAL EFFECTOR BINDS PLANT DNA AND MODULATES TRANSCRIPTION	164

LIST OF TABLES

Table	Page
 <u>Chapter II</u>	
1 Common reference genes used in this study for comparison against custom selected reference genes	43
 <u>Chapter III</u>	
1 Features of the CEs investigated in this study	66
2 Summary of “biological process” GO terms enriched in the WGCNA gene sets	70
S1 List of deregulated genes across the experiment with log ₂ -transformed fold changes (FC) and false discovery rates (FDR) for each transgenic line.....	109
S2 Percentage of identity and similarity, presented as “ID (SIM)”, calculated with pairwise sequence alignment of CEs using Needle	109
S3 Summary of metabolomic analysis in negative mode of extractions with 20% and 80% methanol combined	111
S4 Metabolites assigned and deregulated in each sample separated by category..	112
S5 Parameters used for bioinformatic analyses	116
S6 Sequencing results and alignment summary.....	117

LIST OF FIGURES

Figure	Page
1.1 Diagram of the primary (A) and secondary (B) plant cell walls	2
1.2 The zigzag model explains the dynamic interaction between plant immunity and pathogen attack	4
1.3 PAMP detection by RLK/RLP complexes	6
1.4 The <i>Moniliophthora perniciosa</i> MpChi is an inactive chitinase that binds to chitin	8
1.5 Subcellular localization of effectors or candidate effector proteins from several rust fungi.....	9
1.6 Proposed model of Brg1 l function.....	13
1.7 Modes of effector recognition by R proteins	15
1.8 Role of RIN4 and RPM1 in the plant- <i>Pseudomonas syringae</i> interaction	17
1.9 Model representing the role of Pto in plant resistance	18
1.10 Updated zig zag model showing that ETI amplifies PTI responses	20
1.11 Schematic overview of types III (right) and IV (left) secretion systems.....	22
1.12 Pipeline for selection of candidate effectors of <i>M. larici-populina</i> from list of small secreted proteins.....	25
1.13 A) Subcellular localization of <i>Mlp</i> CEs in <i>Arabidopsis</i> . B) 37347 accumulates in the plasmodesmata.....	28
1.14 Subcellular localization of <i>Mlp</i> CEs in <i>Nicotiana benthamiana</i>	29
1.15 Rust fungi life cycles	31
1.16 <i>M. larici-populina</i> 's life cycle.....	32
1.17 Representation of asexual cycle of <i>Puccinia</i> spp. on wheat.....	33
1.18 Expression of CE-encoding genes in <i>Arabidopsis thaliana</i> led to increased susceptibility to <i>H. arabidopsidis</i>	35
1.19 Impact of CEs (numbers) on plant susceptibility to bacterial and oomycete infection	36

Chapter II

1	Evaluation of covariance distribution in the three transcriptome data sets a) among a set of 14 commonly used reference genes and b) a set of 104 reference genes proposed by Zhuo <i>et al.</i> (2016).....	44
2	Comparison the four sets of reference genes in relation to covariance level and log ₂ TPM for A) Mlp37347 vs Control, B) Mlp124499 vs Control and C) Mlp124499 vs Mlp37347	46
3	Comparison of the four sets of reference genes in relation the distribution of log ₂ fold Change by -log ₁₀ adjusted p-value for A) Mlp37347 vs Control, B) Mlp124499 vs Control and C) Mlp124499 vs Mlp37347	47
4	Comparison of custom selected reference genes (blue border) and commonly used reference genes (red border) with geNorm ranking, NormFinder stability index and covariance for a) Mlp37347 vs Control, b) Mlp124499 vs Control and c) Mlp124499 vs Mlp37347.....	48
	Additional File 1. Covariance level for each of the 30 genes selected by Czechowski <i>et al.</i> (2005) for each permutation (A: Mlp37347 vs Control; B: Mlp124499 vs Control; C: Mlp124499 vs Mlp37347)	57

Chapter III

1	<i>In planta</i> expression of candidate fungal effector results in important deregulation at the transcriptome level.....	68
2	Heatmap of genes deregulated in each CE-expressing transgenic line	72
3	Hierarchical clustering of gene deregulation groups effectors independently of amino acid sequence homology	73
4	Effectors converge on deregulating the same metabolic pathways while others display unique patterns	75
5	Metabolic composition of samples in number of formulas (A) and relative abundance of compounds (B).....	77
6	(A) Metabolites down-regulated (left) are enriched in highly unsaturated and phenolic compounds while peptides are over-represented among those up-regulated (right).....	78
7	Hierarchical clustering based on metabolite deregulation groups effectors independently of amino acid sequence homology and gene deregulation patterns are not correlated to metabolite deregulation patterns in CE- expressing lines.....	80

S1	Magnitude of impact of CE on the plant's transcriptome is independent of its level of expression	97
S2	Magnitude of impact of CE on the plant's metabolome is independent of its level of expression, considering either the absolute number of deregulated metabolites (triangles, linear regression results in blue) or the ratio of metabolites deregulated by those identified (circles, linear regression results in red).....	98
S3	Principal component analysis of the replicates of 14 transgenic lines expressing candidate effectors from <i>Melampsora larici-populina</i> attached to GFP and a control line expressing only GFP (black dots).....	99
	File S1. Figure 1. Photosynthesis.....	100
	File S1. Figure 2. Cysteine and methionine metabolism	100
	File S1. Figure 3. Glutathione metabolism	101
	File S1. Figure 4. Starch and sucrose metabolism	102
	File S1. Figure 5. Glyoxylate and dicarboxylate metabolism.....	103
	File S1. Figure 6. Phenylpropanoid biosynthesis.....	104
	File S1. Figure 7. Ribosome.....	105
	File S1. Figure 8. MAPK signaling pathway	106
	File S1. Figure 9. Plant-hormone signal transduction.....	107
	File S1. Figure 10. Plant-pathogen interaction.....	108
	File S1. Figure 11. Circadian rhythm.....	109

LIST OF ABBREVIATIONS AND ACRONYMS

AAT	Amino Acid Transporter
ABRC	<i>Arabidopsis</i> Biological Resources Center
ACT2	Actin 2
ACT7	Actin 7
ACT8	Actin 8
ADC	Arginine Decarboxylase
APT1	Adenine Phosphoribosyltransferase 1
ARC	Apaf1, R protein and CED4
ARD1	NADH ⁺ dependent d-arabitol dehydrogenase
ARN	Acide Ribonucléique
Aux/IAA	Auxin/Indoleacetic Acid Proteins
Avr	Avirulence factor
BAK1	BRI1-Associated Kinase 1
BAX	BCL2 Associated X, Apoptosis Regulator
BED	Drosophila BEAF and DREF finger
BIK1	<i>Botrytis</i> -Induced Kinase 1
Brg11	hrpB-Regulated 11
BSMV	Barley Stripe Mosaic Virus
CC	Coiled-Coil domain
CEs	Candidate effectors
ChIP	Chromatin Immunoprecipitation
Cmu1	Chorismate mutase 1
coIP	co-Immunoprecipitation

CPL	CAZymes, Proteases and Lipases
CRN	Crinkler
CSEP	Candidate Secreted Effector Protein
DAFS	Data-Adaptive Flag Method for RNA-sequencing Data
DEG	Differentially Expressed Gene
DOC	Dissolved Organic Carbon
EC	Effector Candidate
EDC	mRNA Decapping Protein
EF1 α	Elongation Factor 1 α
EFR	Elongation Factor-Tu Receptor
EF-Tu	Elongation Factor-Tu
EGaseA	Endoglucanase A
eGFP	enhanced Green Fluorescent Protein
EHM	Extrahaustorial Matrix
eIF4A	Eukaryotic Translation Initiation Factor 4A-1
ET	Ethylene
ETI	Effector-Triggered Immunity
ETS	Effector-Triggered Susceptibility
FC	Fold Change
FLS2	Flagellin Sensitive 2
FPKM	Fragments per Kilobase per Million
GFP	Green Fluorescent Protein
GIP1	Glucanase Inhibitor Protein 1
GLK1p	Glucokinase 1 protein
GO	Gene Ontology

GUS	Beta Glucuronidase
HIGS	Host Induced Gene Silencing
HMC	Haustrorium Mother Cell
HR	Hypersensitive Response
HXT1	Hexose transporter 1
ID	Integrated Domain
INRA	Institut National de la Recherche Agronomique
INV1p	Invertase 1 protein
Isc	Isochorismate synthase
ISR	Induced Systemic Resistance
JA	Jasmonic Acid
KEGG	Kyoto Encyclopedia of Genes and Genomes
MAD1	Mannitol Dehydrogenase
MAPK	Mitogen-Activated Protein Kinase
MAPKKK	MAPK Kinase Kinase
MKK	MAPK Kinase
Mlp	<i>Melampsora larici-populina</i>
MoHTR	<i>M. oryzae</i> Host Transcription Reprogramming
MPa	mega Pascal
mRNA	messenger Ribonucleic Acid
MS	Mass Spectrometry
NBS-LRR	Nucleotide-Binding-Site Leucine-Rich Repeat
NDUFA8	Nicotinamide Adenine Dinucleotide-Ubiquinone Oxidoreductase 19kDa subunit
NGS	Next-Generation Sequencing
NMR	Nuclear Magnetic Resonance

NPR1	Non-expressor of Pathogenesis Related 1
OPT	Oligopeptide Transporter
P/MAMPs	Pathogen/Microbe-Associated Molecular Patterns
Pgt	<i>Puccinia graminis</i> f.sp. <i>tritici</i>
PI3P	Phosphatidylinositol-3-Phosphate
PMA1p	Plasma Membrane H ⁺ -ATPase 1 protein
PNPi	<i>Puccinia</i> NPR1 interactor
PR	Pathogenesis-Related proteins
PRPs	Proline-Rich cell wall structural Proteins
PRR	Pattern-Recognition Receptors
PstSCR1	<i>P. striiformis</i> f. sp. <i>tritici</i> Small Cysteine-Rich candidate effector
PsXEG1	<i>P. sojae</i> xyloglucan endoglucanase 1
PsXLP1	PsXEG1-like protein 1
PTI	Pattern-Triggered Immunity
PWL	Prevents pathogenicity toward Weeping Lovegrass
qPCR	quantitative Polymerase Chain Reaction
R protein	Resistance protein
RBOHD	Respiratory Burst Oxidase Homologue protein D
RIN4	RPM1-interacting protein 4
RLK	Receptor-Like Kinases
RLP	Receptor-Like Proteins
RNA	Ribonucleic Acid
RNAseq	RNA sequencing
ROS	Reactive Oxygen Species
RPKM	Reads per Kilobase per Million

RPM1	Resistance to <i>P. syringae</i> pv. Maculicola 1
RPS2	Resistance to <i>Pseudomonas syringae</i> 2
RPS4	Resistance to <i>P. syringae</i> 4
RRS1	Resistant to <i>R. solanacearum</i> 1
RTP	Rust-transferred proteins
RT-qPCR	Reverse Transcription quantitative Polymerase Chain Reaction
RXLR-dEER	Arg-X-Leu-Arg and Asp-Glu-Glu-Arg
SA	Salicylic Acid
SAR	Systemic Acquired Resistance
SCF	Skp, Cullin, F-box
See1	Seedling Efficient Effector 1
SERK	Somatic Embryogenesis Receptor Kinases
SGT1	Suppressor of G2 allele of <i>skp1</i>
SNP	Single Nucleotide Polymorphism
SOBIR1	Suppressor Of BIR1-1
SRFA	Suwannee River Fulvic Acid
SSPs	Small Secreted Proteins
SWEET	Sugar Will Eventually be Exported Transporters
T3SS	Type Three Secretion System
TAL	Transcription Activator-Like
TGA2.2	TGA transcription factor 2.2
TIR	Toll/Interleukin-1 receptor/Resistance protein
TPM	Transcripts per Million
TTK1	Tin2-targeting kinase 1
TUB2	Tubulin β -2/ β -3 chain

TUB6	β -Tubulin 6
TUB9	Tubulin β -9 chain
UBQ10	Polyubiquitin10
UBQ11	Polyubiquitin 11
UBQ4	Polyubiquitin 4
UBQ5	Ubiquitin Extension Protein 5
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
WGCNA	Weighted Correlation Network Analysis

CHAPTER I

INTRODUCTION

1.1 Importance of plant diseases

Plants, as any other organism, face threats in the form of pests and pathogens. This includes viruses (Scholthof *et al.*, 2011), bacteria (Mansfield *et al.*, 2012), oomycetes (Kamoun *et al.*, 2015), fungi (Dean *et al.*, 2012), nematodes (Jones *et al.*, 2013), insects (Oliveira *et al.*, 2014) and even other plants (Strange & Scott, 2005). The impact of these pathogens and pests on food security, economy, and our societies can be extreme, as in the case of the Irish Potato Famine in the 1840s, when a million people died and another 1.5 million emigrated Ireland due to the potato blight epidemic, caused by the oomycete *Phytophthora infestans* (Strange, 1993). Another example is the Bengal Famine of 1943, which was caused by the brown spot disease in rice (*Cochliobolus miyabeanus*) and led to the death of 2 million people (Strange, 1993). To control plant diseases, we need to combine many approaches, including the use of pesticide, resistant plants, biological control agents and the combat of virulence mechanisms of the pathogens (Strange & Scott, 2005). Hence, it is imperative to decipher how pathogens cause diseases and how the plants can become resistant to pathogens if we are to prevent other plant epidemic disasters.

1.2 Plant defenses

To be able to feed on a plant, pathogens must pass through physical and chemical barriers. The surface of aerial plant organs, including leaves and stems, is covered with the cuticle, that faces the outside, and cutin, facing the plant epidermis (Yeats & Rose, 2013) while the roots are covered with suberin (Strange, 1993). While the cutin is made of fatty acids, suberin is composed of phenolic compounds and monobasic, hydroxylated, and dibasic acids. The epidermal cells are under the cutin/suberin and they are protected

by cell walls. Growing epidermal cells only have primary cell walls (**Figure 1.1A**), made of cellulose (a polymer of β -1-4-glucan) and extensin, a glycoprotein (Strange, 1993). Once the cells stop growing, they produce another layer of cell wall, which lies between the primary cell wall and the plasma membrane. The secondary cell wall (**Figure 1.1B**) is more rigid than the primary cell wall, because it contains lignin, a polymer of coumaryl, coniferyl, and sinapyl alcohols (Kang *et al.*, 2019; Strange, 1993). These structures confer major physical barriers, but if the pathogen finds a way to enter, through stomata or open wounds, it still must fight the plant's chemical defenses.

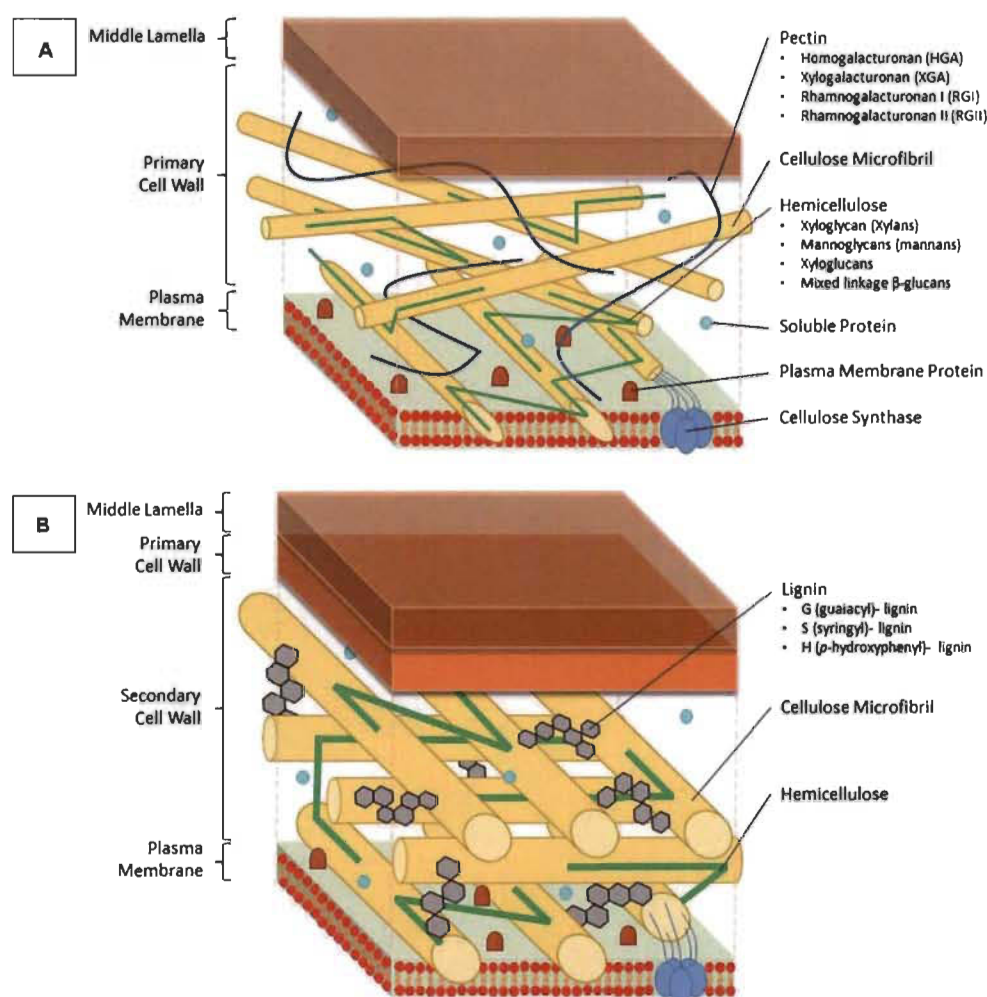


Figure 1.1. Diagram of the primary (A) and secondary (B) plant cell walls.

Their main difference is the presence of lignin in the secondary cell wall, replacing the pectin. The cellulose microfibrils in the secondary cell wall are stacked in a more organized manner compared to the primary cell wall. Source: Loix *et al.* (2017).

Plants constitutively accumulate specialized metabolites that confer resistance against pathogens. These are known as phytoanticipins (VanEtten *et al.*, 1994) and they can be phenolic compounds, quinones, unsaturated lactones, cyanogenic glucosides, saponins, terpenoids, or stilbenes (Strange, 1993). For instance, some onion cultivars produce two types of phenolic compounds: catechol and protocatechuic acid. These metabolites are anti-fungal, conferring resistance against *Colletotrichum circinans* (Clark & Lorbeer, 1975). Aliphatic glucosinolates are phytoanticipins produced by *Arabidopsis* plants. These compounds limit herbivory and participate in non-host resistance against certain *Pseudomonas syringae* pathovars (Fan *et al.*, 2011).

If the invading microorganism can pass through and survive these constitutive defenses, it then faces the plant immune system. Plants do not have mobile immune cells, thus each plant cell has to defend itself (Jones & Dangl, 2006). The dynamics of the interaction between the plant immune system and pathogen attacks is described in the Zigzag model proposed by Jones & Dangl (2006) (**Figure 1.2**). The first tier of plant immunity is activated by Pathogen/Microbe-Associated Molecular Patterns (P/MAMPs). This defense is known as Pattern-Triggered Immunity (PTI), and may include plant cell death, transient production of Reactive Oxygen Species (ROS) and activation of Mitogen-Activated Protein Kinase (MAPK) cascades, among other mechanisms. Certain adapted pathogens can secrete molecules (effectors) into the host to deactivate this immune response. This is known as Effector-Triggered Susceptibility (ETS), and it allows adapted pathogens to grow and reproduce in the host. Plant cultivars/ecotypes may have Resistance (R) proteins able to recognize directly or indirectly specific effectors, inducing Effector-Triggered Immunity (ETI), which blocks pathogen growth. In these cases, the effector is known as an Avirulence factor (Avr), since its presence makes the pathogen avirulent in the resistant host. ETI is robust and prolonged in relation to PTI, but these two tiers of defense use very similar gene sets and hormonal pathways (Tsuda & Katagiri, 2010). Through mutation and selection, new pathogen races may appear with mutated *Avr* genes, no longer recognized by the R proteins, or new effectors that disable the previous ETI, leading to ETS. On the plant side, evolution may result in changes in

R genes, resulting in proteins able to detect new versions of effectors, causing again ETI. The steps of the zigzag model will be further discussed below.

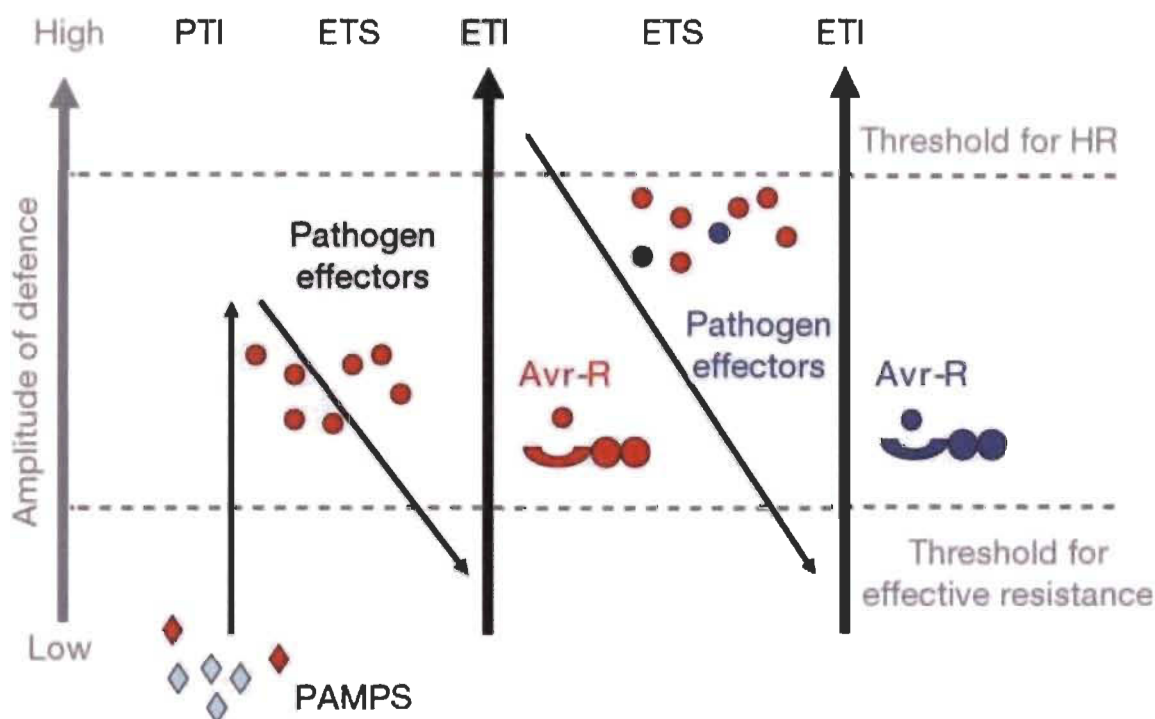


Figure 1.2. The zigzag model explains the dynamic interaction between plant immunity and pathogen attack.

Microorganisms liberate PAMPs which may activate PTI, a transient and weak defense response. Adapted pathogens can counterattack by secreting effectors, leading to ETS. In some cases, plants may be able to detect these effectors (Avirulence factors) with R proteins, which will then trigger ETI. ETI is stronger, more prolonged and robust than PTI and may cause localized cell death. Evolution may lead to the appearance of pathogen races with new effectors that disable the ETI response or new versions of the Avr that are no longer recognized by the respective R protein, leading to ETS. This process may then select for plant varieties with R proteins able to detect the new effectors, causing again ETI. Source: Jones & Dangl (2006).

1.2.1 Pattern-Triggered Immunity

PAMPs are molecules, relatively conserved across big groups of microorganisms (pathogens or not), which may induce defense responses in both plants and animals (Nurnberger & Kemmerling, 2009). Chitin (from fungi), flagellin (from Gram-negative bacteria) and beta-glucans (from oomycetes) are all examples of PAMPs (Nurnberger & Kemmerling, 2009). These molecules are recognized by Pattern-Recognition Receptors (PRRs) at the plant cell surface. PRRs can be Receptor-Like Kinases (RLKs, enzymes

with a receptor domain to the outside of the cell, a transmembrane domain; and a kinase domain on the inside) or Receptor-Like Proteins (RLPs, proteins similar to RLKs, but without a catalytic domain inside the cell) (Buendia *et al.*, 2018; Fritz-Laylin *et al.*, 2005; Nurnberger & Kemmerling, 2009).

When RLKs bind to their ligands, they will often heterodimerize to form complexes with co-receptors (such as Somatic Embryogenesis Receptor Kinases, SERKs). For instance, the SERK BRI1-Associated Kinase 1 (BAK1) associates with Flagellin Sensitive 2 (FLS2) and with Elongation Factor-Tu (EF-Tu) Receptor (EFR) in *Arabidopsis* (Couto & Zipfel, 2016). Similarly, RLPs, lacking cytoplasmic kinase domains, must interact with a RLK (for instance the Suppressor Of BIR1-1, SOBIR1), which, after ligand binding, recruits SERKs or similar co-receptors (Jamieson *et al.*, 2018). The complex, formed of RLK-co-receptor or RLP-RLK-co-receptor, activates Botrytis-Induced Kinase 1 (BIK1) and related proteins (**Figure 1.3**). BIK1 in turn activates Ca^{2+} and other ion channels and phosphorylates Respiratory Burst Oxidase Homologue protein D (RBOHD), which produces ROS (Couto & Zipfel, 2016). The accumulation of ROS in the apoplast leads to cross-linking of Proline-Rich cell wall structural Proteins (PRPs) and induction of lignin biosynthesis, which makes the cell wall less easy to penetrate. It also induces the biosynthesis of phytoalexins (defense molecules produced and accumulated in response to an infection) and deployment of hydrolytic enzymes (Dixon *et al.*, 1994). RLK-co-receptor and RLP-RLK-co-receptor complexes also activate MAPK cascades, that phosphorylate transcription factors, transducing the cell surface signals to the nucleus. These signaling cascades lead to production of toxic molecules (phytoalexins), enzymes (with anti-fungal, anti-oomycete and anti-bacterial activities, such as Pathogenesis-Related (PR) proteins and defensins), and reinforcement of physical barriers (Luna *et al.*, 2011).

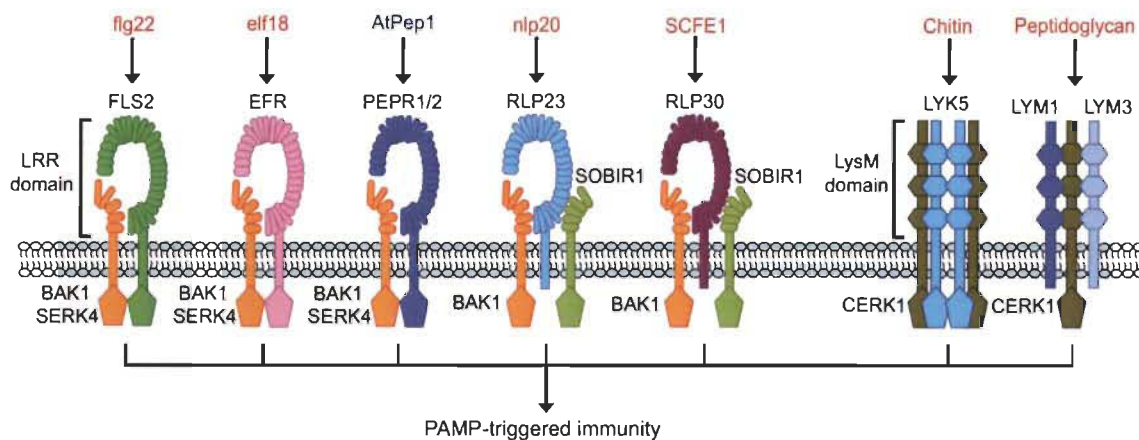


Figure 1.3. PAMP detection by RLK/RLP complexes.

Names in red indicate the recognized PAMP. *AtPep1* is a Damage-Associated Molecular Pattern (DAMP), it is a host peptide originated from the degradation of PROPEP1 and induces PTI (Yamaguchi *et al.*, 2006). Source: He *et al.* (2018).

1.2.2 Effector-Triggered Susceptibility

PTI prevents plants from succumbing to each organism attempting to infect them. Yet, adapted pathogens can escape recognition or block the activation of defense responses. This is because they secrete molecules into the host that modulate the host functions to facilitate feeding and reproduction and suppress or evade host immunity. These molecules are known as effectors (Hogenhout *et al.*, 2009) and this phenomenon is termed ETS. Effectors have been detected in viruses (Leisner & Schoelz, 2018), bacteria (Büttner, 2016), oomycetes (Stassen & Van den Ackerveken, 2011), fungi (Lo Presti *et al.*, 2015), insects (Rodriguez & Bos, 2013) and nematodes (Mejias *et al.*, 2019).

Effectors are molecules secreted into the host; therefore, they may act in the apoplast or inside the plant cell. Plant-interacting fungi use effectors to protect their cell wall from plant chitinases, or to bind chitin and prevent its detection by PRRs (Han *et al.*, 2019; Mentlak *et al.*, 2012; Van Esse *et al.*, 2007; Volk *et al.*, 2019). The fungus *Moniliophthora perniciosa* (causal agent of witches' broom disease from cocoa trees) uses a chitinase-like protein, MpChi, to bind to chitin in the apoplast (Fiorin *et al.*, 2018). By sequestering chitin, MpChi prevents this PAMP from binding to plant chitin receptors, which would lead to activation of plant defense (**Figure 1.4**). *MpChi* is expressed exclusively during the biotrophic stage of the infection process and, although its catalytic site is inactive,

it can bind to chitin and it binds weakly to chitosan (Fiorin *et al.*, 2018). Treatment of *Nicotiana tabacum* cells with MpChi and chitin oligomers attenuated medium alkalisation and led to fewer deregulated genes in comparison to *N. tabacum* cells treated only with chitin oligomers (Fiorin *et al.*, 2018). Another example is the necrotrophic wheat pathogen *Zymoseptoria tritici*, which has three LysM effector proteins (Mg1LysM, Mg1xLysM, and Mg3LysM) capable of binding to chitin. These effectors can protect the beneficial fungus *Trichoderma viridae* hyphae from plant chitinases and are necessary for full virulence of this pathogen (Marshall *et al.*, 2011; Tian *et al.*, 2020). Wheat plants infected with *Z. tritici* strains lacking Mg3LysM or Mg1xLysM had less fungal biomass than plants infected with the wild type fungus. Nevertheless, only plants infected with Mg3LysM deletion *Z. tritici* strains had smaller necrotic areas (Tian *et al.*, 2020) and Marshall *et al.* (2011) also showed that infection with this strain resulted in higher expression of wheat defense genes *TaPR1* and *TaChitinase* compared to infection with wild type strains. Oomycetes do not have chitin, but their cell wall is composed of β -glucans that are targeted by plant apoplastic endoglucanases. The Endoglucanase A (EGaseA) from soybean is an apoplastic enzyme that interacts with oomycete cell wall, liberating glucans that act as elicitors of plant defense. However, the plant pathogen oomycete *Phytophthora sojae* effector Glucanase Inhibitor Protein 1 (GIP1) inhibits EGaseA, which prevents immune activation (Rose *et al.*, 2002).

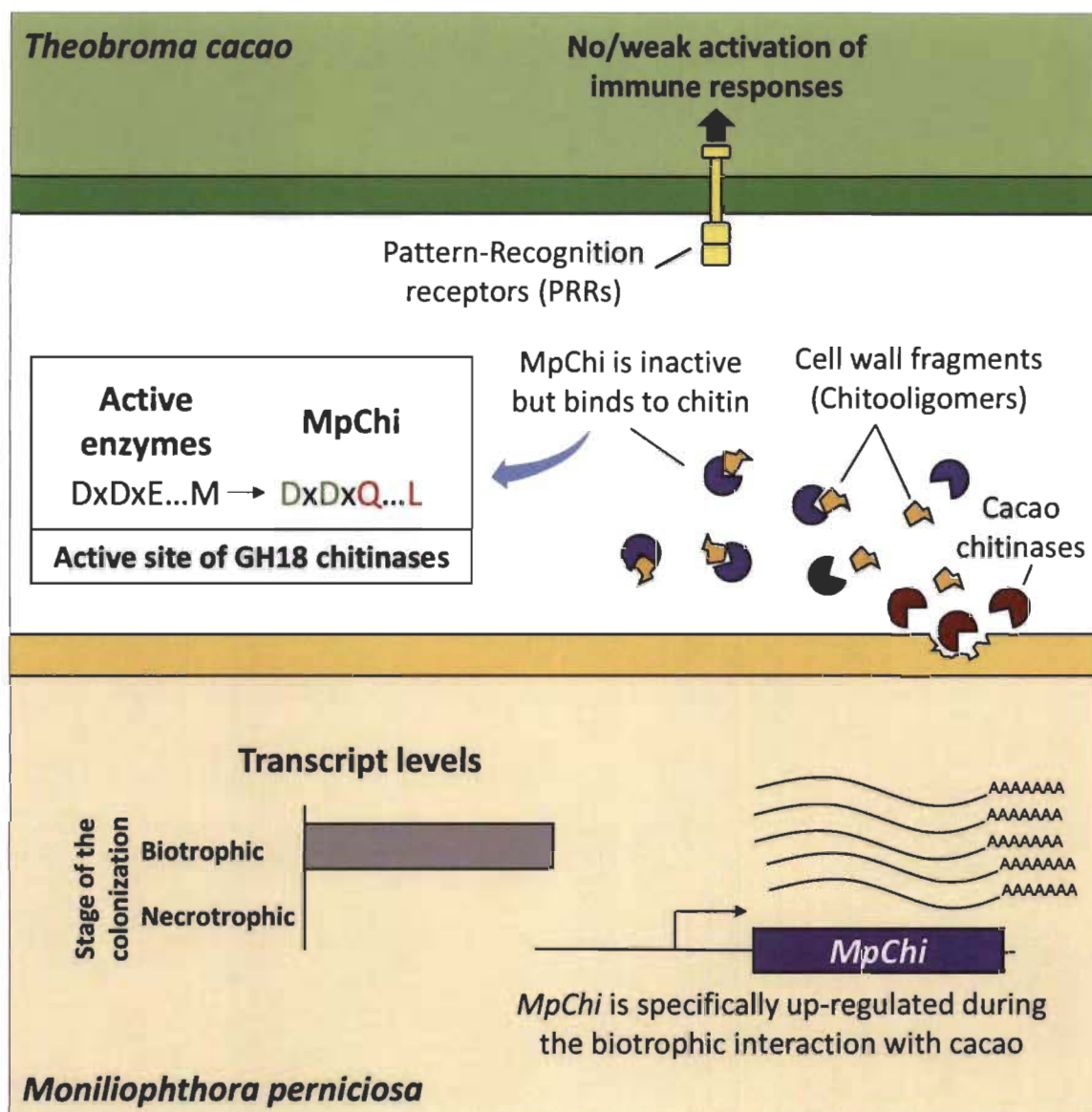


Figure 1.4. The *Moniliophthora perniciosa* MpChi is an inactive chitinase that binds to chitin. MpChi is expressed during the biotrophic phase of the infection, sequestering fungal cell wall fragments, which prevents activation of PTI in the plant. Source: Fiorin *et al.* (2018).

There are apoplastic effectors that do not play a role in protection of the pathogen, this is the case of *P. sojae*'s effector *P. sojae* xyloglucan endoglucanase 1 (PsXEG1) (Ma *et al.*, 2015). This effector hydrolyses xyloglucan and β -glucan to form reducing sugars. It is still unknown whether the oomycete uses these sugars as nutrients or if PsXEG1 is used to breach the host cell wall, or both. However, PsXEG1's activity is inhibited by the soybean apoplastic protein GmGIP1 (Ma *et al.*, 2017). To counter-attack, *P. sojae* secretes a paralog of PsXEG1, *P. sojae* XEG1-like protein 1 (PsXLP1), which has no

hydrolase activity but binds more strongly to GmGIP1 (Ma *et al.*, 2017). In the absence of GmGIP1, PsXEG1 contributes to soybean's susceptibility to *P. sojae* without interacting with the plant immunity, and the decoy PsXLP1 recovers PsXEG1's virulence in the presence of GmGIP1 (Ma *et al.*, 2017).

Effectors have been found inside the plant cell in multiple cellular compartments (**Figure 1.5**). For instance, fungal effectors and Candidate Effectors (CEs) target chloroplasts (Germain *et al.*, 2018; Petre *et al.*, 2015; Tang *et al.*, 2020; Xu *et al.*, 2019), cytosol (Djamei *et al.*, 2011a; Germain *et al.*, 2018; Petre *et al.*, 2015; Tanaka *et al.*, 2014), endoplasmic reticulum (Robin *et al.*, 2018), Golgi bodies (Robin *et al.*, 2018), microtubules (Robin *et al.*, 2018), mitochondria (Petre *et al.*, 2015), nucleus (Petre *et al.*, 2015) and nucleolus (Ahmed *et al.*, 2018; Robin *et al.*, 2018), plasmodesmata (Germain *et al.*, 2018), peroxisomes (Robin *et al.*, 2018) and tonoplast (Madina *et al.*, 2020).

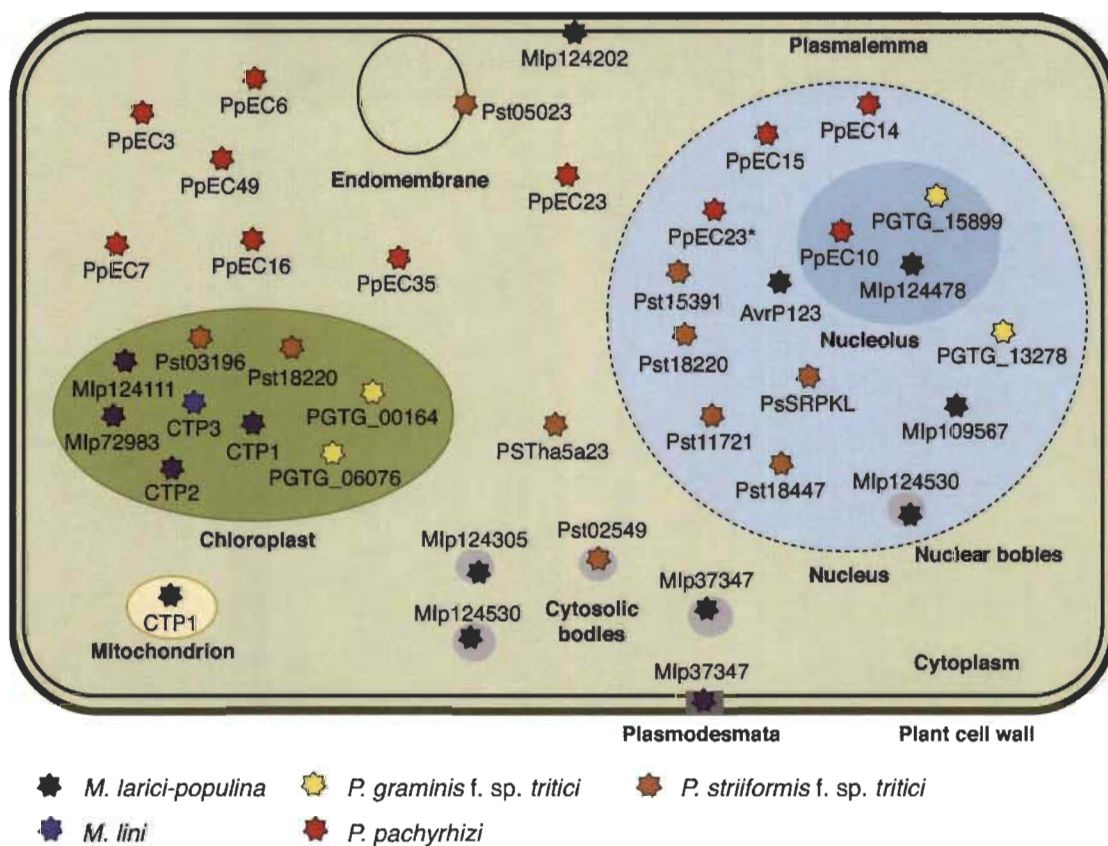


Figure 1.5. Subcellular localization of effectors or candidate effector proteins from several rust fungi.

Source: Lorrain *et al.* (2018b).

There are several effectors that target the host plasma membrane and are involved with plant immunity or other processes. *Pseudomonas syringae* pathovars have several effectors that localize to the host plasma membrane and interact with Resistance to *P. syringae* pv. Maculicola 1 (RPM1)-interacting protein 4 (RIN4) (Ray *et al.*, 2019), including AvrB, AvrRpt2 and HopF2. AvrB interacts with and phosphorylates RIN4 (Mackey *et al.*, 2002), but also it interacts with HSP90, RAR1, and MAPK4 (Cui *et al.*, 2010), leading to phosphorylation of MAPK4 and inducing jasmonic acid (JA)-mediated defense responses (Cui *et al.*, 2010). It is possible that AvrB uses the antagonism of JA and salicylic acid (SA) mediated responses to promote host susceptibility, since *P. syringae* secretes coronatine, an analogue of JA-Ile (functional form of JA) to suppress SA-mediated responses (Geng *et al.*, 2014). AvrRpt2 cleaves RIN4 and induces turn-over of Auxin/Indoleacetic acid (Aux/IAA) proteins, which are repressors of auxin-responsive genes (Cui *et al.*, 2013). Auxin was shown to repress SA-mediated responses and susceptible *AvrRpt2*-expressing *Arabidopsis* plants infected with *P. syringae* pv. *tomato* had lower expression of pathogenesis-related (PR) genes induced by SA (Chen *et al.*, 2004). Thus, the impact of these two effectors, AvrB and AvrRpt2, is similar: suppression of SA-mediated responses to promote host susceptibility. However, other responses related to auxin may be beneficial for the pathogen, such as cell wall loosening (Perrot-Rechenmann, 2010), which could facilitate obtention of nutrients by the bacteria, its induction of ethylene (ET) biosynthesis (Abel *et al.*, 1995), a hormone that causes increased susceptibility to *P. syringae* and *Xanthomonas campestris* pv. *campestris* (Bent *et al.*, 1992), or stomatal opening (Irving *et al.*, 1992), that could facilitate the dispersion of the pathogen. Finally, HopF2 has multiple functions. It interacts with BAK1, preventing BAK1-phosphorylation of BIK1 (Wu *et al.*, 2011), and it ADP-ribosylates MAPK kinase 5 (MKK5), attenuating MPK3 and MPK6 phosphorylation (Wang *et al.*, 2010), suppressing PTI. HopF2 also interacts with RIN4, preventing AvrRpt2-cleavage of RIN4 (Wilton *et al.*, 2010), and forms a complex with RIN4 and the P-type ATPase2, preventing pathogen-induced stomatal closure (Hurley *et al.*, 2014). Thus, HopF2 promotes susceptibility by suppressing PTI and preventing ETI.

The fungus *Ustilago maydis* causes corn smut and is a model organism for the study of biotrophic fungal plant pathogens (Dean *et al.*, 2012). *U. maydis* has three well-characterized effectors that target the host cytosol: Chorismate mutase 1 (Cmul), Tin2 and Seedling efficient effector 1 (See1). The accumulation of Cmul in the plant's cytosol was corroborated by infecting plants with a Cmul-knockout *U. maydis* strain complemented with Cmul-human influenza hemagglutinin and performing immunolocalization (Djamei *et al.*, 2011a). Also, by transiently expressing Cmul tagged with mCherry and without signal peptide in maize, it was shown that Cmul can spread to other cells using the plasmodesmata (Djamei *et al.*, 2011a), although in this case the fluorescence was also seen in the plant nuclei. Cmul converts chorismate into prephenate, which is used for the synthesis of phenylalanine (Tounekti *et al.*, 2013). Thus, Cmul reduces the pool of chorismate available in the chloroplast for the synthesis of SA, attenuating defense responses (Djamei *et al.*, 2011a). In relation to Tin2, doubts remain with regards to its precise localisation. This is because the C-terminal residues of this protein are crucial for its function: Tin2 mutant protein with the five last residues deleted and the mutant protein with five added alanine residues to the C-terminus are not able to complement the Δ tin2 mutant fungus (Tanaka *et al.*, 2014). Thus, the addition of tag fluorescent proteins or polyhistidine tag renders Tin2 not functional. Infection of maize with transgenic *U. maydis* carrying the construct Tin2₁₋₂₅-mCherry-Tin2₂₆₋₂₀₇ (Tin2₁₋₂₅ is Tin2 signal peptide) results in accumulation of mCherry in the apoplast, whereas transgenic maize carrying the construct mCherry-Tin2₂₆₋₂₀₇ shows nucleocytosolic fluorescence (Tanaka *et al.*, 2014), which is the subcellular localization of free mCherry. Nevertheless, the cytosolic localization of this effector is supported by its interaction with *Zea mays* Tin2-targeting kinase 1 (TTK1), a cytoplasmic protein. Tin2 stabilizes TTK1, leading to the induction of anthocyanin biosynthesis genes, increased accumulation of anthocyanins, and the suppression of lignin biosynthesis genes (Tanaka *et al.*, 2014). Finally, See1 is a nucleocytosolic effector that interacts with Suppressor of G2 allele of skp1 (SGT1) (Redkar *et al.*, 2015), a protein that is part of two complexes: Heat-shock protein 90 (HPS90)-RAR1-SGT1, essential in NLR-mediated immune responses, and Skp, Cullin, F-box (SCF) E3 ubiquitin ligase, involved in the degradation of cell cycle-related proteins. By interacting with SGT1, See1 prevents the phosphorylation of this

protein and maintains it in the nucleus, where SGT1 degrades cell cycle-related proteins, inducing DNA synthesis, leading to tumor formation (Villajuana-Bonequi *et al.*, 2019).

Although the effectors described above modulate the host's transcriptome, this is done indirectly since they do not interact with the host DNA. However, there are effectors able to directly modulate the host transcriptome. Transcription Activator-Like (TAL) effectors are bacterial proteins that induce host gene expression. *Arabidopsis* plants infected with *P. syringae* pv. *tomato* DC3000 showed increased expression of genes encoding sugar transporter (Sugar Will Eventually be Exported Transporters, SWEET), and this induction in expression was dependent on the type three secretion system (T3SS, the mechanism by which effectors are secreted from the bacterial cell into the host cell, discussed in detail in section 1.3), indicating that effectors manipulate the expression of SWEET genes in *Arabidopsis* (Chen *et al.*, 2010). This strategy is also used by *Xanthomonas* spp. The TAL effector PthXo1 from *X. oryzae* pv. *oryzae* activates the expression of the rice gene *OsSWEET11* (Yang *et al.*, 2006), while *X. axonopodis* pv. *manihotis* uses TAL20 to induce the expression of the cassava gene *MeSWEET10a* (Cohn *et al.*, 2014). These TAL effectors are used by the pathogens to modulate the plant transcriptome to increase the concentration of sugars in the apoplast, an indirect modulation of the metabolome, making them more accessible to the bacteria. Finally, pathogens can use effectors to modulate the host to inhibit growth of competitors. The TAL effector hrpB-regulated 11 (Brg11), from *Ralstonia solanacearum*, induces expression of Arginine Decarboxylase (ADC)-encoding gene, which is responsible for the synthesis of polyamines, including agmatine and putrescine (Wu *et al.*, 2019). This leads to increased resistance to *P. syringae* in tomato and *Arabidopsis* (Wu *et al.*, 2019). Thus, Brg11 is used by *R. solanacearum* to modulate the plant's metabolome and make it inhospitable to *R. solanacearum* competitors (**Figure 1.6**).

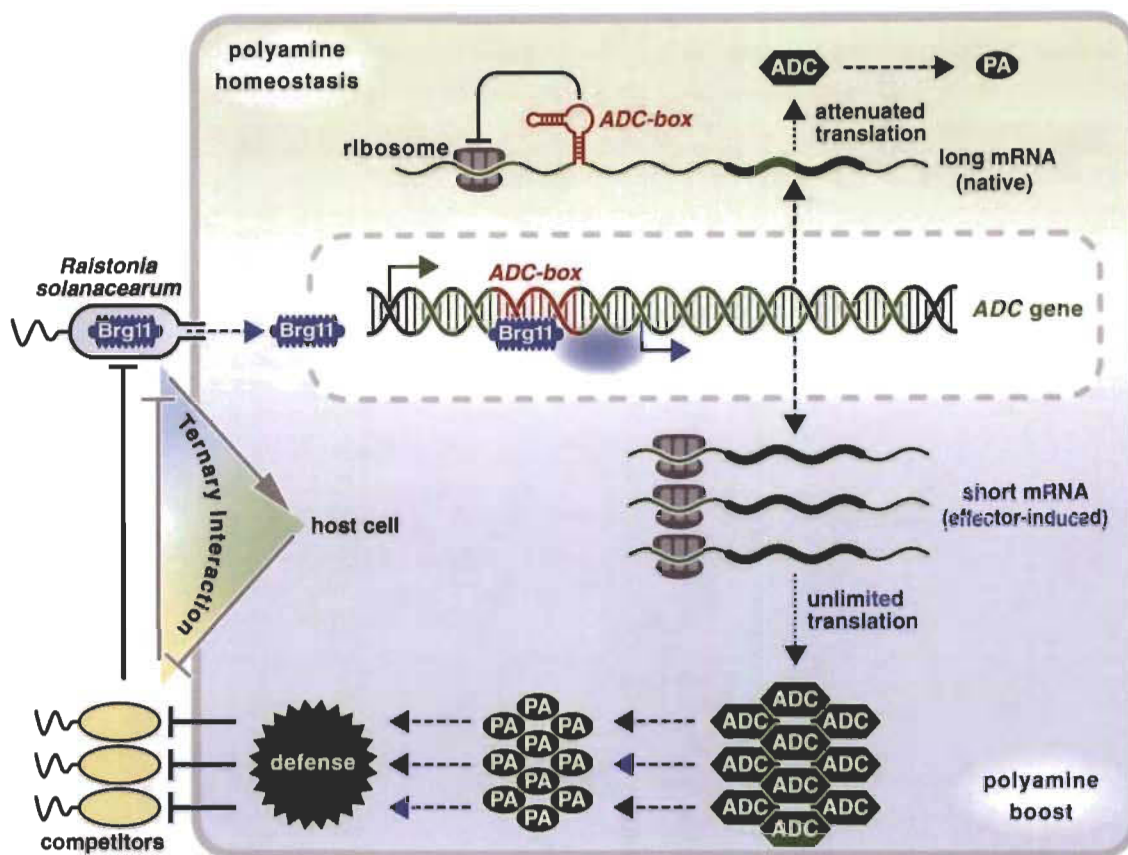


Figure 1.6. Proposed model of Brg11 function.

Brg11 induces expression of short mRNA of ADC, which has no attenuation in translation. This leads to increased production of polyamines, which block growth of *Ralstonia solanacearum*'s competitor bacteria. PA, polyamines. ADC, arginine decarboxylase. Source: Wu *et al.* (2019).

Bacteria are not the only plant pathogens able to modulate the host transcriptome by interacting with its DNA. CgEP1 is an effector from *Colletotrichum graminicola*, a fungus that causes anthracnose in maize, and it is necessary for development of the disease in this plant (Vargas *et al.*, 2016). Once in the host cell, CgEP1 is translocated into the nucleus and is able to bind maize DNA. Using chromatin immunoprecipitation (ChIP), Vargas *et al.* (2016) found 58 possible gene targets of CgEP1, including a MAPKKK, a protein phosphatase and a proteasome subunit, indicating that this effector may bind DNA to modulate the expression of these genes. Another example is the rust fungal CE Mlp124478 from the poplar rust *Melampsora larici-populina*. Transient expression of Mlp124478, without its signal peptide, tagged with green fluorescent protein (GFP) in *N. benthamiana* and constitutive expression of the same construct in *Arabidopsis* showed that this protein accumulates in the host cytosol, nucleoplasm and nucleolus (Ahmed *et al.*, 2018;

Petre *et al.*, 2015). Expression of *Mlp124478* in *Arabidopsis* induced plant susceptibility to *Hyaloperonospora arabidopsidis* infection, but this function did not depend on the nucleolar localization, and resulted in down-regulation of genes involved with response to viruses and bacteria, response to brassinosteroid, cell wall organization and response to ethylene (Ahmed *et al.*, 2018). With ChIP-PCR, the authors showed that this CE binds TGA1a transcription factor-binding site from one of the most up-regulated genes in *Arabidopsis* expressing *Mlp124478*, suggesting this CE modulates the plant's transcriptome by binding TGA1a-binding site (Ahmed *et al.*, 2018). Finally, two rice blast fungus effectors, *M. oryzae* Host Transcription Reprogramming 1 (MoHTR1) and MoHTR2, accumulate in rice nuclei and bind to the promoter region of their target genes, related to plant immunity, and repress their expression, making the plant more susceptible to hemibiotrophic and less susceptible to necrotrophic pathogens (Kim *et al.*, 2020).

1.2.3 Effector-Triggered Immunity

Plants sense pathogens and their effectors by using R proteins. The most common class of R proteins is the Nucleotide-Binding-Site Leucine-Rich Repeat (NBS-LRR). They may recognize the presence of some effectors in the plant cell and activate defense responses that are considered stronger and more robust than PTI, known as ETI (Tsuda *et al.*, 2009). There are several different mechanisms by which R proteins recognize effectors (**Figure 1.7**): direct recognition, guard model, decoy model, and through integrated domains.

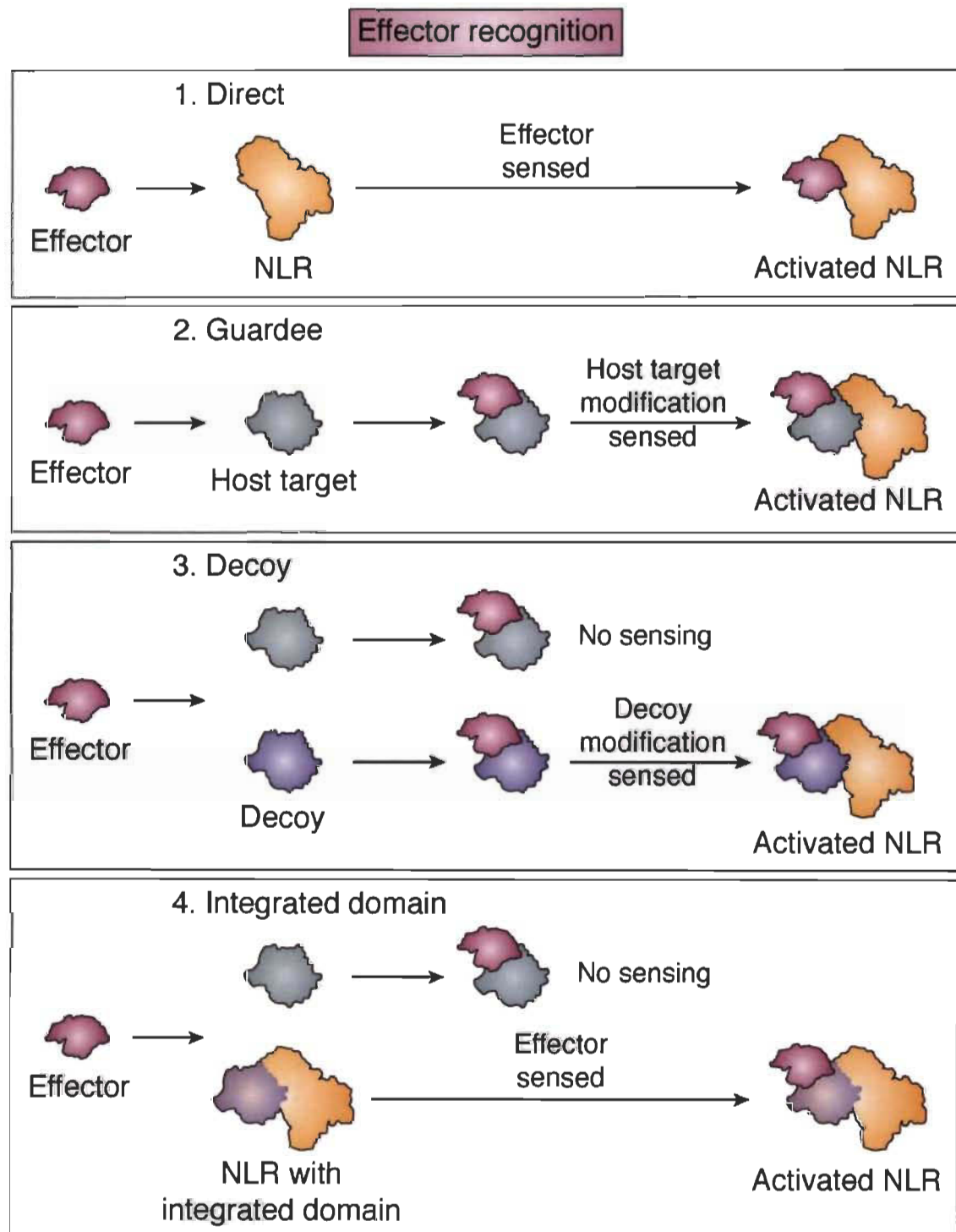


Figure 1.7. Modes of effector recognition by R proteins.

Direct: effector binds to R protein. Guard: effector modifies a target protein in the host, modification is detected by R protein. Decoy: effector modifies a host protein (the decoy) that mimics its target, modification is detected by R protein. Integrated decoy: effector interacts with R protein domain that is similar to its target, which activates the R protein. Source: Bentham *et al.* (2020).

The AvrL567 proteins are recognized through direct interaction with the R proteins L5, L6 and/or L7 (Bernoux *et al.*, 2011; Dodds *et al.*, 2006). Transient expression of these effector variants in flax plants containing either *L5*, *L6*, or *L7* genes led to different symptoms. On L5 and L6 plants, AvrL567-A, F, L, and J caused necrotic responses, but only chlorosis in L7 plants (Dodds *et al.*, 2006), indicating these effectors interact with the R proteins *in planta*, albeit weakly with L7. AvrL567-D is the only variant that interacts strongly with L7, causing necrosis in L6 and L7 plants (Dodds *et al.*, 2006). Mutations in the effector may prevent recognition by an R protein without impacting its virulence function. Besides, effector-encoding genes evolve fast, as they are typically found in gene-sparse regions of the genome (Dong *et al.*, 2015): telomeres, associated with transposable elements or in AT-rich regions (Sanchez-Vallet, 2018). Thus, pathogens may fast overcome resistance derived by direct detection of an effector (van der Hoorn & Kamoun, 2008).

R proteins may also detect effectors indirectly, by guarding their target proteins (**Figure 1.8**). This defense mechanism exerts two opposing selective pressures on effector-encoding genes, because a mutation that diminish the effector-target interaction will render the pathogen less virulent, but it will also reduce or prevent activation of host defense responses (Cesari, 2018; Hoorn *et al.*, 2002). As effectors converge on targeting key proteins (Mukhtar *et al.*, 2011), plants use fewer R proteins to protect themselves against multiple effectors if they guard the targets, instead of recognizing the effectors directly. For example, AvrB and AvrRpm1 both target RIN4 and the result of this interaction is plant susceptibility (Ray *et al.*, 2019). Nevertheless, plants expressing the *RPM1* gene are not susceptible, instead ETI is activated when AvrB/AvrRpm1-RIN4-RPM1 are all present in the plant (Mackey *et al.*, 2002). This is because RIN4 interacts with RPM1 and the modification of RIN4 by either AvrB or AvrRpm1 leads to RPM1 activation (Mackey *et al.*, 2002). Thus, by guarding an effector target, the plant is able to use a single R protein against two unrelated effectors (**Figure 1.8**).

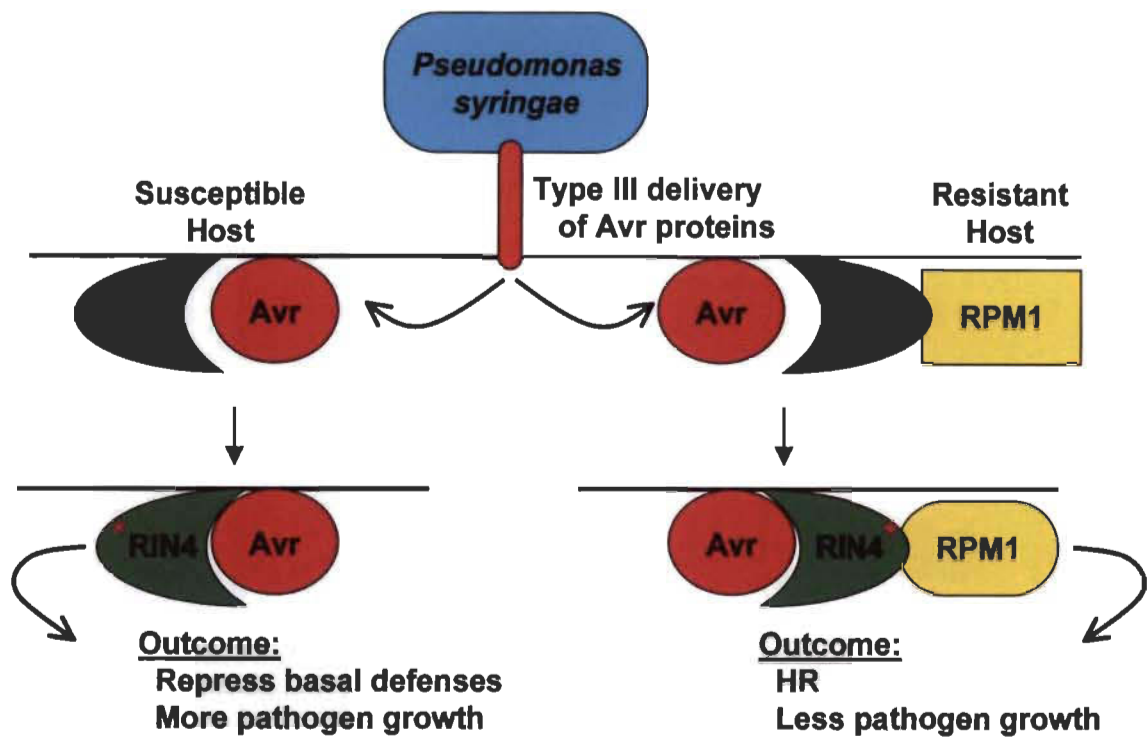


Figure 1.8. Role of RIN4 and RPM1 in the plant-*Pseudomonas syringae* interaction.
Avr represents either AvrB or AvrRPM1. Source: Mackey *et al.* (2002).

In some cases, the protein guarded by an R protein is not the virulence target of the cognate effector, in other words, the interaction between the effector and the guarded protein does not help the virulence of the pathogen in the absence of the R protein (van der Hooft & Kamoun, 2008). These proteins are called decoys, they are similar to the real targets of an effector, but non-functional, and their sole purpose is to trap the effector in a recognition event (van der Hooft & Kamoun, 2008). This is the case of Pto, a protein bound to the membrane with a kinase domain that is structurally similar those of the PRRs FLS2 and EFR (**Figure 1.9**). In tomato plants lacking the kinase protein Pto or the resistance protein Prf, the effectors AvrPto and AvrPtoB from *P. syringae* interact with and inhibit these PRRs, preventing PTI responses (Xiang *et al.*, 2008). However, when both Pto and Prf are present, AvrPto and AvrPtoB also interact with Pto, which leads to the activation of Prf and the induction of ETI (Chang *et al.*, 2000; Mucyn *et al.*, 2006).

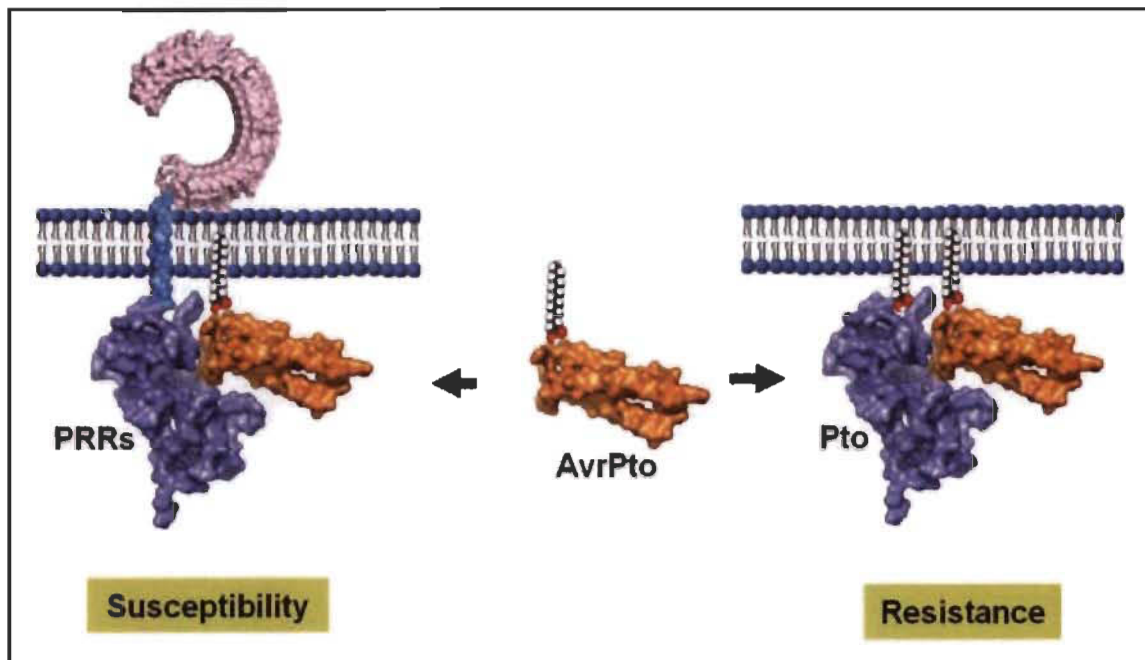


Figure 1.9. Model representing the role of Pto in plant resistance.

AvrPto interacts with receptor-like kinases, blocking PTI. Pto is similar to PRRs, thus AvrPto interacts with it, but it is not a RLK, thus the interaction does not block PTI. Source: Zong *et al.* (2008).

The typical architecture of resistance proteins is: an N-terminal Toll/Interleukin-1 receptor/Resistance protein (TIR) or a Coiled-Coil domain (CC) followed by a Nucleotide Binding (NB), Apaf1, R protein and CED4 (ARC) 1 and ARC2 (also known as NB-ARC or NBS domain), and a C-terminal LRR domain. However, resistance genes have been identified with multiple architectures. For example, in *Populus trichocarpa*, there are genes encoding the NBS domain alone, the only the NBS-LRR region, and other combinations of these domains, such as TIR/CC-NBS and TIR-CC-NBS-LRR (Kohler *et al.*, 2008). In *Arabidopsis*, the resistance protein TIR-NBS2 interacts with EXO70B1 and is necessary for activated defense responses against the powdery mildew *Golovinomyces cichoracearum* (Zhao *et al.*, 2015), however its exact function is still unknown.

Other domains also appear in NLR genes, including *Drosophila* BEAF and DREF finger (BED) domain (Germain & Seguin, 2011; Kohler *et al.*, 2008; Kroj *et al.*, 2016; Marchal *et al.*, 2018), protein kinase, WRKY, and NAC domains (Kroj *et al.*, 2016;

Sarris *et al.*, 2016). The occurrence of NLRs with extra domains are the result of gene fusion events and this type of resistance proteins is common among angiosperms (Sarris *et al.*, 2016). These integrated domains (IDs) serve as sensors for effectors. Thus, NLR-IDs can recognize effectors directly, but the effector is less likely to escape recognition, since it would mean not being able to recognize its target. Another benefit NLR-IDs confer to the plant is that they may detect multiple effectors. This is the case for Resistant to *R. solanacearum* 1 (RRS1, *A. thaliana*). This is an R protein with an integrated WRKY domain (TIR-NBS-LRR-WRKY) that acts with another R protein, Resistance to *P. syringae* 4 (RPS4) and they accumulate mostly in the nucleus (Huh *et al.*, 2017). The WRKY domain of RRS1 acts as a decoy, trapping AvrRPS4, from *Pseudomonas syringae* (Deslandes *et al.*, 2003; Sarris *et al.*, 2015), and PopP2, from *R. solanacearum* (Sarris *et al.*, 2015), in a recognition event. RRS1 then derepresses RPS4, which induces ETI. In *Arabidopsis*, RRS1 and RPS4 are also necessary for resistance against *Colletotrichum higginsianum*, however the effector recognized in this case is still unknown and thus it is not clear if the integrated domain participates in the recognition (Narusaka *et al.*, 2009).

1.2.4 Interaction between PTI and ETI

The outcome of PTI and ETI may be very different, but these two layers of immunity have many similarities. In both cases there is activation of MAPK cascade (Nicaise *et al.*, 2009). Yet, treatment of *Arabidopsis* with the PAMP flg22 showed MPK3 and MPK6 activation only during the first 10 min, while MPK3 and MPK6 remained activated from three to 10 hours post-induction of *in planta* expression of the avirulence gene AvrRpt2 (which induces ETI through the R protein Resistance to *P. syringae* 2, RPS2) (Tsuda *et al.*, 2013). Interestingly, *in planta* expression of AvrRps4 (which induced ETI through RPS4) did not cause activation of MPK3, MPK6, MPK4 or MPK11 (Ngou *et al.*, 2020). On the other hand, there are three peaks of ROS burst when both PTI and ETI are activated, but they are less intense during PTI alone. During ETI caused by AvrRps4, only the third peak is detectable, while no ROS burst is detected during ETI caused by AvrRpt2 (Ngou *et al.*, 2020). Similarly, callose deposition only increases during PTI and PTI+ETI,

not during ETI induced by AvrRps4 (Ngou *et al.*, 2020). However, MPK3 and MPK6 activation and ROS burst were prolonged, ROS burst was more intense and callose deposition was increased during PTI+ETI, and ETI activation led to increased accumulation of BIK1, BAK1, RBOHD and MPK3, as well as induction of expression of the genes encoding these proteins and MPK6. These data suggest that plant defense cannot happen through activation of ETI alone, however ETI replenishes PTI signaling components, prolonging and enhancing PTI responses (**Figure 1.10**) (Ngou *et al.*, 2020; Yuan *et al.*, 2020).

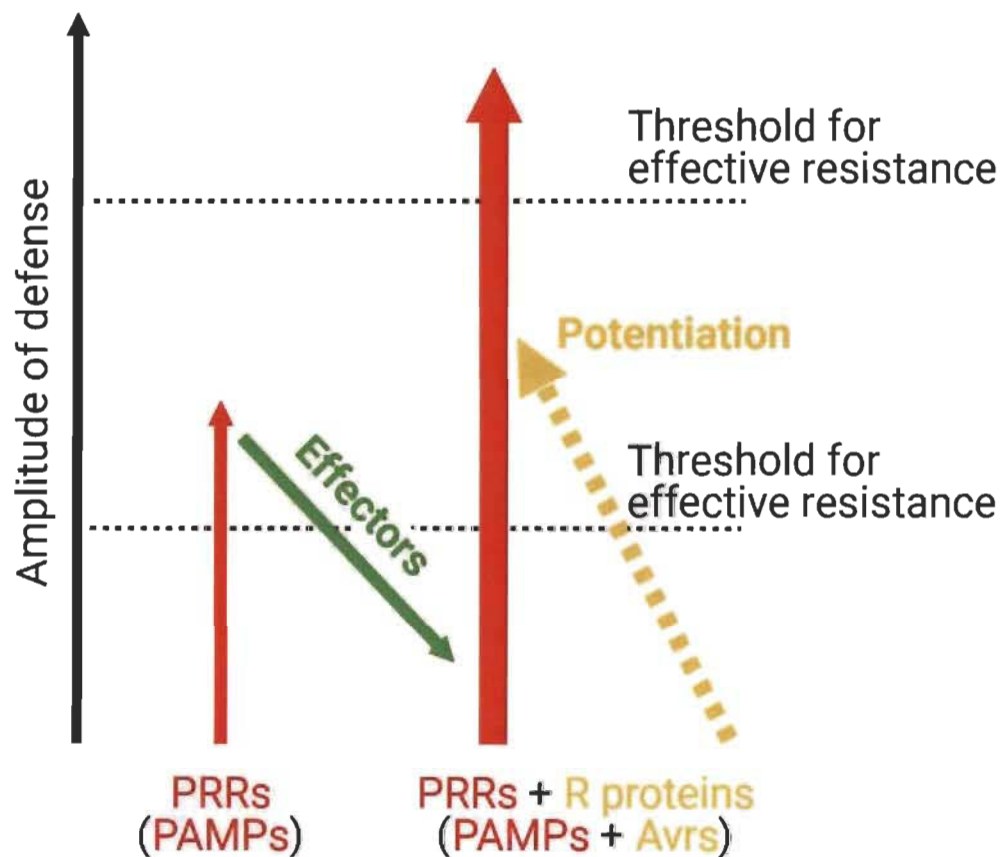


Figure 1.10. Updated zig zag model showing that ETI amplifies PTI responses.
Adapted from: Ngou *et al.*, 2020.

1.2.5 ETI and necrotrophic pathogens

Plant pathogens are grouped into three categories: biotrophic pathogens feed and reproduce on live tissues, necrotrophic pathogens kill the plant cells and feed on the dead tissue and hemibiotrophic organisms have an initial biotrophic phase and then switch to

necrotrophy. Necrotrophic and hemibiotrophic pathogens can trick the plant immune system and activate ETI to cause cell death (Lorang *et al.*, 2007; Wang *et al.*, 2014; Zhang *et al.*, 2018; Zuppini *et al.*, 2005). For instance, ToxA is a toxin present in *Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, and *Bipolaris sorokiniana* (McDonald *et al.*, 2019) that interacts with the NBS-LRR protein Toxin sensitive 1 (Tsn1) from wheat, leading to plant cell death (Adhikari *et al.*, 2009). Another example is *Cochliobolus victoriae*, which secretes the toxin victorin into its host. Victorin is recognized by the resistance protein LOV1, leading to PR1 gene expression, camalexin accumulation, and cell death (Gilbert and Wolpert, 2013).

1.3 Effectoromics

Effectors are key players in the outcome of plant-microbe interactions. Thus, to improve plant immunity, we need to understand how pathogens use these molecules. First, there are different types of effectors: proteins (De Wit *et al.*, 2009; Rocafort *et al.*, 2020; Wawra *et al.*, 2012b), secondary metabolites (Collemare *et al.*, 2019), and small non-coding RNAs (Wang *et al.*, 2015). Although effector biology has focused on proteinaceous effectors, their identification and functional characterization are still challenging. Bacteria secrete effectors mainly through T3SS and type four secretion system (T4SS, **Figure 1.11**) (Cunnac *et al.*, 2009). T3SS is made of a basal body that starts with an ATPase in the cytoplasm and crosses both the inner and outer membranes, a needle/pilus that originates from the basal body and crosses multimeric pores (known as secretin) in the outer membrane and reaches the extracellular space, and a translocon that is assembled when the bacterium is in contact with the host cell and forms a pore in the host membrane allowing the passage of effectors into the host cytoplasm (Büttner & He, 2009; Green & Mecsas, 2016). Examples of effectors of plant pathogenic bacteria secreted using T3SS are: AvrRpm1 (*Pseudomonas syringae* (Mackey *et al.*, 2002)), AvrXa27 (*Xanthomonas oryzae* (Gu *et al.*, 2005)), GALA (*Ralstonia solanacearum* (Angot *et al.*, 2006)), and HsvB (*Pantoea agglomerans* (Nissan *et al.*, 2006)).

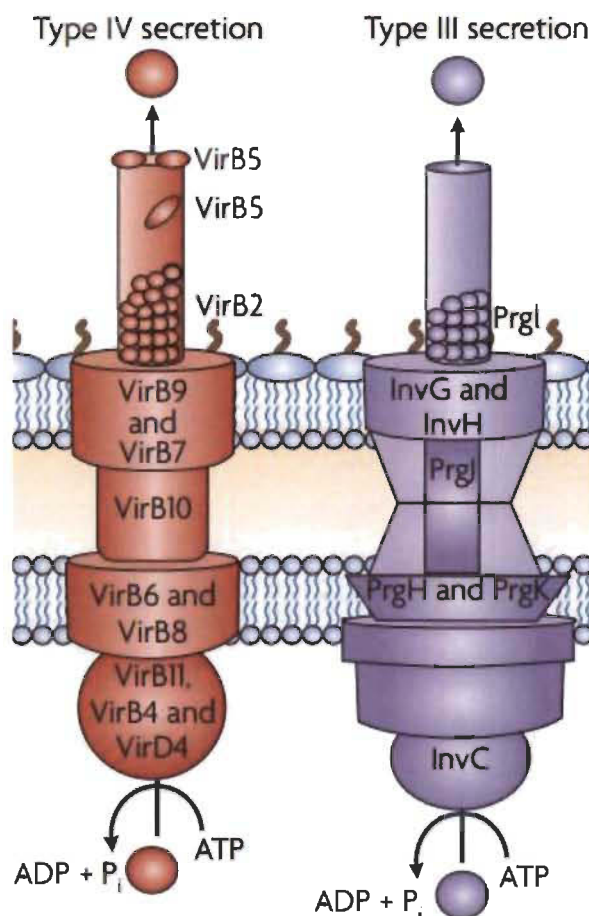


Figure 1.11. Schematic overview of types III (right) and IV (left) secretion systems.

From bottom to top: bacterial cytoplasm, inner cell membrane, periplasm, outer cell membrane, extracellular space. Source: Fronzes *et al.* (2009).

The T4SS (**Figure 1.11**) is present both in Gram-positive and Gram-negative bacteria (Fronzes *et al.*, 2009) and it has been extensively studied in the causal agent of the crown gall disease, *Agrobacterium tumefaciens* (Green & Mecsas, 2016; Zhu *et al.*, 2000). The T4SS is made of 12 proteins: a type IV coupling protein (the ATPase VirD4), the inner membrane complex (the ATPases VirB4 and VirB11 and the transmembrane proteins VirB3, VirB6 and VirB8), the outer membrane complex (VirB2, VirB5, VirB7, VirB9 and VirB10) (Christie *et al.*, 2014; Green & Mecsas, 2016). In *A. tumefaciens*, this secretion system is known to secrete the complex VirD2-Transfer DNA, VirE2 and VirF, which act as effector proteins/molecules in the host cell (Zhu *et al.*, 2000). The Type four effector protein A from *Sinorhizobium meliloti* and *S. medicae* is another example of effector secreted by the T4SS (Nelson *et al.*, 2017), which in this case act in the

establishment of a beneficial interaction between these bacteria and *Medicago truncatula*. The T4SS is also used by animal pathogens, such as *Helicobacter pylori* (Segal *et al.*, 1999), *Bordetella pertussis* (Covacci & Rappuoli, 1993) and *Legionella pneumophila* (Ninio & Roy, 2007).

The translocation of effector molecules from fungi and oomycete symbionts into plants is still a mystery. Secretion systems as the ones found in bacteria have not been discovered in these other groups of microbes. Nevertheless, characterized effectors from both fungi and oomycete usually have signal peptides (Kamoun, 2007), allowing them to be secreted from the pathogen cell. Also, the N-terminal region of oomycete effectors have domains that may play a role in their secretion by the microbe and uptake by the plant host. Dou *et al.* (2008) showed that the Arg-X-Leu-Arg and Asp-Glu-Glu-Arg (RXLR-dEER) motif of the *Phytophthora sojae* effector Avr1b is implicated in its translocation into soybean cells and is able to do so even in the absence of the pathogen. Later, Kale *et al.* (2010) showed that the RXLR motif bound to phosphatidylinositol-3-phosphate (PI3P) in the host cell membrane, which allowed the uptake of the effector into the plant cell by endocytosis. This study also discovered the fungal effector motif Arg/Lys/His-X-Leu/Met/Ile/Phe/Tyr/Trp-X ([R/K/H]X[L/M/I/F/Y/W]X motif). It is present in AvrL567 (*M. lini*), Avr-Pita (*Magnaporthe oryzae*), Avr1-3 (*Fusarium oxysporum*) and AvrLm6 (*Leptosphaeria maculans*) and also bound to PI3P. Nevertheless, the interaction of RXLR motif and PI3P and the role of the RXLR-dEER motif in effector uptake is still debated (Ellis & Dodds, 2011). For instance, the effector Avh5, from *P. sojae*, needed both its N-terminal (RXLR motif) and its basic C-terminal region for cell entry (Sun *et al.*, 2013). On the other hand, Avr3a (*P. infestans*) was shown to interact with PI3P through an α 1 helix in the “effector” domain (C-terminal region of the effector sequence) not the RXLR region (Yaeno *et al.*, 2011), although Wawra and colleagues showed that the denatured Avr3a interacts with phospholipids, thus the interaction would be meaningless in the context of the infection (Wawra *et al.*, 2012a). The Crinkler (CRN) protein family is another group of oomycete effectors. They have the Leu-X-Leu-Phe-Leu-Ala-Lys (LXLFLAK) motif in their N-terminal region which was shown to be responsible for their translocation to the host cell (Schornack *et al.*, 2010)

and a variety of C-terminal domains that lead to the accumulation of GFP in the nucleus of *N. benthamiana* (Stam *et al.*, 2013).

Effectors are a group of protein with many different functions (chitinase, transcription factor, protease, etc.) and that accumulate in different subcellular localizations (apoplast, cytoplasm, nucleus, discussed in detail in section 1.2.2). Because of this, searching for effectors in the genome, transcriptome or proteome of microbes is challenging. In the case of bacteria, since these proteins are secreted by T3SS and T4SS, the N-terminal motif that targets proteins to these secretion systems can be used to select candidate effectors (CEs) in bacterial proteomes. This strategy is also used for oomycetes: genomes are screened for proteins containing LXLFLAK or RXLR motifs (Ai *et al.*, 2020; Armitage *et al.*, 2018; Oh *et al.*, 2009; Rojas-Estevez *et al.*, 2020; Tabima & Grünwald, 2019). In the case of fungal effectors, the screening is not as straightforward, because motifs as the ones in bacterial and oomycete effectors have not been found in fungi. As a result, many screening pipelines (for example **Figure 1.12**) have been designed with criteria including: presence of signal peptide (since these proteins must be secreted from the microorganism to the host) and absence of transmembrane motifs (because signal peptides could then indicate that these proteins are destined the plasma membrane, not to be secreted), signatures of positive selection (indicating genes under the pressure of host resistance), exclusive or induced expression in infection structures, unknown function, small size and high cysteine content (Carreón-Anguiano *et al.*, 2020; Duplessis *et al.*, 2011a; Hacquard *et al.*, 2012; Petre *et al.*, 2015; Saunders *et al.*, 2012). These criteria are based on characteristics of known effector proteins (Stergiopoulos & de Wit, 2009; Tyler & Rouxel, 2013).

A CANDIDATE EFFECTOR PROTEINS SELECTION

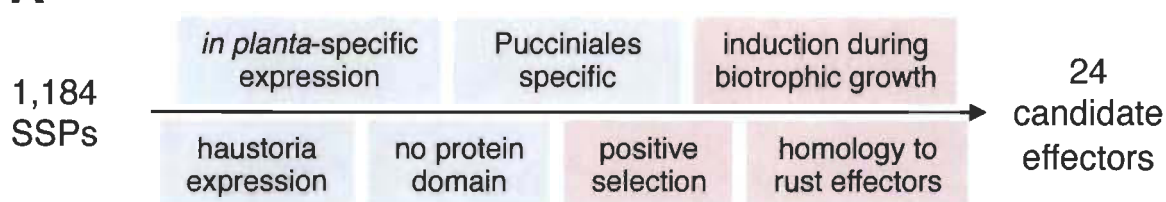


Figure 1.12. Pipeline for selection of candidate effectors of *M. larici-populina* from list of small secreted proteins.

From the genome annotation, the authors selected proteins of size < 300 amino acids, containing a signal peptide and no transmembrane domain (small secreted proteins, SSPs). The seven-step pipeline was used to filter the most promising candidate effectors in *M. larici-populina*. Blue boxes indicate steps done through an automated pipeline; red boxes indicate manual pipeline. Source: Petre *et al.* (2015).

Nevertheless, these pipelines may overlook true effectors (Kanja & Hammond-Kosack, 2020). For example, a pipeline that only selected small proteins (length < 300 amino acids (Duplessis *et al.*, 2011a; Joly *et al.*, 2010)) with high cysteine content (cysteine content > 3% (Saunders *et al.*, 2012)) and containing a signal peptide would not consider at least three *bona fide* effectors. Isochorismate synthase from the oomycete *Phytophthora sojae* and fungus *Verticillium dahliae* (PsIsc1 and VdIsc1, respectively) are two unconventionally secreted effectors that do not present signal peptides in their N-termini and have low cysteine content (PsIsc1 = 1.9%, VdIsc1 = 0.5% (Liu *et al.*, 2014)). AvrM is an effector of 314 amino acids from *Melampsora lini* with only one cysteine (0.3% (Catanzariti *et al.*, 2006)).

Following identification of CEs, effector biologists need to functionally characterize these proteins to determine their roles in plant-pathogen interaction. First, it is necessary to confirm that the selected CE is translocated to the host's apoplast or cytoplasm. The translocation of several fungal effectors has been corroborated by transforming the fungus with an effector:tag protein construction and following the infection using microscopy: *Ustilago maydis* Tin2 (tagged with mCherry or iLOV (Tanaka *et al.*, 2015)), *U. maydis* Cmu1 (tagged with mCherry and HA (Djamei *et al.*, 2011b)), *Magnaporthe oryzae* AVR-Pita1 and Prevents pathogenicity toward Weeping Lovegrass (PWL1 and PWL2, tagged with eGFP or mRFP (Khang *et al.*, 2010)). Another way of detecting effectors in the plant during the infection is by using antibodies against the effector being

studied, which was done for *Xanthomonas* AvrBs3 (Szurek *et al.*, 2002), *Myzus persicae* Mp10 (Mugford *et al.*, 2016) and several *M. larici-populina* CEs (Hacquard *et al.*, 2012).

To evaluate the role of a CE in the infection, it is possible, in some cases, to infect the host with CE-encoding gene knockout strains of the pathogen. This was done with *Fusarium oxysporium* f. sp. *lycopersici* to establish that the effector SIX1, which induces resistance in tomato plants containing the I-3 resistance gene, is necessary for full virulence in susceptible tomato plants (Rep *et al.*, 2005) and to identify SIX3/Avr2 as the effector recognized by I-2 resistance protein in tomato, which leads to hypersensitive response (Houterman *et al.*, 2009). When knockout is not an option, as is the case for obligate biotrophic pathogens, it is sometimes possible to use host-induced gene silencing to knockdown genes of interest. This technique has been used to characterize CEs from the barley powdery mildew *Blumeria graminis* f. sp. *hordei* (Aguilar *et al.*, 2016), the wheat stripe rust *Puccinia striiformis* f. sp. *tritici* (Liu *et al.*, 2016; Xu *et al.*, 2019), the cucurbit powdery mildew *Podosphaera xanthii* (Martínez-Cruz *et al.*, 2018). Nevertheless, it is possible that the lack of the CE in the pathogen, through either knockdown or knockout techniques, results in minimal impact to the pathogen because of functional redundancy in the repertoire of effectors (Tyler, 2017).

Finally, the CE can be heterologously expressed in another plant pathogen that is then used to infect a plant, or it can be expressed directly in a plant, constitutively or transiently (Lorrain *et al.*, 2018a). There are issues with this technique. Namely, if the effector's host target is not present in the non-host plant used, results may not reflect the reality in the true host (Fudal *et al.*, 2018; Motteram *et al.*, 2009). For example, the effector may not accumulate in the same subcellular compartment or it may interact with another protein, not present in the true host (Robin *et al.*, 2018). In the case of AvrLm1 from *Leptosphaeria maculans*, which infects canola, the effector interacts with *Arabidopsis*'s MAPK9, a protein with two homologs in canola (Ma *et al.*, 2018). Although the authors of the study investigated only one of the homologs, it is not possible to know if the two have the same function or if both are targeted by AvrLm1. Regarding the organism used for the heterologous expression, an effector from an eukaryotic pathogen expressed in a

bacterium might not be properly folded or post-translationally modified (Snelders *et al.*, 2020). However, *in planta* expression of an effector-encoding gene is also problematic, because protein interaction and subcellular localization results may be caused by forcing an apoplastic effector inside the plant cell, because the gene is expressed in high levels (Robin *et al.*, 2018) or the fusion of a fluorescent protein may alter its folding. On the other hand, heterologous expression of CEs allows to study of these proteins without the interference of the rest of the pathogen's secretome, which prevents redundant effectors from hiding the impact of the CE under study. The technique also makes possible the fast screening of CEs with all the resources available for model plants.

Several CEs from *M. larici-populina* (*Mlp*) CEs were studied in *Arabidopsis thaliana* (Germain *et al.*, 2018) or in *Nicotiana benthamiana* (Petre *et al.*, 2015), showing they accumulate in multiple cellular compartments, including the chloroplast, the nucleus and nucleolus and the mitochondria (**Figure 1.13** and **Figure 1.14**, respectively). Also, by constitutively expressing *Mlp* CEs in *Arabidopsis*, Germain *et al.* (2018) showed that all the 14 CEs studied caused ETS to either *P. syringae*, or *Hyaloperonospora arabidopsidis* or to both pathogens. Likewise, Petre *et al.* (2016) found the subcellular localization of seven CEs from *P. striiformis* f. sp. *tritici* by transiently expressing them in *N. benthamiana*.

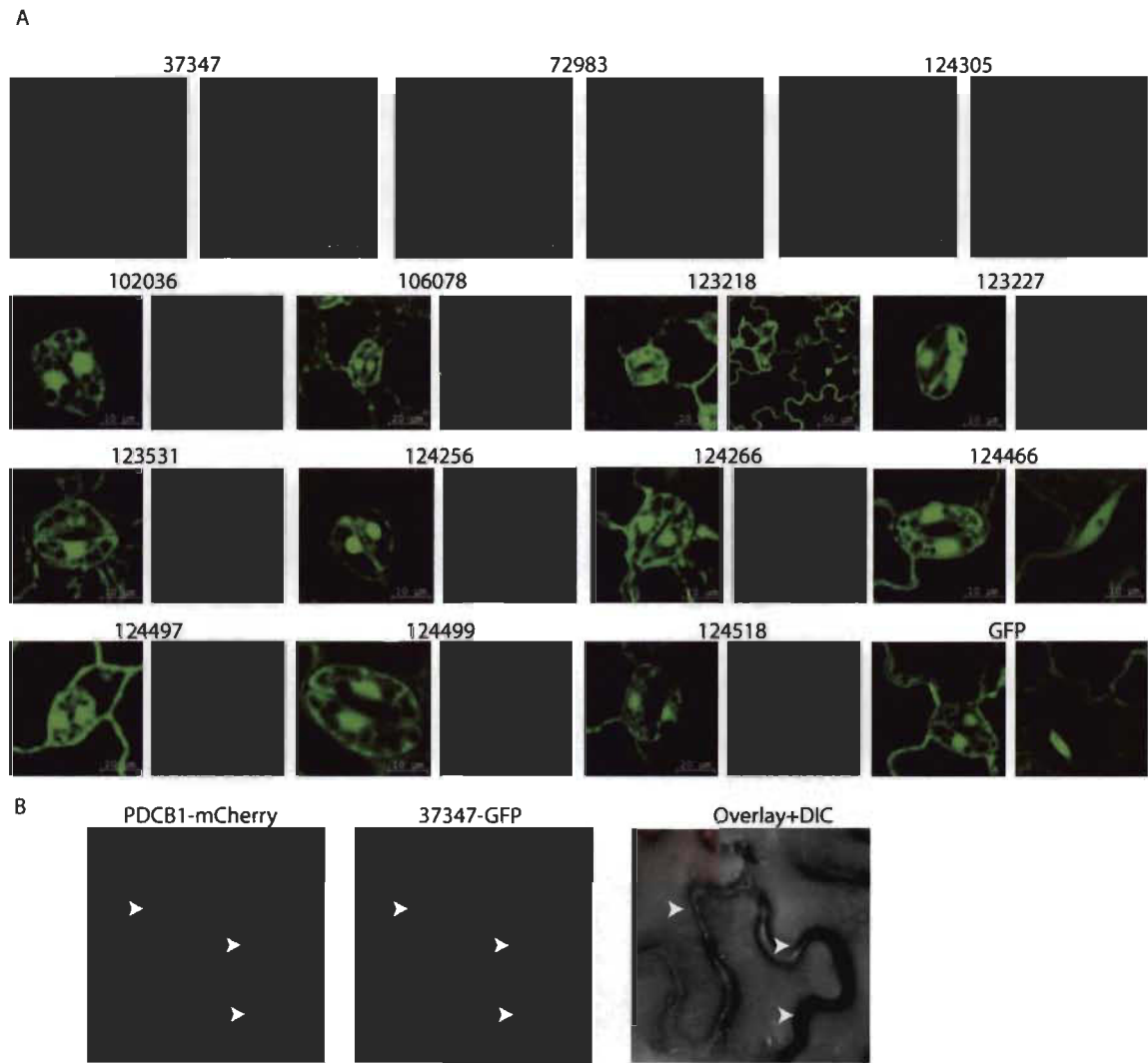


Figure 1.13. A) Subcellular localization of *Mlp* CEs in *Arabidopsis*. B) 37347 accumulates in the plasmodesmata.

Source: Germain *et al.* (2018).

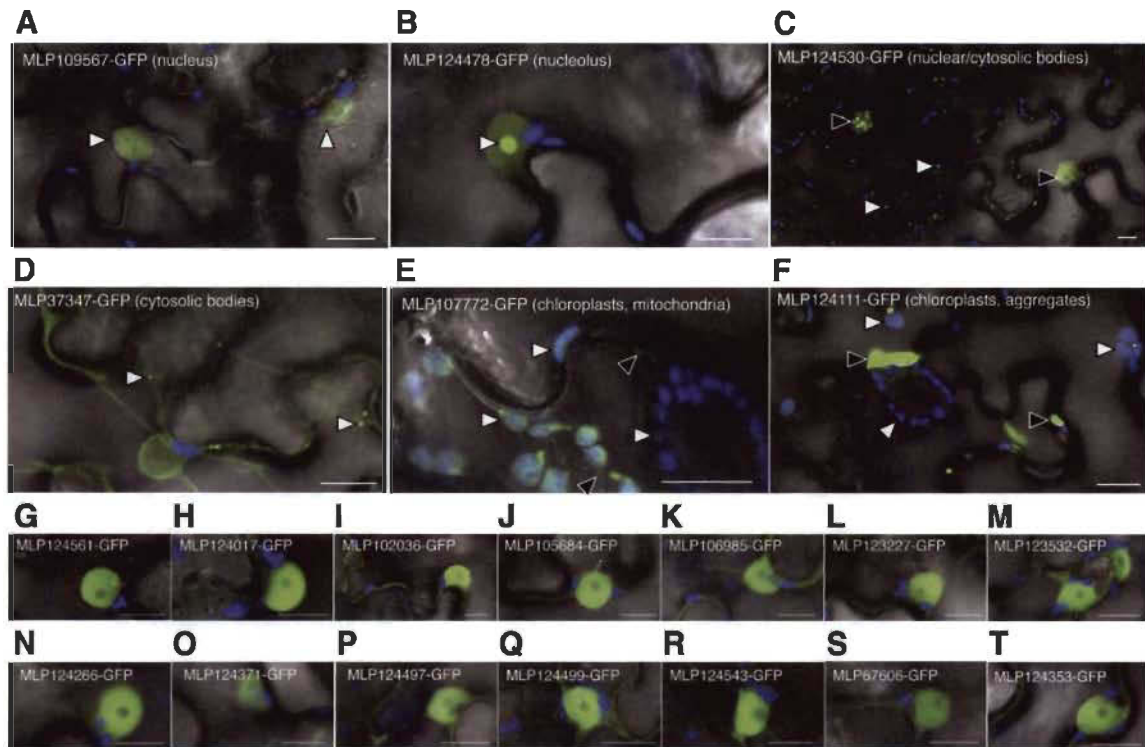


Figure 1.14. Subcellular localization of *Mlp* CEs in *Nicotiana benthamiana*. Blue indicates autofluorescence from the chloroplast. Source: Petre *et al.* (2015).

1.4 Rust fungi

The rust fungi (order Pucciniales, phylum Basidiomycota) are the biggest group of plant pathogens, with around 8000 species (Aime *et al.*, 2017; Lorrain *et al.*, 2019). Rust fungi are obligate plant pathogens, meaning they are unable to feed, grow and reproduce on anything but their living hosts, although growth induction of initial infection structures has been reported for *Uromyces vignae*, the cowpea rust fungus (Heath, 1990), and for *P. graminis* f.sp. *tritici*, the wheat stem rust fungus (Wiethölter *et al.*, 2003). Pucciniales species are among the most important fungal plant pathogens (Dean *et al.*, 2012), infecting ferns, gymnosperms and almost all families of angiosperms (Aime *et al.*, 2017), but individually they have narrow host ranges. Important crops are infected by rust fungi, including: apple (cedar-apple rust, *Gymnosporangium juniperi-virginianae*), barley (*P. coronata* var. *hordei*), coffee (*Hemileia vastatrix*), maize (*P. sorghi*), oat (oat crown rust, *P. coronata*), soy (*Phakopsora pachyrhizi*), sugarcane (brown rust caused by *P. melanocephala* and orange rust caused by *P. kuehnii*), and wheat (leaf rust caused by

P. triticina, *P. graminis* f.sp. *tritici* which is the causal agent of stem rust and also infects barberry, and stripe rust caused by *P. striiformis*). Trees used for the paper industry and for wood are also among the species infected by rust fungi: ash tree (*P. sparaganioides*), eucalyptus (*P. psidii*), larch, *Pseudotsuga* and poplar (*Melampsora*), *Picea* spp. (*Chrysomyxa* and *Melampsorella*), and *Pinus* spp. (*Cronartium* and *Coleosporium*).

1.4.1 Life cycle

In addition to their dependency on their hosts, another characteristic of rust fungi is their complex life cycle (Lorrain *et al.*, 2019), illustrated in **Figure 1.15**. The most complex life cycle among the rust fungi includes the production of five types of spores (macrocyclic) and the infection of two unrelated host species (heteroecious). In macrocyclic and heteroecious species, the basidiospores (haploid, n) infect one host species, where it develops the spermogonia/pycnia which produces spermatia/pycniospores (n). Spermatia/pycniospores of different mating types fuse cytoplasm, forming dikaryotic ($n+n$) aecia. Aeciospores ($n+n$) infect another host species and develop uredinia ($n+n$), which reproduce asexually. Karyogamy occurs in telia (which are formed from urediniospores) and it produces teliospores (diploid, $2n$). These form basidia ($2n$), in which meiosis occurs, producing basidiospores.

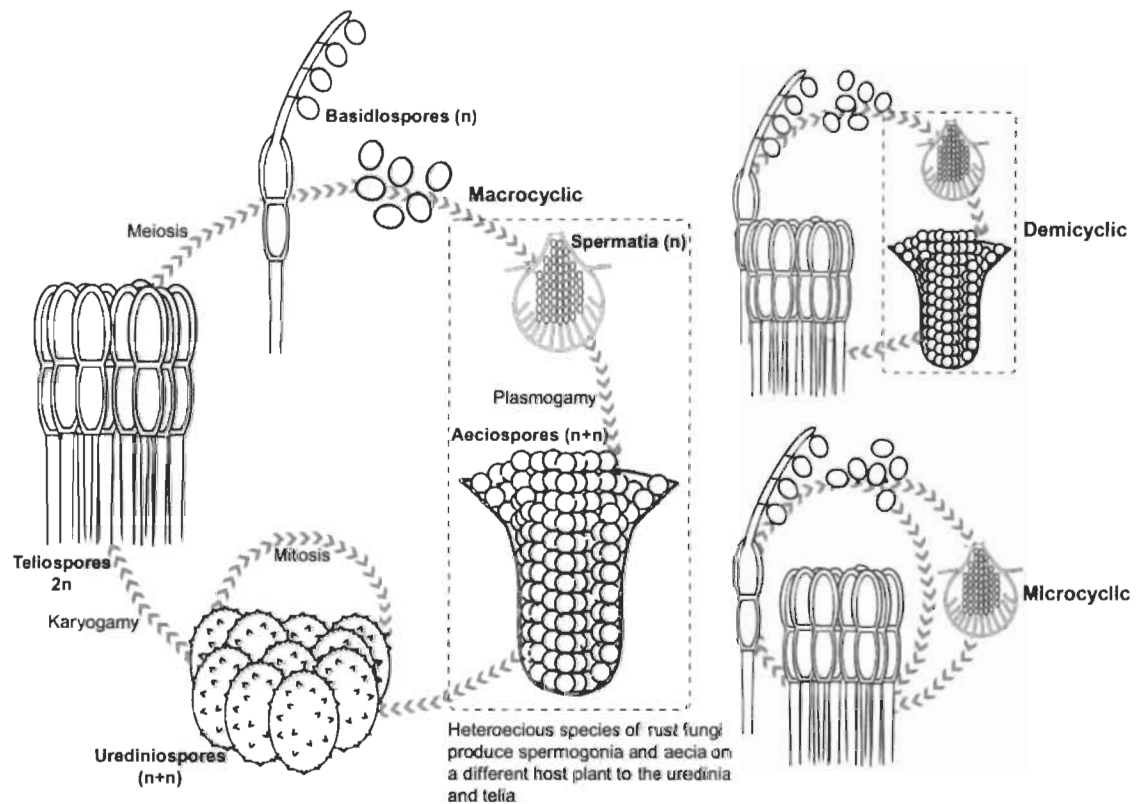


Figure 1.15. Rust fungi life cycles.

Source: Aime *et al.* (2017).

The poplar rust fungus, *M. larici-populina*, is an example of a macrocytic and heteroecious rust fungus (**Figure 1.16**). It infects larch during spring with basidiospores and develops both spermatogonia and aecia in this plant, then aeciospores infect poplar and develop uredinia. Urediniospores repeatedly infect poplar during summer and then develop into telia in autumn. Teliospores overwinter on dead poplar leaves and germinate in spring forming basidia. *P. asparagi* is a macrocytic autoecious species, producing all five spores in one host, the asparagus (Elmer, 2001). Demicyclic species lack uredinia and urediniospore, thus they do not reproduce asexually (Parmelee & Malloch, 1972; Salama, 2009). Finally, microcyclic species only produce teliospores and basidiospores (Ono, 2002), although they may also produce spermatia, and they are all autoecious (Aime *et al.*, 2017).

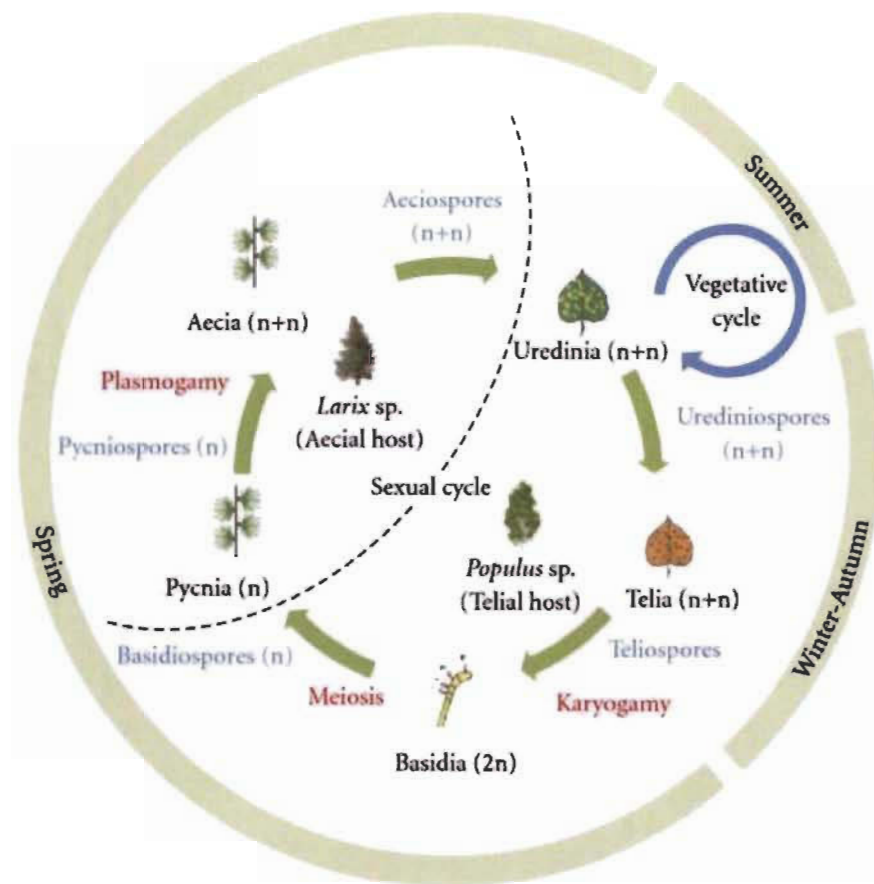


Figure 1.16. *M. larici-populina*'s life cycle.

Source: Hacquard *et al.* (2011).

1.4.2 Infection process

Of the five spore types rust fungi produce, three can infect plants: aeciospores, urediniospores and basidiospores (Hahn, 2000). Infection by the two dikaryotic spore types is similar, typically involving penetration of the leaf through the stomata and formation of several specialized cells, while basidiospores usually epidermal cells and develop less differentiated mycelia (Hahn, 2000; Lorrain *et al.*, 2019). The infection process is better characterized for the urediniospores and is illustrated in **Figure 1.17**. These spores attach themselves to the leaf and germinate forming a germ tube, which grows over the leaf until it finds a stomatal aperture (Hahn, 2000). Over the stomata, it forms an appressorium that uses turgor pressure to push the guard cells open (Mendgen, 1975) and develops a penetration hypha that enters through the stomata and forms a substomatal vesicle (Lorrain *et al.*, 2019). This vesicle elongates into an infection

hypha that searches for mesophyll cells and differentiates into a haustorium mother cell (HMC) when in close contact with a host cell. The HMC forms a penetration peg that breaks through the host cell wall, using mechanical pressure and cell wall degrading enzymes, and invaginates the host cell membrane, forming the haustorium. This cell constitutes the most important site of communication between the pathogen and its host: through it the pathogen absorbs host nutrients and secretes effectors to control the host metabolism and defense (Garnica *et al.*, 2014; Lorrain *et al.*, 2019). The haustorium is separated from the plant cell cytoplasm by the extrahaustorial matrix, an amorphous mixture of carbohydrates and proteins sealed from the rest of the apoplast by the haustorial neckband (a bridge between fungal and plant membrane, rich in phosphorous and iron), and the extrahaustorial membrane, which is a modified plant cell membrane lacking ATPase activity (Baka *et al.*, 1995; Heath, 1976).

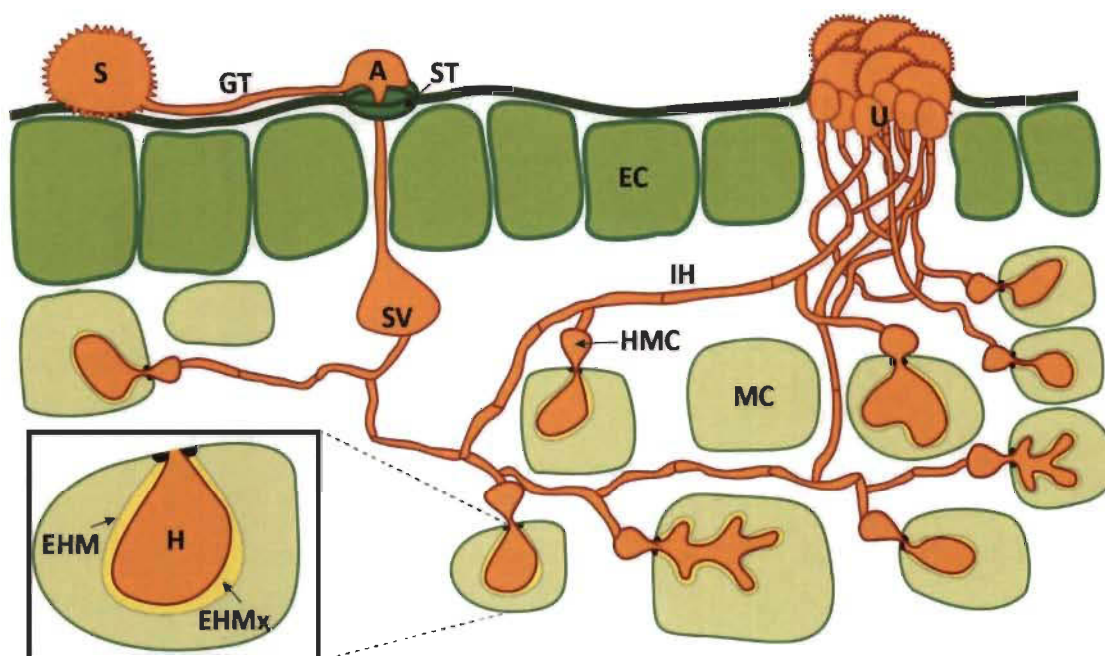


Figure 1.17. Representation of asexual cycle of *Puccinia* spp. on wheat.

S, spore. GT, germ tube. A, appressorium. ST, stomata. SV, substomatal vesicle. H, haustorium. EHM, extrahaustorial membrane. EHMx, extrahaustorial matrix. IH, infection hypha. HMC, haustorial mother cell. MC, mesophyll cell. EC, epidermal cell. U, urediniospores. Black dots at the top of each haustorium represent the haustorial neckband. Source: Garnica *et al.* (2014).

1.5 Research objectives

Effectors are key players in the interaction between an adapted pathogen and its host. They may protect the pathogen in the apoplast (Fiorin *et al.*, 2018; Marshall *et al.*, 2011; Tian *et al.*, 2020), prevent activation of PTI (Ma *et al.*, 2017), facilitate the access to nutrients (Chen *et al.*, 2010; Cohn *et al.*, 2014; Yang *et al.*, 2006) or even prevent the growth of competitors in the plant (Wu *et al.*, 2019). On the other hand, effectors are exposed, so the host might recognize them or detect their activities (Jones & Dangl, 2006), activating defense. Thus, the interaction between effectors and the plant's defense determines if the infection is successful or not. However, plant pathology has focused on effectors that are detected by the host, the avirulence proteins. Cultivars of major crops carrying specific R protein-encoding genes have been used worldwide, for instance wheat carrying the resistance gene *Sr38* was used in Australia, United Kingdom and United States in 1999 (Pretorius *et al.*, 2000). However, the widespread use of cultivars in which resistance to a pathogen is based on a single gene, together with the rapid life cycle of pathogens, as exemplified by the rust fungi, leads to selection of new pathogen races that are virulent on previously resistant plants. This is the case of the new stem rust (*Puccinia graminis* f.sp. *tritici*) race Ug99 (discovered in Uganda in 1998 and described in 1999), which is virulent to wheat varieties carrying the resistant genes *Sr31* and *Sr38*, among others (Singh *et al.*, 2011). In 2006, this stem rust race had already spread from Uganda to Sudan, North Africa, Yemen, Middle East and West-South Asia (Singh *et al.*, 2008), becoming a threat to wheat production worldwide.

It is evident that the research for resistant proteins in the plants and avirulence genes in the pathogens is not enough to help protect plants. To develop mechanisms of durable resistance, we need to understand compatible plant-microbe interactions, those in which the pathogen is virulent in the host. Effectors are essential in these interactions, however in rust fungi very few CEs have been confirmed as *bona fide* effectors (Petre *et al.*, 2014), and the virulence function has not been determined for any rust effector. Omics tools have been previously used to characterize CE functions (discussed in detail in section 1.2.2), for instance: the transcriptomic study of the impact of MoHTR1 and MoHTR2 on rice showed these CEs deregulate immunity-related genes (Kim *et al.*, 2020); the gene

expression analysis and metabolite quantification of plant tissue infected with Brg11-expressing bacteria revealed increased accumulation of polyamines and upregulation of a gene involved in polyamine biosynthesis (Wu *et al.*, 2019); and the transcriptomic and targeted metabolomic study of maize infected with wild type and Δ tin2 *U. maydis* showed that this effector upregulates the biosynthesis of anthocyanins (Tanaka *et al.*, 2014).

Thus, for this research we used omics tools, transcriptomics and untargeted metabolomics, to screen the impact of 14 CEs from the poplar leaf rust, *M. larici-populina* (a model organism for obligate biotrophic plant pathogens), in *Arabidopsis*. This model plant that cannot be infected by rust fungi, but it was shown to develop ETS to *P. syringae* or to *H. arabidopsidis* (**Figure 1.18** and **Figure 1.19**) when exposed to the CEs selected here (Germain *et al.*, 2018; Madina *et al.*, 2020). We separated the work into two chapters, one is published in a peer-reviewed journal (dos Santos, Desgagné-Penix, & Germain, 2020a) and the other is available as a preprint in bioRxiv (dos Santos, Pelletier, Séguin, Guillemette, Hawkes *et al.*, 2020b) and will be submitted shortly.

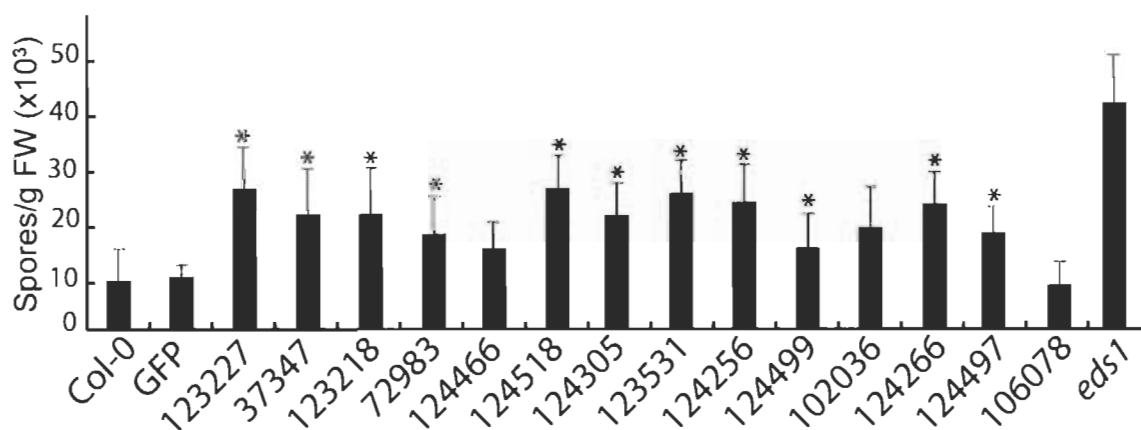


Figure 1.18. Expression of CE-encoding genes in *Arabidopsis thaliana* led to increased susceptibility to *H. arabidopsidis*.

Growth of the oomycete in the CE-expressing lines (numbers in the x-axis indicate the CE expressed in the transgenic line) was compared against *A. thaliana* Col-0 GFP. Source: Germain *et al.*, 2018.

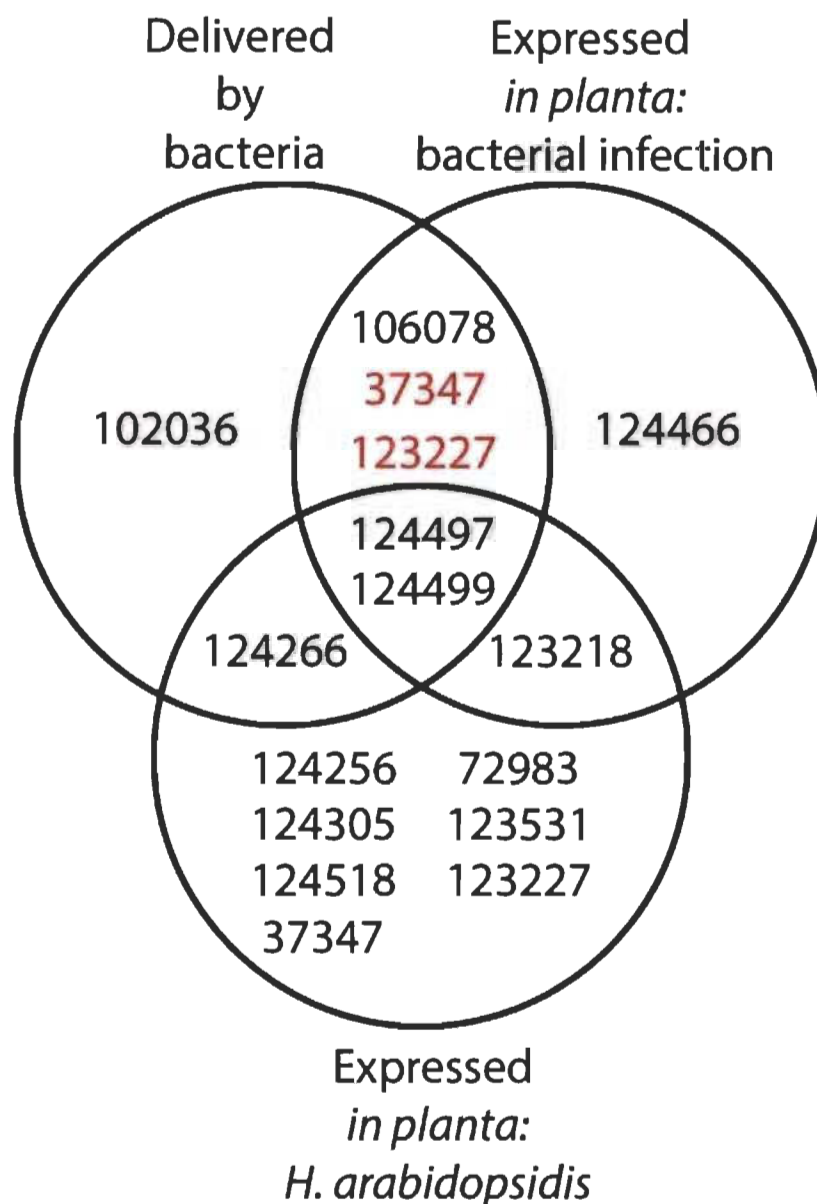


Figure 1.19. Impact of CEs (numbers) on plant susceptibility to bacterial and oomycete infection.

CEs were expressed either in bacteria and secreted into the plant (left circle) or constitutively *in planta* (right and bottom circles). Black and red numbers CE-expressing lines with increased and decreased susceptibility to pathogen infection, respectively. Source: Germain *et al.*, 2018.

Given that the difference between the control line and the CE-expressing lines in this work was the insertion of a single gene, we expected to find subtle changes in the transcriptome of each transgenic line. Thus, we needed to ensure that the data was well normalized in order to not miss weakly deregulated genes. We found the literature lacked user-friendly methods for selection of reference genes or methods to select reference genes for organisms or conditions that have not been extensively studied previously. This led us

to develop a method for selection of reference genes based exclusively on RNA sequencing data, which was published in BMC genomics (dos Santos *et al.*, 2020a) and is presented as **Chapter II**. This method was used for the transcriptomic analysis presented in **Chapter III**. The latter comprises the core work of this project, in which we performed full-transcriptomic analysis, sequence alignment and metabolome analysis using mass spectrometry to describe the impact of the CEs studied both on the transcriptome and the metabolome of *Arabidopsis* (dos Santos *et al.*, 2020b). A discussion is presented in **Chapter IV**.

Two papers to which I collaborated as a coauthor are presented as annexes. **Annex A** is a review on rust fungi biology published in New Phytologist in 2019 (Lorrain *et al.*, 2019). **Annex B** is the functional characterization of the CE Mlp124478 published in Scientific Reports in 2018 (Ahmed *et al.*, 2018).

CHAPTER II

CUSTOM SELECTED REFERENCE GENES OUTPERFORM PRE-DEFINED REFERENCE GENES IN TRANSCRIPTOMIC ANALYSIS

Karen Cristine Gonçalves dos Santos, Isabel Desgagné-Penix and Hugo Germain

Published on January 10th 2020 in BMC Genomics.

Correction to: **BMC Genomics** 21, 35 (2020). <https://doi.org/10.1186/s12864-019-6426-2>

Following publication of the original article, the author reported that although they used the word “covariance” throughout the manuscript, the actual statistics used for selection of reference genes was the “coefficient of variation”.

Note that this does not affect the method, nor the scripts present in GitHub (<https://github.com/KarenGoncalves/CustomSelection>) to process the datasets.

The original article has been corrected.

2.1 Author contributions

dos Santos, Desgagné-Penix and Germain designed the work; dos Santos performed the experiments; dos Santos and Germain wrote the paper; Desgagné-Penix and Germain revised the paper and all authors approved the manuscript.

2.2 Résumé de l'article

Introduction

Le séquençage d'ARN permet la mesure de l'expression génique à une résolution non satisfaite par les « oligoarrays » ou RT-qPCR. Il est cependant nécessaire de normaliser les données de séquençage par taille des bibliothèques, la taille et la composition des transcrits,

entre autres facteurs, avant de comparer les niveaux d'expression. L'utilisation de gènes de contrôle interne ou des « spike-ins » est préconisée dans la littérature pour mettre à l'échelle le nombre de lectures, mais les méthodes de choix des gènes de référence sont principalement ciblées sur les études RT-qPCR et nécessitent un ensemble de contrôles candidats présélectionnés ou des gènes cibles présélectionnés.

Résultats

Ici, nous rapportons un « pipeline » basé sur R pour sélectionner des gènes de contrôle interne basés uniquement sur le nombre de lectures et la taille des gènes. Cette nouvelle méthode normalise d'abord le nombre de lectures en transcrits par million (TPM), puis exclut les gènes faiblement exprimés en utilisant le script DAFS pour calculer le seuil. Elle sélectionne ensuite comme références les gènes avec la covariance TPM la plus faible. Nous avons utilisé cette méthode pour choisir des gènes de référence personnalisés pour l'analyse d'expression différentielle de trois ensembles de transcriptomes à partir de plantes transgéniques *Arabidopsis* exprimant des protéines effectrices fongiques hétérologues marquées avec la GFP (en utilisant la GFP seule comme contrôle). Les gènes de référence personnalisés ont montré une covariance et un niveau de dérégulation inférieurs ainsi qu'une gamme plus large de niveaux d'expression que les gènes de référence couramment utilisés. Lorsqu'ils sont analysés avec NormFinder, les gènes de référence typiques et personnalisés ont été considérés comme des contrôles internes appropriés, mais les gènes sélectionnés personnalisés ont été exprimés de manière plus stable. geNorm a produit un résultat similaire dans lequel la plupart des gènes sélectionnés personnalisés se classaient plus haut (c'est-à-dire étaient exprimés de manière plus stable) que les gènes de référence couramment utilisés.

Conclusion

La méthode proposée est innovante, rapide et simple. Comme elle ne dépend pas de l'annotation du génome, elle peut être utilisée avec n'importe quel organisme et ne nécessite pas de candidats de référence présélectionnés ou de gènes cibles qui ne sont pas toujours disponibles.

Mots-clés : séquençage nouvelle génération, Genes constitutifs pour qPCR, R script

2.3 Full article in English: Custom selected reference genes outperform pre-defined reference genes in transcriptomic analysis

Abstract

Background

RNA sequencing allows the measuring of gene expression at a resolution unmet by expression arrays or RT-qPCR. It is however necessary to normalize sequencing data by library size, transcript size and composition, among other factors, before comparing expression levels. The use of internal control genes or spike-ins is advocated in the literature for scaling read counts, but the methods for choosing reference genes are mostly targeted at RT-qPCR studies and require a set of pre-selected candidate controls or pre-selected target genes.

Results

Here, we report an R-based pipeline to select internal control genes based solely on read counts and gene sizes. This novel method first normalizes the read counts to Transcripts per Million (TPM) and then excludes weakly expressed genes using the DAFS script to calculate the cut-off. It then selects as references the genes with lowest TPM covariance. We used this method to pick custom reference genes for the differential expression analysis of three transcriptome sets from transgenic *Arabidopsis* plants expressing heterologous fungal effector proteins tagged with GFP (using GFP alone as the control). The custom reference genes showed lower covariance and fold change as well as a broader range of expression levels than commonly used reference genes. When analyzed with NormFinder, both typical and custom reference genes were considered suitable internal controls, but the custom selected genes were more stably expressed. geNorm produced a similar result in which most custom selected genes ranked higher (i.e. were more stably expressed) than commonly used reference genes.

Conclusions

The proposed method is innovative, rapid and simple. Since it does not depend on genome annotation, it can be used with any organism, and does not require pre-selected reference candidates or target genes that are not always available.

Keywords: Next-generation sequencing, Housekeeping genes for qPCR, R script

Background

RNAseq is a technique used since the pioneer studies of Lister *et al.* (2008) (*Arabidopsis thaliana*), Nagalakshmi *et al.* (2008) (*Saccharomyces cerevisiae*), Wilhelm *et al.* (2008) (*Schizosaccharomyces pombe*), and Mortazavi *et al.* (2008) (*Mus musculus*). This technique allows the combination of transcript discovery and expression level quantification in a single assay and has an unlimited dynamic range of detection compared to microarray or RT-qPCR (Wang *et al.*, 2009; Zhao *et al.*, 2014).

For differential expression studies, the gene expression values must be comparable between samples, which means that count data should be normalized for sequencing depth and other biases such as transcript length, GC content and transcript coverage. Reads/Fragments per Kilobase per Million (RPKM or FPKM) and Transcripts per Million (TPM) both normalize count data by transcript length and sequencing depth (Pachter, 2011; Wagner *et al.*, 2012), but they may give biased results in the presence of highly expressed genes or when a lot of the genes are expressed in only one sample. This is because one differentially expressed gene shifts the sequencing effort distributed to the others and all genes appear to be differentially expressed (Robinson & Oshlack, 2010; Wolf, 2013; Zhuo *et al.*, 2016). Other methods such as relative log expression (DESeq2) and trimmed mean of M-values (edgeR) can work with the carry-over effect of highly expressed genes (Wolf, 2013).

The comparison of different softwares for RNAseq analysis is a recurrent subject in the literature (Evans *et al.*, 2018; Rapaport *et al.*, 2013; Sonesson & Delorenzi, 2013) and

many authors argue over the benefits of using housekeeping genes or spike-in controls to scale the count data, yet the evaluation of the reference genes used for RNAseq data analysis is not as common. When using internal or external control genes, the normalization is first performed on the controls and the result is used to normalize the other genes. The use of external spike-ins is advocated for introducing little error into the read counts, allowing identification of global shifts in gene expression (Lovén *et al.*, 2012; Lutzmayer *et al.*, 2017; Taruttis *et al.*, 2018). However, reports have shown mixed performances with different normalization methods (Risso *et al.*, 2014), resulting in high false discovery rates and false positive rates (Paepe, 2015). These may show differences in amplification depending on the type of tissue studied or the protocol for mRNA enrichment (Qing *et al.*, 2013).

One alternative for external spike-ins is the use of internal control genes, as it is done in qPCR studies. Typical control genes are actin, tubulin, elongation factor 1, polyubiquitin and ribosomal RNAs, though the stability of expression of several of those is dependent on the conditions studied (Gutierrez *et al.*, 2008). To solve this issue, different algorithms were proposed to find stably expressed genes, mostly for qPCR applications, but they need a set of predefined genes of interest (RefGenes, (Hruz *et al.*, 2008)) or a set of pre-selected candidate reference genes (geNorm, (Vandesompele *et al.*, 2002); NormFinder, (Andersen *et al.*, 2004); BestKeeper, (Pfaffl *et al.*, 2004)). The most frequent approach is to take previously identified stably expressed genes, as done by Zhuo *et al.* (2016), this however does not ensure that the selected genes will show stable expression in the studied organism and conditions.

Here we propose a simple and fast method to identify the most stably expressed genes for each experimental condition. Our method is aimed at differential expression studies and represents a simple way to select custom reference genes for any species or any type of experiments, so they can be used in the normalization step of differential expression analysis algorithms, and does not necessitate spike-ins. It alleviates the problem inherent to predefined reference genes, which may not be stably expressed across experimental set-ups and are applicable to a single species.

Results

Initially three RNAseq transcriptomes were generated using *Arabidopsis* transgenic plants expressing GFP alone (control) or GFP-fused to fungal effector genes (Mlp37347 and Mlp124499). We tested the normalization of our RNAseq data using two sets of reference genes: commonly used reference genes (**Table 1**) and the 104 stably expressed *Arabidopsis* genes proposed by Zhuo *et al.* (2016). The first set of reference genes was assessed for stability in three different permutations of the transcriptome sets as shown in **Figure 1A** (panel 1: Mlp37347 vs Control, panel 2: Mlp124499 vs Control, panel 3: Mlp124499 vs Mlp37347). In each case, high levels of covariance, ranging from 4.9% (NDUFA8 in Mlp124499 vs Mlp37347) to 41.5% (tubulin 6 in Mlp124499 vs Mlp37347) were obtained. Next, we performed the same analysis using the 104 genes proposed by Zhuo *et al.* (2016). For the three permutations of the transcriptome sets, important fluctuations in the covariance were observed ranging from 2.9 to 49% (**Figure 1B**). Finally, we did the same for the set of 30 genes selected by Czechowski *et al.* (2005) for several plant tissues (**Additional File 1**). These results demonstrate that neither the commonly used reference genes, nor the 104 reference genes proposed by Zhuo *et al.* (2016) were stably expressed in our conditions.

Table 1. Common reference genes used in this study for comparison against custom selected reference genes.

Symbol	Name	ATG
Actin 2	ACT2	AT3G18780
Actin 7	ACT7	AT5G09810
Actin 8	ACT8	AT1G49240
Adenine phosphoribosyltransferase 1	APT1	AT1G27450
Elongation factor 1- α	EF1 α	AT5G60390
Eukaryotic translation initiation factor 4A-1	eIF4A	AT3G13920
NADH-ubiquinone oxidoreductase 19-kDa subunit	NDUFA8	AT5G18800
Tubulin β -2/ β -3 chain	TUB2	AT5G62690
β -tubulin 6	TUB6	AT5G12250
Tubulin β -9 chain	TUB9	AT4G20890

Symbol	Name	ATG
Polyubiquitin	UBQ4	AT5G20620
Ubiquitin extension protein	UBQ5	AT3G62250
Polyubiquitin	UBQ10	AT4G05320
Polyubiquitin	UBQ11	AT4G05050

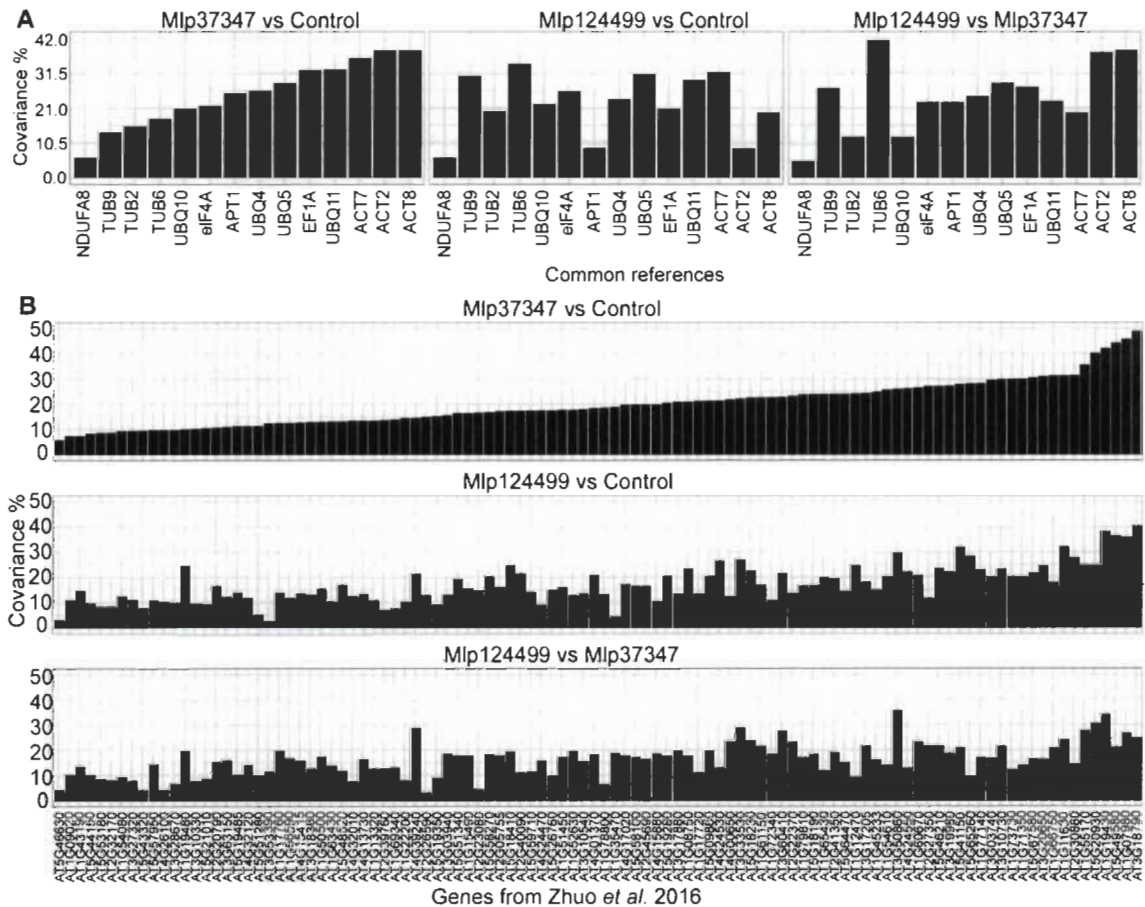


Figure 1. Evaluation of covariance distribution in the three transcriptome data sets a) among a set of 14 commonly used reference genes and b) a set of 104 reference genes proposed by Zhuo *et al.* (2016).

In order to search for more stably expressed genes, we developed a custom method to select reference genes using only one's own RNAseq data. We first used a R function to transform the count data into Transcripts per Million (Slowikowski, 2016) and calculate the average TPM and covariance for each gene. We then used the DAFS function (George & Chang, 2014) to calculate a cut-off for the exclusion of weakly expressed genes.

Finally, the 0.5% remaining genes with lowest covariance were selected as reference genes (R-package “CustomSelection” (dos Santos *et al.*, 2019)). This pipeline is thereafter referred to as the custom selection script.

To test the developed method, we used the same transcriptome sets described in **Figure 1** (the list of selected genes for each analysis is available in **Additional File 2 – Table S1**). For each transcriptome set, we show in **Figure 2** the average expression in \log_2 TPM and covariance of the common reference genes (Common), the set of 30 genes from Czechowski *et al.* (2005), the set of 104 genes from Zhuo *et al.* (2016) and the genes selected using the CustomSelection package (dos Santos *et al.*, 2019) (Custom script). In all pairings the custom selected reference genes show broader range of expression levels and lower covariance than the other sets. Next, we performed a differential expression analysis with DESeq2 (Love *et al.*, 2014) without control genes. We show in **Figure 3** the \log_2 -transformed fold change by the $-\log_{10}$ -transformed adjusted p-value for each gene set. We can see that the set of genes selected with the custom script shows lower fold change in all cases. We also compared the results of DESeq2 using no reference gene or the four sets indicated above for each permutation. As is shown in **Additional File 2 – Table S2**, in all the permutations the analysis without the use of references gives higher number of up-regulated genes than the analyses that use any of the reference sets while resulting in a lower number of down-regulated genes, possibly indicating a shift to downregulation that is not detected without reference genes.

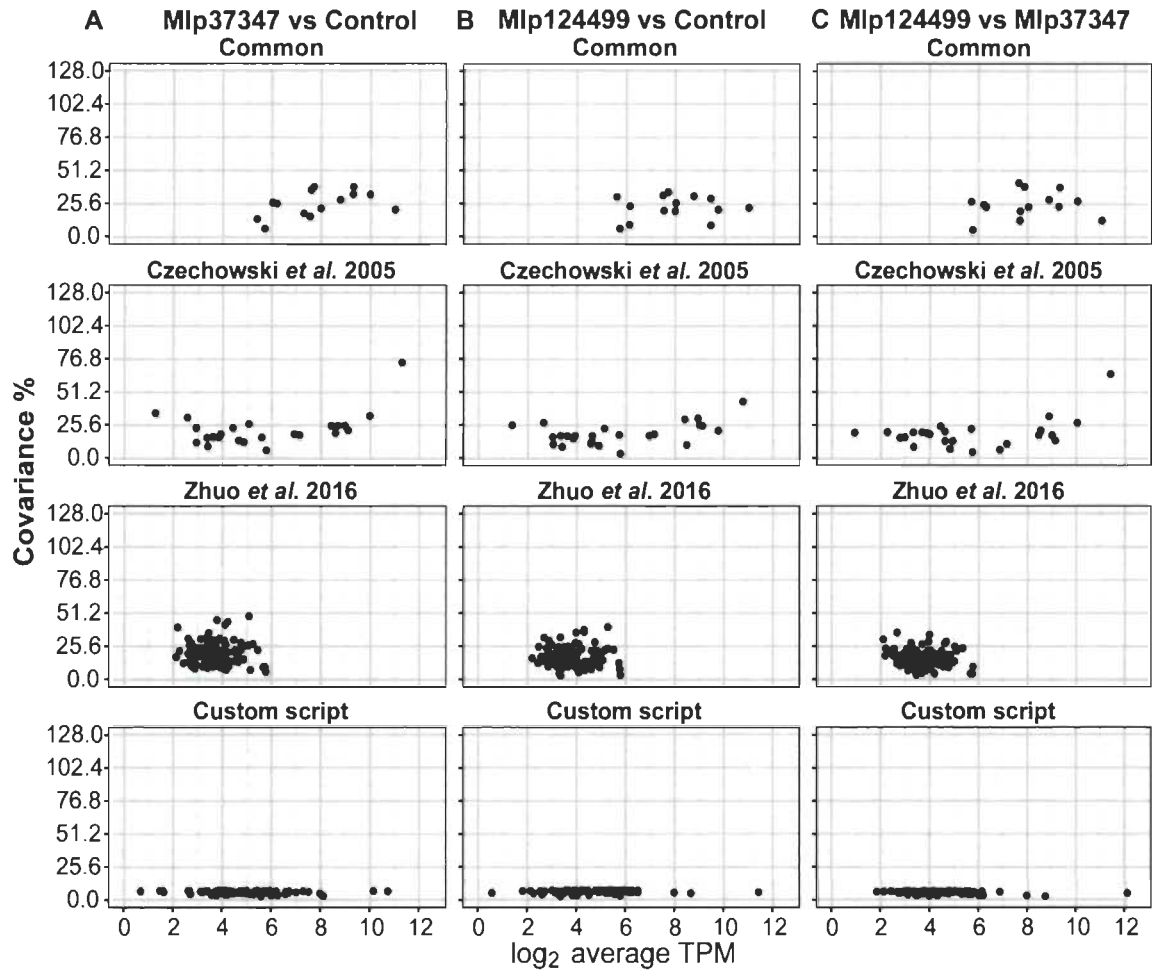


Figure 2. Comparison the four sets of reference genes in relation to covariance level and log₂ TPM for A) Mlp37347 vs Control, B) Mlp124499 vs Control and C) Mlp124499 vs Mlp37347.

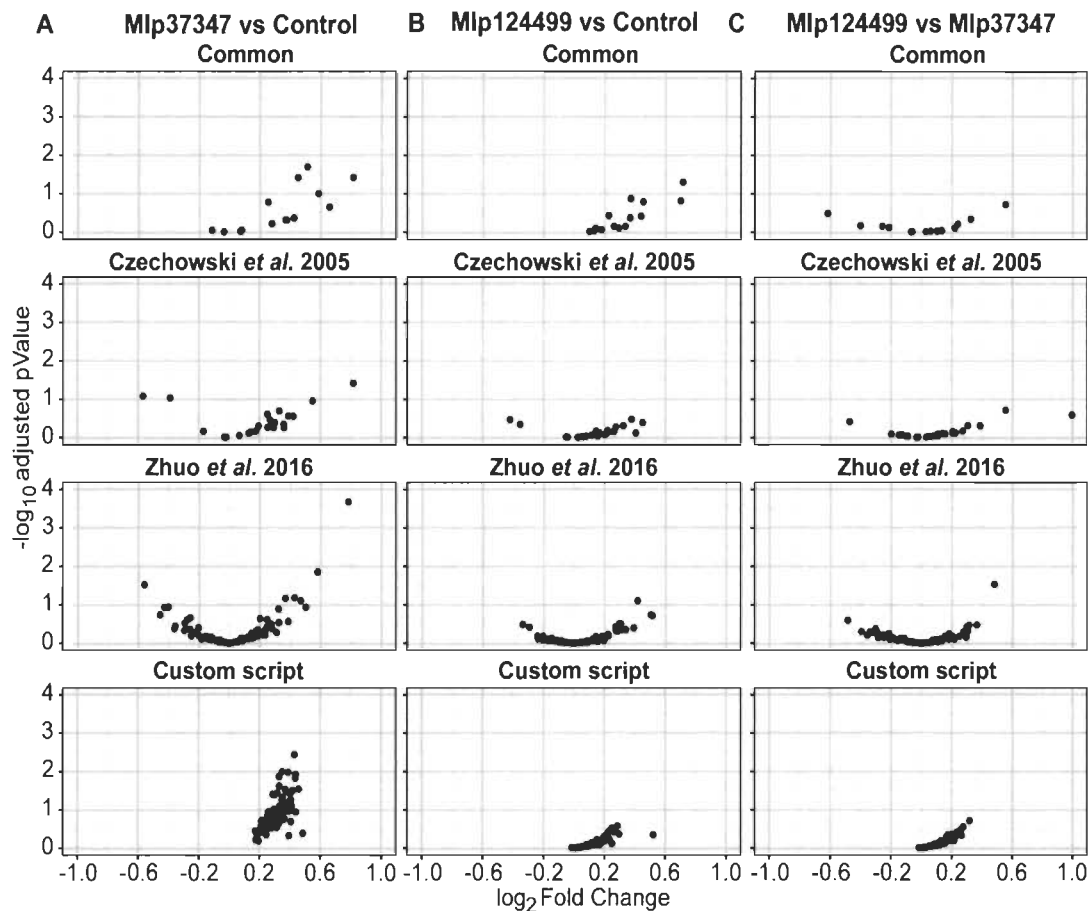


Figure 3. Comparison of the four sets of reference genes in relation the distribution of \log_2 fold Change by $-\log_{10}$ adjusted p-value for A) Mlp37347 vs Control, B) Mlp124499 vs Control and C) Mlp124499 vs Mlp37347.

To further test the stability of the custom reference genes in our experiment, we used NormFinder (Andersen *et al.*, 2004) and geNorm (Vandesompele *et al.*, 2002) to compare the four sets of reference genes using \log_2 transformed TPM values. The complete result is presented in the **Additional File 2 – Tables S3-S5**. We present in **Figure 4** the comparison of the set of common reference genes against the custom selected reference genes. The gene AT5G18800 (NDUFA8) which is in the set of common references was selected by the custom script in all three permutations and is shown with a purple border. Both sets of genes (custom and common references) were under the stability threshold of NormFinder (0.5), meaning that the software considers them suitable reference genes, however the custom selected genes (shown with a blue border) were more stable than the commonly used genes (shown in red, **Figure 4**). This was also the case for most genes tested with geNorm.

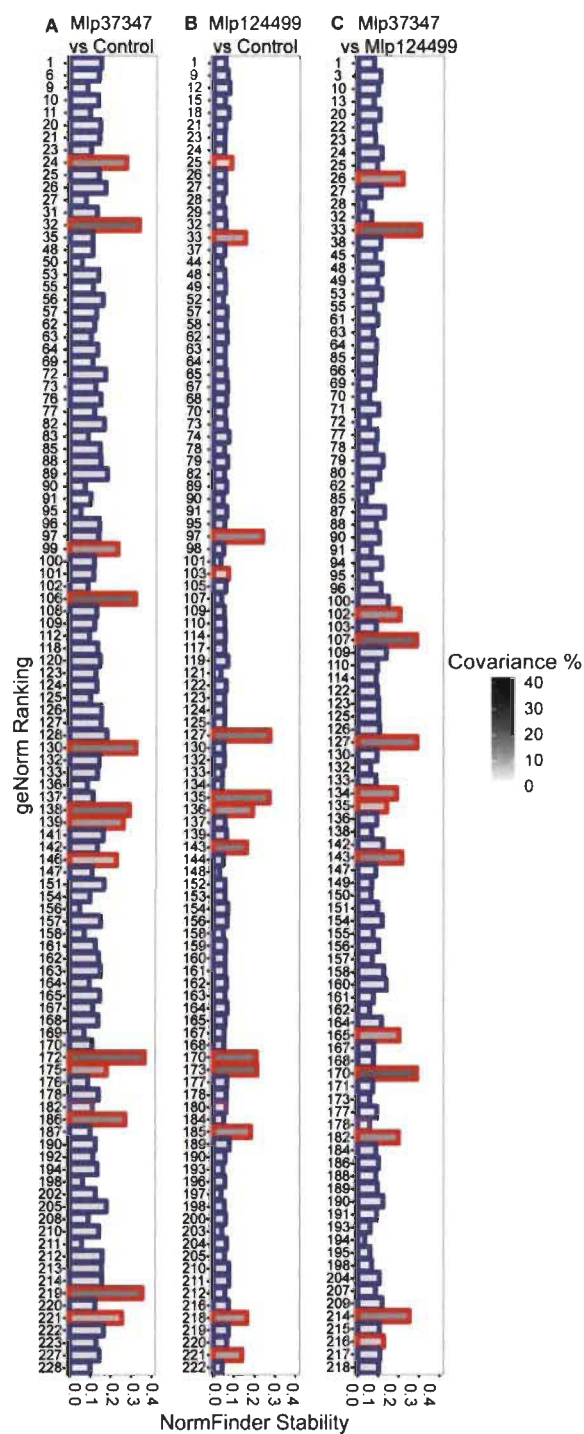


Figure 4. Comparison of custom selected reference genes (blue border) and commonly used reference genes (red border) with geNorm ranking, NormFinder stability index and covariance for a) Mlp37347 vs Control, b) Mlp124499 vs Control and c) Mlp124499 vs Mlp37347.

Discussion

The use of reference genes in RNAseq studies is suggested in the literature (Lovén *et al.*, 2012; Lutzmayer *et al.*, 2017; Taruttis *et al.*, 2018), yet the methods for the selection of these genes are designed for qPCR data and require a set of pre-selected reference or target genes or the selection of conditions similar to that of one's own experiment (Andersen *et al.*, 2004; Hruz *et al.*, 2008; Pfaffl *et al.*, 2004; Vandesompele *et al.*, 2002), which are not always available. As there is no previous transcriptomic study of plants constitutively expressing fungal effectors and since the information available on these effectors is scarce (Germain *et al.*, 2018), it is not possible to know a priori their function and which host genes are impacted by the presence of these fungal proteins. For these reasons, we propose a new R-package which enables the selection of custom reference genes regardless of the organism used or of the experimental conditions.

The method developed here only requires information available from the RNAseq analyses. It uses Transcripts per Million (Slowikowski, 2016) as a proxy for the expression level and the DAFS algorithm (George & Chang, 2014) to exclude genes with low counts, which may be inactive (Hart *et al.*, 2013). We first assessed whether the most commonly used reference genes (**Table 1**) or two sets of published reference genes for *Arabidopsis* (Czechowski *et al.*, 2005; Zhuo *et al.*, 2016) were indeed stably expressed in our experimental conditions. As demonstrated in **Figure 1** and **Additional File 1**, the three sets of reference genes show a high level of covariance in our experimental conditions, indicating that they were not suitable reference genes for our differential expression analysis.

Having a high level of variability in the expression of the reference genes results in skewed quantitative analysis and may cause the loss of some differentially expressed genes which show modest variation in gene expression (Gutierrez *et al.*, 2008). In relation to the reference gene sets, there is minimal overlap between sets published and the ones selected in this article (maximum of 5 genes shared between our set and the set of Zhuo *et al.* (2016) and 2 genes shared between our set and the set of Czechowski *et al.* (2005), shown in **Additional File 2 – Tables S3-S5**, column J). However, there is

extensive overlap in the deregulated genes (up- and down-regulated as shown in **Additional File 2 – Table S2**). This fact demonstrates that all three sets perform well in detecting deregulated genes, however having a reference gene set with lower covariance results in the finding of more de-regulated genes (**Additional File 2 – Table S2** downregulated) since more subtle deregulation can be detected.

Thus, to alleviate the bias inherent to the use of inappropriate reference genes, we devised a R-based pipeline to select custom reference genes for one's own experimental data. As presented in **Figure 2** and **Figure 3**, in all the pairings of the data used, the custom selected reference genes outperformed the other sets of reference genes in their expression stability, presenting lower fold changes and lower covariances. Our method allows the selection of genes more stably expressed and the selection of more genes as references (the final number is user defined, with the default setting being 0.5% of the expressed genes), giving more reference points, hence more robustness, to the normalization of genes expressed at different levels. The advantage of having a user-defined threshold is that when there is extensive variation in the data, a stringent threshold may result in the selection of few or no genes as references. On the contrary, extremely homogenous data would result in a very large reference gene set, for this reason a user-defined threshold is preferable.

Conclusions

Our results show the need for a new R-based pipeline for the selection of custom reference genes in transcriptomic studies. Our method can be applied to any organism and to any type of experimental conditions, and can easily be implemented or modified in R. This tool provides an alternative to spike-in controls and represents an improvement over pre-defined reference genes which may not be stably expressed in one's own experimental conditions.

Methods

Initial *Arabidopsis thaliana* Columbia-0 were obtained from *Arabidopsis* Biological Resources Center (ABRC). *Arabidopsis* transgenic plants expressing GFP alone (Control) or fused to a candidate secreted effector protein of the fungus *Melampsora larici-populina* (Mlp37347 or Mlp124499), obtained in our laboratory (Germain *et al.*, 2018), were used for the transcriptome analysis.

RNA was extracted from pooled aerial tissue of 2-week-old soil-grown plants, doing four replicates per genotype, with the Plant Total RNA Mini Kit (Geneaid) using RB buffer following manufacturer's protocol. The samples were treated with DNase, then RNA quality was assessed using agarose gel electrophoresis. Libraries were generated with the NeoPrep Library Prep System (Illumina) using the TruSeq Stranded mRNA Library Prep kit (Illumina) and 100 ng of total RNA following manufacturer's recommendations. The libraries were then sequenced with Illumina HiSeq 4000 Sequencer paired-end reads of 100 nt.

Libraries were trimmed using Trimmomatic (Williams *et al.*, 2016) (LEADING:4 TRAILING:4 SLIDINGWINDOW:4:20 MINLEN:20) and then the surviving paired reads were aligned to the TAIR10 assembly of the genome of *A. thaliana* with TopHat v2.0.14 (Kim *et al.*, 2013) in Galaxy (Afgan *et al.*, 2016) (default options, with average mate inner distance varying for each replicate (**Additional File 2 – Table S6.**) and standard deviation of mate inner distance of 50 base pairs). The general information of the sequencing results and mapping data is presented in **Additional File 2 – Table S6.**, the dataset was deposited in NCBI under BioProject PRJNA528094. Further analyses were done using R software v.3.2.5. Genomic ranges of *Arabidopsis* transcripts were obtained from Ensembl plants (Durinck *et al.*, 2005) with GenomicFeatures and overlaps of sequencing reads with the transcripts were counted using GenomicAlignments (Lawrence *et al.*, 2013), using options for paired-end reads and union mode.

We transformed the counts into TPM (Słowikowski, 2016) and calculated the cut-off for active genes with DAFS (George & Chang, 2014). We considered as reference

the 0.5% of the active genes with the lowest covariance (R package “CustomSelection” (dos Santos *et al.*, 2019)). Next, we used DESeq2 (Love *et al.*, 2014) to confirm that the selected genes were not deregulated. Finally, we used geNorm (Vandesompele *et al.*, 2002) and NormFinder (Andersen *et al.*, 2004) to compare the custom selected reference genes against three sets of genes (a list of 14 commonly used housekeeping reference genes (**Table 1**), the reference genes selected by Czechowski *et al.* (2005) and the 104 reference genes selected by Zhuo *et al.* (2016)), using TPM values for the expression levels.

Description of the R-package. This package has 4 functions, “Counts_to_tpm” (to convert read counts into TPM values using a named vector with gene lengths) and the read count data frame with the samples as the column names and the genes as row names, “DAFS” (uses the data frame of TPM values, first object of the result from “Counts_to_tpm” to get the threshold for expressed genes), “gene_selection” (uses the data frame of TPM and the result from “DAFS” output a data frame with the selected reference genes, their average TPM and the covariance of the TPM values) and “customReferences” (calculates internally “Counts_to_tpm”, “DAFS” and “gene_selection” outputs the result from “gene_selection”). The package also includes to datasets for testing: a data frame of counts created with the data used in this article and a named vector with the lengths of genes from *Arabidopsis*. A Wiki, which is the file README.md of this package, describes a workflow to get the read counts from raw read files.

Availability of data and materials

The dataset used herein was deposited in NCBI-SRA under BioProject PRJNA528094.

Acknowledgements

We thank Melodie B. Plourde for revising the manuscript.

References

- Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., . . . Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, 44(W1), W3-W10. doi: 10.1093/nar/gkw343
- Andersen, C. L., Ledet-Jensen, J., & Ørntoft, T. (2004). Normalization of real-time quantitative RT-PCR data: a model based variance estimation approach to identify genes suited for normalization applied to bladder- and colon-cancer data-sets. *Cancer Research*, 64(15), 5245-5250. doi: 10.1158/0008-5472.CAN-04-0496
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., & Scheible, W.-R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiology*, 139(1), 17. doi: 10.1104/pp.105.063743
- Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, 21(16), 3439-3440. doi: 10.1093/bioinformatics/bti525
- Evans, C., Hardin, J., & Stoebel, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*, 19(5), 776-792. doi: 10.1093/bib/bbx008
- George, N. I., & Chang, C. W. (2014). DAFS: a data-adaptive flag method for RNA-sequencing data to differentiate genes with low and high expression. *BMC Bioinformatics*, 15, 92. doi: 10.1186/1471-2105-15-92
- Germain, H., Joly, D. L., Mireault, C., Plourde, M. B., Letanneur, C., Stewart, D., . . . Séguin, A. (2018). Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Molecular Plant Pathology*, 19(1), 191-200. doi: 10.1111/mpp.12514
- Gutierrez, L., Mauriat, M., Guénin, S., Pelloux, J., Lefebvre, J. F., Louvet, R., . . . Van Wuytswinkel, O. (2008). The lack of a systematic validation of reference genes: A serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnology Journal*, 6(6), 609-618. doi: 10.1111/j.1467-7652.2008.00346.x
- Hart, T., Komori, H. K., LaMere, S., Podshivalova, K., & Salomon, D. R. (2013). Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics*, 14(1), 778. doi: 10.1186/1471-2164-14-778

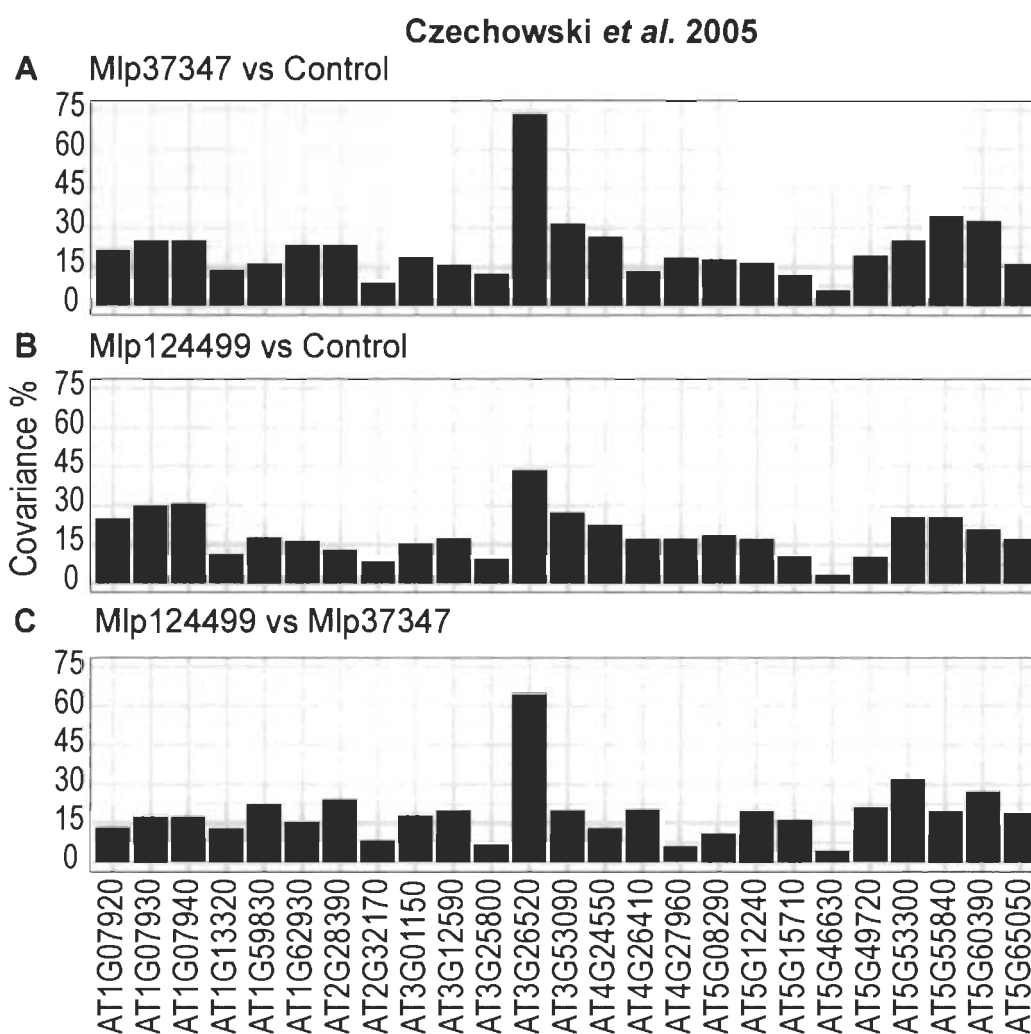
- Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., . . . Zimmermann, P. (2008). Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Advances in Bioinformatics*, 2008, 420747. doi: 10.1155/2008/420747
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, 14(4), R36. doi: 10.1186/gb-2013-14-4-r36
- Lawrence, G. J., Huber, M. L. W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., . . . Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, 9, e1003118. doi: 10.1371/journal.pcbi.1003118
- Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., & Ecker, J. R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell*, 133(3), 523-536. doi: 10.1016/J.CELL.2008.03.029
- Love, M. I., Anders, S., & Hu, W. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15, 550. doi: 10.1186/s13059-014-0550-8
- Lovén, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., . . . Young, R. A. (2012). Revisiting global gene expression analysis. *Cell*, 151(3), 476-482. doi: 10.1016/j.cell.2012.10.012
- Lutzmayer, S., Enugutti, B., & Nodine, M. D. (2017). Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-Seq data. *Scientific Reports*, 7, 5913. doi: 10.1038/s41598-017-06174-3
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7), 621-628. doi: 10.1038/nmeth.1226
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881), 1344-1349. doi: 10.1126/science.1158441
- Pachter, L. (2011). Models for transcript quantification from RNA-Seq [PREPRINT]. *arXiv*, doi: arXiv:1104.3889v2

- Paepe, K. D. (2015). *Comparison of methods for differential gene expression using RNA-seq data* (Thèse de Dissertation inédite). Universiteit Gent, Gand.
- Pfaffl, M. W., Tichopad, A., Prgomet, C., & Neuvians, T. P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnology Letters*, 26(6), 509-515. doi: 10.1023/B:BILE.0000019559.84305.47
- Qing, T., Yu, Y., Du, T., & Shi, L. (2013). mRNA enrichment protocols determine the quantification characteristics of external RNA spike-in controls in RNA-Seq studies. *Science China life sciences*, 56(2), 134-142. doi: 10.1007/s11427-013-4437-9
- Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., . . . Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome biology*, 14, 3158. doi: 10.1186/gb-2013-14-9-r95
- Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature biotechnology*, 32(9), 896-902. doi: 10.1038/nbt.2931
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, 11, R25. doi: 10.1186/gb-2010-11-3-r25
- Soneson, C., & Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14, 91. doi: 10.1186/1471-2105-14-91
- Taruttis, F., Feist, M., Schwarzfischer, P., Gronwald, W., Kube, D., Spang, R., & Engelmann, J. C. (2018). External calibration with *Drosophila* whole-cell spike-ins delivers absolute mRNA fold changes from human RNA-Seq and qPCR data. *BioTechniques*, 62(2), 53-61. doi: 10.2144/000114514
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, 3(7), research0034. doi: 10.1186/gb-2002-3-7-research0034
- Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, 131, 281-285. doi: 10.1007/s12064-012-0162-3

- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57-63. doi: 10.1038/nrg2484
- Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., . . . Bähler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, *453*, 1239-1245. doi: 10.1038/nature07002
- Williams, C. R., Baccarella, A., Parrish, J. Z., & Kim, C. C. (2016). Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics*, *17*, 103. doi: 10.1186/s12859-016-0956-2
- Wolf, J. B. W. (2013). Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular Ecology Resources*, *13*(4), 559-572. doi: 10.1111/1755-0998.12109
- Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One*, *9*(1), e78644. doi: 10.1371/journal.pone.0078644
- Zhuo, B., Emerson, S., Chang, J. H., & Di, Y. (2016). Identifying stably expressed genes from multiple RNA-Seq data sets. *PeerJ*, *4*, e2791. doi: 10.7717/peerj.2791

Supplementary material

Tables S1, S2, S3, S4 and S5 (Additional File 2) are not shown here, but are available in xls format at: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-019-6426-2>



Additional File 1. Covariance level for each of the 30 genes selected by Czechowski *et al.* (2005) for each permutation (A: Mlp37347 vs Control; B: Mlp124499 vs Control; C: Mlp124499 vs Mlp37347).

Additional File 2 – Table S1. TAIR IDs of custom selected references for each transcriptome permutation.

Additional File 2 – Table S2. DESeq2 results summary of analysis without reference genes or with different reference sets (Custom selected, from Czechowski *et al.* (2005), from Zhuo *et al.* (2016) or Commonly used references). Table presents the number of genes found up- and down-regulated in two analyses.

Additional File 2 – Table S3 to S5. Summary of the results of several analyses for all the genes evaluated in this article: Column A: TAIR ID; Column B: ranking calculated with geNorm with the function “selectHKs” from the R package “NormqPCR”; Column C: average TPM value; Column D: covariance of the TPM values; Column E: the difference of expression of a gene between two samples calculated with NormFinder; Column F: the common standard deviation of the expression of a gene between two samples calculated with NormFinder; Column G: stability measure from NormFinder; Column H: log2-transformed fold change of each gene calculated with DESeq2 without using reference genes; Column I: adjusted p value of the gene deregulation calculated with DESeq2 without using reference genes; Column J: sources that identified the gene as a reference, when more than one source selected the gene as reference they are separated by a “;”.

Additional File 2 – Table S3. Permutation Mlp37347 vs Control;

Additional File 2 – Table S4. Permutation Mlp124499 vs Control;

Additional File 2 – Table S5. Permutation Mlp124499 vs Mlp37347.

Additional File 2 – Table S6. Metadata of samples used: replicate identification, number of sequenced reads, average length of the separation between two paired reads, number of reads after trimming and filtering and number of aligned reads for each of the 4 replicates of the three samples used in this study.

Sample name	#Reads	Mate inner distance	#Surviving Reads	#Aligned Reads
Control_1	12 959 539	50	11 420 407	6 578 769
Control_2	17 628 091	60	16 441 057	15 326 908
Control_3	17 505 072	55	15 952 366	14 251 736
Control_4	44 449 775	84	40 740 349	38 447 039
Mlp37347_1	27 784 075	115	26 345 883	25 136 138
Mlp37347_2	23 439 580	110	21 914 868	20 722 973
Mlp37347_3	14 033 172	110	13 268 922	12 674 770
Mlp37347_4	16 264 602	70	14 934 782	13 919 261
Mlp124499_1	12 422 097	65	11 615 353	10 818 003
Mlp124499_2	14 079 283	65	13 059 459	11 939 187
Mlp124499_3	16 445 351	60	14 910 023	13 122 045
Mlp124499_4	16 980 712	65	15 844 378	14 703 575

#Reads = sequenced reads.

#Surviving Reads = number of reads left after trimming.

Aligned Reads = number of reads mapped against TAIR10 assembly of *Arabidopsis*.

CHAPTER III

DIFFERENTIAL ALTERATION OF PLANT FUNCTIONS BY HOMOLOGOUS FUNGAL CANDIDATE EFFECTORS

Karen Cristine Gonçalves dos Santos, Gervais Pelletier, Armand Séguin,
François Guillemette, Jeffrey Hawkes, Isabel Desgagné-Penix, Hugo Germain

Preprint available at bioRxiv (doi: 10.1101/2020.10.30.363010) on October 30th 2020

3.1 Author contributions

dos Santos: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – Original draft preparation, Writing – Review & editing.

Pelletier: Methodology, Resources, Writing – Review & editing.

Séguin: Resources, Writing – Review & editing.

Guillemette: Methodology, Resources, Software, Writing – Review & editing.

Hawkes: Methodology, Resources, Writing – Review & editing.

Desgagné-Penix: Conceptualization, Methodology, Resources, Supervision, Writing – Review & editing.

Germain: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – Review & editing.

3.2 Résumé de l'article

Les champignons de la rouille sont des agents pathogènes des plantes qui provoquent des épidémies menaçant la production d'espèces végétales importantes, telles que le blé, le soja, le café et le peuplier. *Melampsora larici-populina* (*Mlp*) provoque la rouille du peuplier et code pour au moins 1 184 effecteurs candidats (CE), mais leurs fonctions sont mal connues. Dans cette étude, nous avons utilisé des plantes *Arabidopsis* exprimant de manière constitutive les CEs de *Mlp* pour découvrir les processus ciblés par ces protéines fongiques. À cette fin, nous avons séquencé le transcriptome et utilisé la spectrométrie de masse pour analyser le métabolome des plantes d'*Arabidopsis* exprimant individuellement l'un des 14 EC sélectionnés et d'une lignée témoin. Nous avons trouvé 2 299 gènes dérégulés au cours de l'expérience. Parmi les gènes dérégulés, les voies KEGG « voie de signalisation MAPK » et « interaction plante-pathogène » étaient respectivement surreprésentées dans six et cinq des 14 lignées transgéniques. De plus, les gènes liés à la réponse hormonale et à la défense étaient régulés à la baisse dans toutes les lignées transgéniques. Nous avons en outre observé qu'il y avait 680 métabolites dérégulés dans au moins une lignée transgénique exprimant la CE, avec des composés phénoliques et hautement insaturés enrichis en métabolites régulés à la hausse et des peptides enrichis parmi les métabolites régulés à la baisse. Il est intéressant de noter que nous avons constaté que les lignées transgéniques exprimant des CEs non apparentés avaient des schémas de dérégulation des gènes et des métabolites corrélés, alors que l'expression des CEs appartenant à la même famille dérégulait des gènes et des métabolites différents. Pris ensemble, nos résultats indiquent que la séquence des effecteurs et leur appartenance à des familles ne sont peut-être pas un bon prédicteur de leur impact sur la plante.

Importance : Les champignons de la rouille sont des agents pathogènes des plantes qui menacent la production de cultures importantes, notamment le blé, le soja, le café et le peuplier. Les effecteurs sont utilisés par les agents pathogènes pour contrôler l'hôte, cependant dans le cas de *Melampsora larici-populina*, l'agent causal de la rouille du peuplier, et d'autres champignons de la rouille, ces protéines sont mal connues. Nous avons utilisé des plantes d'*Arabidopsis* exprimant des effecteurs candidats (CE) de

Mlp pour mieux comprendre l'interaction entre cet agent pathogène et ses hôtes. Nous avons constaté que l'expression des CEs non apparentés conduisait à des schémas similaires de dérégulation des gènes et des métabolites, tandis que les lignées transgéniques exprimant des CEs appartenant à la même famille montraient différents groupes de gènes et de métabolites dérégulés. Ainsi, nos résultats suggèrent que l'annotation fonctionnelle des effecteurs basée sur la similarité des séquences peut être trompeuse.

3.3 Full article in English: Differential alteration of plant functions by homologous fungal candidate effectors

Abstract

Rust fungi are plant pathogens that cause epidemics that threaten the production of important plant species, such as wheat, soy, coffee and poplar. *Melampsora larici-populina* (Mlp) causes the poplar rust and encodes at least 1 184 candidate effectors (CEs), however their functions are poorly known. In this study, we used *Arabidopsis* plants constitutively expressing CEs of *Mlp* to discover processes targeted by these fungal proteins. For this purpose, we sequenced the transcriptome and used mass spectrometry to analyze the metabolome of *Arabidopsis* plants expressing individually one of the 14 selected CEs and of a control line. We found 2 299 deregulated genes across the experiment. Among the down-regulated genes, the KEGG pathways “MAPK signaling pathway” and “Plant-pathogen interaction” were respectively over-represented in six and five of the 14 transgenic lines. Moreover, genes related to hormone response and defense were down-regulated across all transgenic lines. We further observed that there were 680 metabolites deregulated in at least one CE-expressing transgenic line, with highly unsaturated and phenolic compounds enriched in up-regulated metabolites and peptides enriched among down-regulated metabolites. Interestingly, we found that transgenic lines expressing unrelated CEs had correlated patterns of gene and metabolite deregulation, while expression of CEs belonging to the same family deregulated different genes and metabolites. Taken together, our results indicate that the sequence of effectors and their belonging to families may not be a good predictor of their impact on the plant.

Importance: Rust fungi are plant pathogens that threaten the production of important crops, including wheat, soy, coffee and poplar. Effectors are used by pathogens to control the host, however in the case of *Melampsora larici-populina*, the causal agent of the poplar rust, and other rust fungi these proteins are poorly known. We used *Arabidopsis* plants expressing candidate effectors (CEs) of *Mlp* to better understand the interaction between this pathogen and its hosts. We found that expression of unrelated CEs led to similar patterns of gene and metabolite deregulation, while transgenic lines expressing CEs

belonging to the same family showed different groups of different genes and metabolites deregulated. Thus, our results suggest that functional annotation of effectors based on sequence similarity may be misleading.

Keywords: Transcriptome, Metabolome, Plant-microbe interactions, Rust fungi, Effector biology, *Melampsora larici-populina*.

Introduction

Plants have to defend themselves against different types of pathogens. Their first line of defense consists of passive barriers, such as the cuticle and cell wall, which prevent pathogens from entering the plant tissue and its cells. Upon successful entry of a pathogen, conserved pathogenic motifs, called Microbe-Associated Molecular Patterns (MAMPs), may be detected and activate the Pattern-Triggered Immunity (PTI) (Henry *et al.*, 2012). PTI includes the transient accumulation of reactive oxygen species (ROS), callose deposition, alteration of hormone networks and activation of defense genes (Bigeard *et al.*, 2015; Luna *et al.*, 2011). Finally, microorganisms secrete effectors into their host to modulate the host metabolism in favor of the pathogen. If detected, these effectors will activate the Effector-Triggered Immunity (ETI), leading to plant cell death in order to avoid pathogen spreading to surrounding cells (Jones & Dangl, 2006).

Rust fungi are the largest group of fungal plant pathogens, infecting ferns, gymnosperms and angiosperms and causing important losses in food production (Aime *et al.*, 2017; Dean *et al.*, 2012). They are obligate biotrophs, produce two to five types of spores and infect one or two unrelated species to complete their life cycle (Aime *et al.*, 2017). To guard themselves against the defense mechanism of two different host species and to be able to feed on them, rust fungi deploy a large arsenal of effectors. To better comprehend the interaction between these pathogens and their hosts, and to provide new mechanisms to target in order to improve plant immunity, it is imperative that we understand how these effectors are secreted into host cells, how they evolve and how they act to promote pathogen growth (Dangl *et al.*, 2013; Hogenhout *et al.*, 2009).

While the precise number of *bona fide* effectors carried by each rust fungi species is unknown, Duplessis and colleagues (Duplessis *et al.*, 2011a) established that the poplar rust (*Melampsora larici-populina*) genome encodes 1 184 small secreted proteins (SSPs) whereas the wheat stripe rust (*Puccinia graminis* f. sp. *tritici*) genome encodes 1 106 SSPs (Duplessis *et al.*, 2011a), which are considered candidate effectors (CEs). These CEs are grouped within families based on sequence homologies (Enright *et al.*, 2002; Saunders *et al.*, 2012). Furthermore, effectors in the same family have been shown to interact with homologous R-proteins (Ravensdale *et al.*, 2010), however the virulence function of these effectors has seldom been investigated.

Previous studies have proposed different criteria to screen the genome of plant pathogenic fungi for high-priority CEs, including having less than 300 amino acids, high cysteine content, being expressed in infection structures during host infection or being detected in the host tissue during infection (Lorrain *et al.*, 2015; Saunders *et al.*, 2012; Sperschneider *et al.*, 2015). Once identified, putative effectors must be functionally characterized. In pathogens that are not obligate biotrophs, this can be achieved by silencing or overexpressing the gene encoding the CEs and analysing the outcome of an infection (Li *et al.*, 2015; Lyu *et al.*, 2016). For rust fungi and other obligate biotrophs, which are not amenable to genetic transformation, this direct investigative approach is not possible. The alternative solution proposed by different research groups is to use heterologous systems, either by transforming model plants to express the CE-encoding gene or by infecting model plants with pathogens able to express these genes (Chaudhari *et al.*, 2014; Lorrain *et al.*, 2018b). This way, it is possible to evaluate if immunity is compromised, as it was shown that effectors expressed in heterologous systems conserve their capacity to alter the plant's susceptibility to pathogens (Ahmed *et al.*, 2018; Bentem *et al.*, 2005; Germain *et al.*, 2018; Houterman *et al.*, 2009; Jamir *et al.*, 2004; Pitino *et al.*, 2016). The stable and transient expression of CEs from *M. larici-populina* in *Arabidopsis thaliana* and *Nicotiana benthamiana*, from *Phakopsora pachirhyzi* in *N. benthamiana* and from *Hyaloperonospora arabidopsidis* in *A. thaliana* allowed the study of their subcellular localization *in planta*, their impact on the growth of different

pathogens and the search for host proteins potentially targeted by CEs (Caillaud *et al.*, 2012; Germain *et al.*, 2018; Kunjeti *et al.*, 2016; Petre *et al.*, 2015).

Still, the impact of CEs in the plant may not be easy to detect or the isolated effect of a single CE may be too subtle to affect pathogen growth. In the study of Germain and colleagues, 14 CEs impacted the growth of *H. arabidopsidis* or *Pseudomonas syringae* pv *tomato*. Eleven of the analyzed CEs displayed nucleocytoplasmic localization *in planta*, providing very limited information on possible host targets or helpers of these protein (Germain *et al.*, 2018). Petre and colleagues found seven CEs of wheat yellow rust fungus (out of 16) with specific accumulation pattern in plant cells (other than nucleocytoplasmic) and discovered specific plant protein interactors for six CEs (Petre *et al.*, 2016). Only three of the 16 CEs studied had both specific accumulation pattern in *N. benthamiana* cells and specific plant protein interactors. Although the pathogen growth readout is informative regarding the impairment of the immune pathway, it is opaque with regards to which pathway has been tampered with or which metabolites are off-balance. Transcriptomic and metabolomic studies of stable transgenic plants expressing CEs have been useful in these cases, since they allow the detection of more subtle changes, unlikely to have a quantifiable impact on pathogen growth on their own (Ahmed *et al.*, 2018; Madina *et al.*, 2020; Plett *et al.*, 2011).

Here we studied the transcriptome and metabolome of 14 transgenic *Arabidopsis* plant lines expressing Mlp CEs known to affect plant susceptibility to pathogen. We identified 2 299 deregulated genes using this approach, including many related to response to biotic and abiotic stress, metabolism of specialized metabolites and plant development. Four lines expressing CEs from different families showed correlated patterns of gene deregulation demonstrating that the current grouping based on sequence homology does not reflect the virulence function of these CEs. We also found important down-regulation of highly unsaturated and phenolic compounds and up-regulation of peptides in almost all CE-overexpressing lines. Overall, our results show a lack of correlation between the sequence similarity of the studied CEs and their overall deregulation of genes or metabolites. Taken together, our results demonstrate that CEs

that have completely different sequences can alter the expression of the same genes sets, while CEs of the same family can target completely different gene sets. Therefore, it is not possible to estimate the function of a CE, its impact on the transcriptome or on the metabolome of the plant, based solely on its sequence, or its similarity to another CE.

Results

In planta expression of candidate fungal effectors results in important deregulation at the transcriptome level

Melampsora larici-populina CEs have been previously studied in heterologous systems for functional characterization (Ahmed *et al.*, 2018; Gaouar *et al.*, 2016; Germain *et al.*, 2018; Madina *et al.*, 2020; Petre *et al.*, 2015). In **Table 1**, we present features of the 14 CEs studied here. Mlp37347 is a homolog of the well studied AvrL567 group from *M. lini* (Dodds *et al.*, 2004; Gan *et al.*, 2010), and accumulates at the plasmodesmata in *Arabidopsis*. Mlp72983 accumulates in the chloroplast (Germain *et al.*, 2018) and Mlp124357 is found in the tonoplast and was shown to interact with *Arabidopsis* and poplar Protein Disulfide Isomerase (Madina *et al.*, 2020). The other 11 CEs selected here have nucleocytoplasmic accumulation, the same as the marker protein GFP used. Although information about these CEs is scarce, all of them impacted *Arabidopsis* susceptibility to either *Pseudomonas syringae* or to *Hyaloperonospora arabidopsidis*.

Table 1. Features of the CEs investigated in this study.

CE	Length (Cysteine)	Family (members)	Subcellular localization ^a	U, P, B, L ^{b,c}
Mlp37347	151 (2)	-	Plasmodesmata	E, HE, E, E
Mlp72983	220 (8)	CPG332-CPG333(13)	Chloroplast	E, HE, E, HE
Mlp102036	107 (0)	CPG2528(5)	Nucleocytoplasmic	E, HE, E, E
Mlp106078	137 (10)	-	Nucleocytoplasmic	E, HE, E, E
Mlp123218	209 (6)	CPG543(7)	Nucleocytoplasmic	E, HE, E, E
Mlp123227	124 (3)	CPG1059(2)	Nucleocytoplasmic	E, HE, E, HE
Mlp123531	102 (8)	CPG4557(3)	Nucleocytoplasmic	E, HE, E, E

CE	Length (Cysteine)	Family (members)	Subcellular localization ^a	U, P, B, L ^{b,c}
Mlp124256	89 (6)	CPG5464(13)	Nucleocytosolic	N, N, E, E
Mlp124266	92 (7)	CPG5464(13)	Nucleocytosolic	N, N, E, E
Mlp124357	98 (6)	CPG4890	Tonoplast	N, N, E, E
Mlp124466	76 (0)	-	Nucleocytosolic	-
Mlp124497	77 (4)	CPGH1(33)	Nucleocytosolic	N, N, N, N
Mlp124499	72 (3)	CPGH1(33)	Nucleocytosolic	N, N, E, HE
Mlp124518	76 (3)	CPGH1(33)	Nucleocytosolic	N, N, E, E

^a Subcellular localization was evaluated in *Arabidopsis* (Germain *et al.*, 2018). ^{b,c} U, P, B, L refers to expression on: U) urediniospores, P) poplar leaves, B) basidiospores or L) larch needles (Lorrain *et al.*, 2018a), where E, HE, and N indicate that the CE is expressed, highly expressed, or was not detected, respectively, and - indicates no data is available.

To better understand the mechanism through which these 14 CEs impact plants, we studied the transcriptome and the metabolome of transgenic *Arabidopsis* plants constitutively expressing them. In total, we found 2 299 differentially expressed genes (DEGs) across the experiment. However, the number of DEGs in each line was variable, from 84 in Mlp106078 to 898 DEGs in Mlp123531 (**Figure 1**), indicating each CE affects the plant transcriptome to a different degree. The list of deregulated genes in each transgenic line is available in **Table S1**. We further assessed if the level of transgene expression could explain the number of DEGs in each sample and plotted the number of deregulated genes per transgenic line against expression level (in transcript per million) of the CE:GFP fusion transcripts. Linear regression shows a poor relation between the two ($R^2 = 0.1016$, **Figure S1**) suggesting that the number of deregulated genes per line depends more on the identity of the expressed CE than on the strength of its expression.

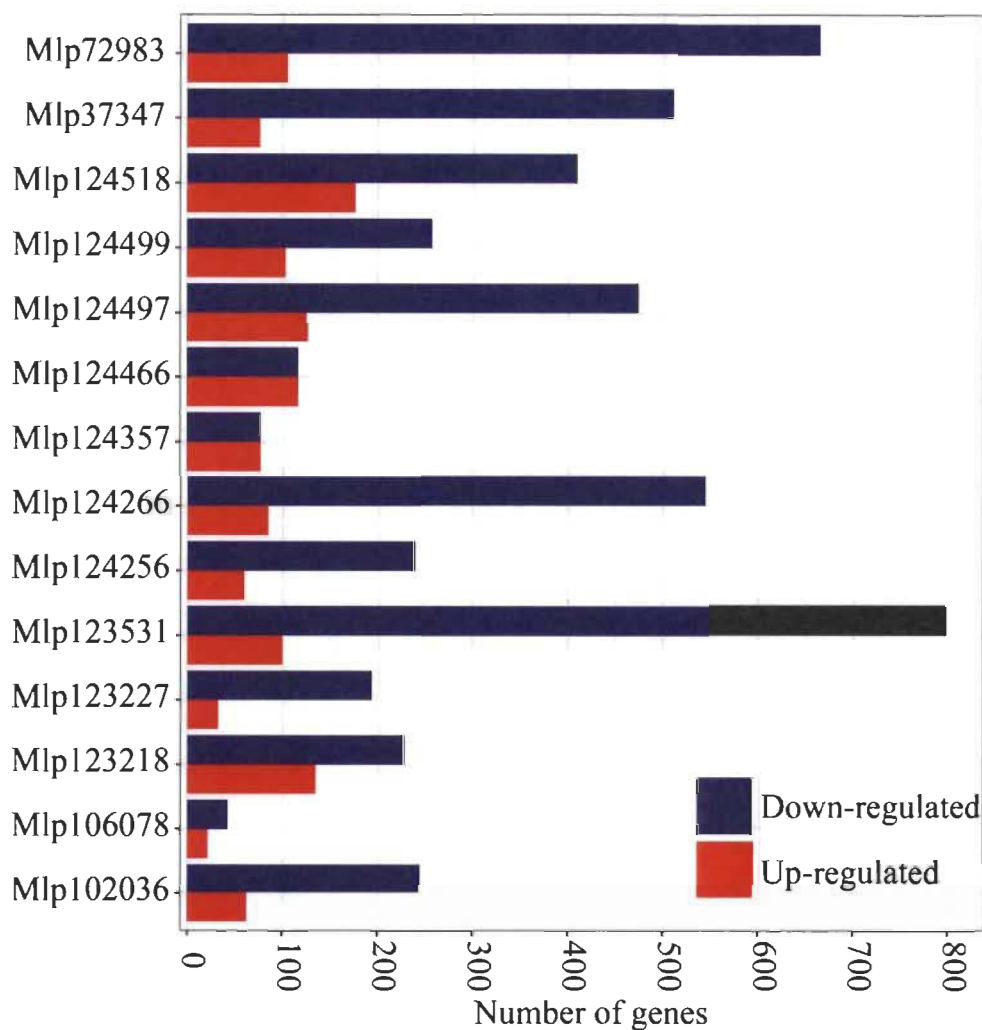


Figure 1. *In planta* expression of candidate fungal effector results in important deregulation at the transcriptome level.

Blue and red bars indicate the number of down- and up-regulated genes, respectively, in each CE-expressing transgenic line compared to the control line. The underlying data for this figure can be found at (dos Santos *et al.*, 2020d).

Hierarchical clustering based on gene expression groups effectors independently of amino acid sequence homology

CEs are typically grouped into families based on their amino acid sequences (Saunders *et al.*, 2012) and it has been shown that R-protein recognize related effectors (Ravensdale *et al.*, 2010). Nevertheless, the virulence activity of effectors from the same family has rarely been studied. To search for gene deregulation patterns of related and unrelated CEs, we used WGCNA to cluster the co-expressed DEGs and Pearson's correlation coefficient to cluster the transgenic lines (**Figure 2**). We found in total 208 GO

terms enriched in the gene sets from WGCNA. A summary is presented in **Table 2**, and the full list of enriched terms is available at (dos Santos *et al.*, 2020d). Set 0 clusters 714 genes deregulated across the 14 transgenic lines, 63.17% of which were down-regulated. Functions enriched in this gene set are related to defense, specialized metabolism, stress, and signaling pathways. Set 1 is composed of down-regulated genes enriched in GO terms related to defense responses and all transgenic lines have down-regulated genes in this set. Of the 379 genes in Set 2, 76.5% were down-regulated and this set is enriched GO terms related to specialized metabolite biosynthesis. In the case of Set 3, 81.8% of the genes were down-regulated, but we did not find enriched GO terms in this gene set. Interestingly, this set is composed of genes with the same pattern of deregulation in 4 transgenic lines expressing effectors without sequence similarity (Mlp72983, Mlp102036, Mlp123218, and Mlp123531, **Table S2**) which accumulate in two separate cell compartments (**Table 1**). Set 4 is related to metabolism and abiotic stress and 77.6% of its genes were down-regulated. Sets 5, 6, and 7 are composed almost exclusively of up-regulated genes (**Table 2**). Set 5 has genes deregulated in most transgenic lines that are related to abiotic stress and development. Set 6 is comprised of up-regulated genes almost exclusively found in the transgenic line Mlp124466 and related to transcription, vascular histogenesis, and response to different types of stress. Finally, Set 7 is made of genes related to photosynthesis and deregulated in the lines Mlp124256 and Mlp124518. In the cases of the Sets 0, 2, 3 and 4, there is mix of genes up and down-regulated, thus the enriched GO terms may be either up or down-regulated, or both. Interestingly, the dendrogram at the top of **Figure 2** shows that CEs belonging to the same family (Mlp124497, Mlp124499 and Mlp124518; Mlp124256 and Mlp124266) fall in separate clusters despite their similarity at the amino acid level (**Table S2**).

Table 2. Summary of “biological process” GO terms enriched in the WGCNA gene sets.

Set	Genes in the set	Up-regulated ^a	Down-regulated ^a	Enriched GO terms
Set 0	714	262	451	<p>Response to water deprivation</p> <p>Cold acclimation; Leaf senescence</p> <p>Response to fungus, to chitin, to ROS</p> <p>Response to salt stress and to hypoxia</p> <p>Defense response to fungus</p> <p>Response to toxic substance</p> <p>Response to nitrogen compound and to ET</p> <p>Isoprenoid, triterpenoid and terpenoid biosynthesis</p> <p>Plant-type cell wall loosening</p> <p>Phosphorelay signal transduction system</p>
Set 1	624	10	615	<p>Response to drug, nitrogen, ROS and ozone</p> <p>Response to SA, JA and karrikin</p> <p>Response to wounding, to herbivore and insect</p> <p>Cellular response to light stimulus and hypoxia</p> <p>Cellular response to acid chemical</p> <p>Defense response (incompatible interaction)</p> <p>Defense response by callose deposition in cell wall</p> <p>Defense response by cell wall thickening</p> <p>SAR and ISR</p> <p>Camalexin, indole phytoalexin and SA biosynthesis</p> <p>Sulfur compound biosynthesis</p> <p>Toxin and phenol-containing compound biosynthesis</p>
Set 2	379	89	290	<p>Response to karrikin, to nutrient levels and to copper ion</p> <p>S-glycoside and unsaturated fatty acid biosynthesis</p> <p>Chlorophyll biosynthesis</p> <p>Tetraterpenoid, terpenoid and carotenoid biosynthesis</p> <p>Isoprenoid, glycosyl and xanthophyll metabolism</p> <p>Sulfur compound, cofactor and leucine biosynthesis</p> <p>Defense response to insect</p> <p>De-etiolation; Chloroplast organization</p>
Set 3	253	47	207	No GO term enriched

Set	Genes in the set	Up-regulated ^a	Down-regulated ^a	Enriched GO terms
Set 4	140	32	109	Response to water deprivation Response to salt stress and to starvation Cellular amino acid catabolism/metabolism ET-activated signaling pathway Indole-containing compound metabolism
Set 5	116	113	4	Circadian rhythm; Starch catabolism Response to cold Regulation of reproductive process Regulation of post-embryonic development
Set 6	40	38	2	Response to hypoxia and to wounding, Response to drug, to chitin and to salt stress Transcription; Phloem or xylem histogenesis
Set 7	32	32	0	Photosynthesis; Proton transmembrane transport

^a Up- and down-regulated indicate the number of genes in the set the are up- or down-regulated in at least one transgenic line, thus there may be genes that are deregulated in both directions in the set because they are deregulated in opposite directions in different samples.

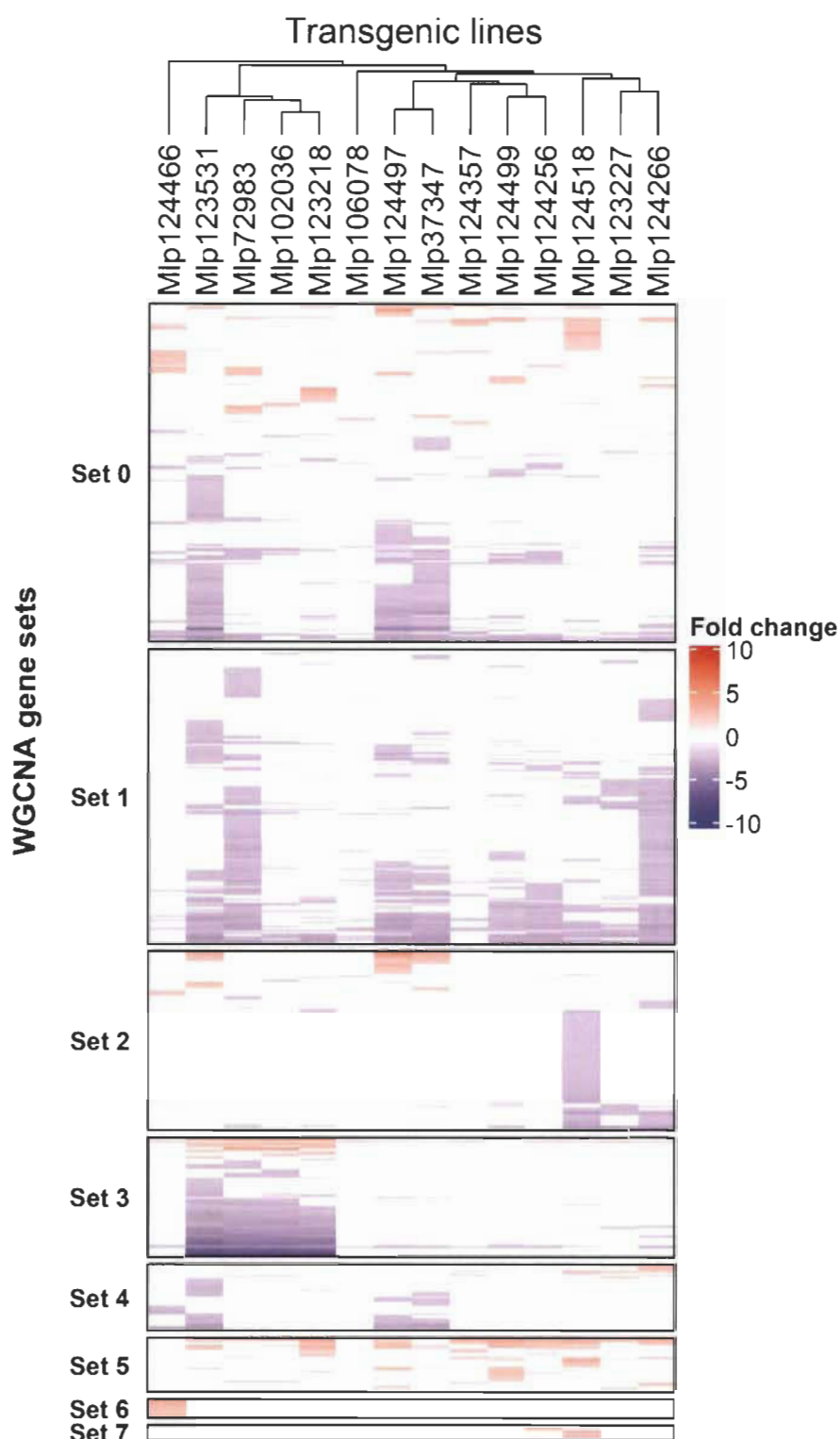


Figure 2. Heatmap of genes deregulated in each CE-expressing transgenic line.

Transgenic lines are displayed as columns and deregulated genes as lines. Sets of co-expressed genes (Sets 0 to 7) were calculated with WGCNA. Transgenic lines were grouped by correlation of gene deregulation using Pearson's correlation coefficient. The underlying data for this figure can be found at (dos Santos *et al.*, 2020d).

To analyze the relation between the sequence of each effector and its influence on the plant transcriptome, we compared the sequence alignment dendrogram to the differential expression dendrogram. After removal of the signal peptide, we aligned the sequences of the studied CEs, and compared the resulting dendrogram with the one obtained from the gene deregulation correlation (**Figure 3**). Pearson's correlation showed that transgenic lines expressing CEs from different families had correlated patterns of gene deregulation. Only one cluster was present in both dendrograms, Mlp102036 and Mlp123218, however this grouping is not supported in the effector sequence dendrogram (bootstrap value 8%) while it is in the gene deregulation dendrogram (bootstrap 100%). This analysis indicates that the sequence similarity between the CEs is not a good predictor of the impact they have on plant gene expression.

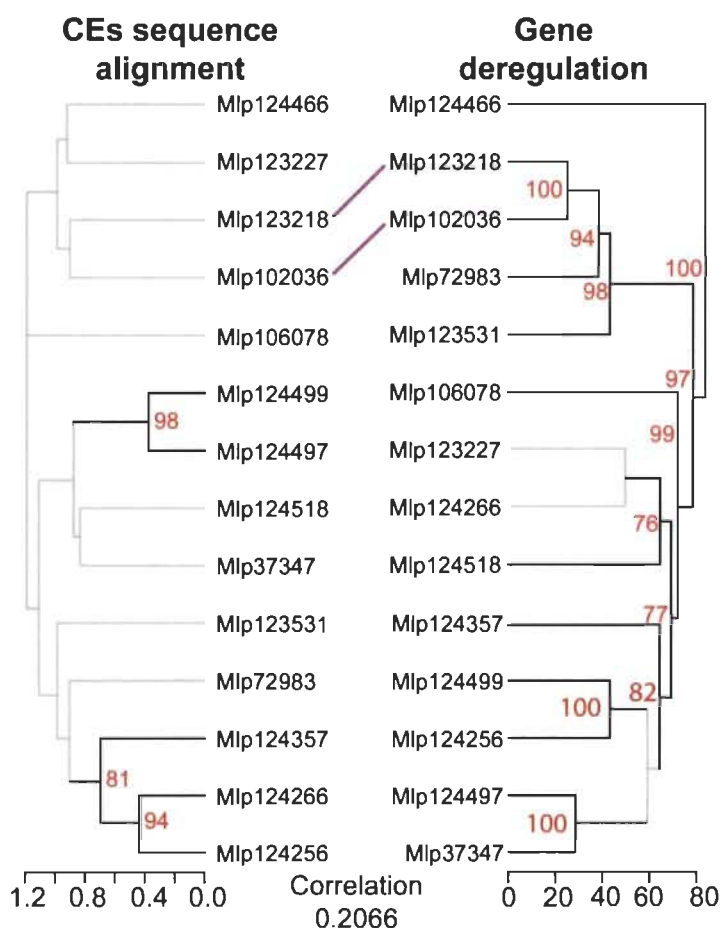


Figure 3. Hierarchical clustering of gene deregulation groups effectors independently of amino acid sequence homology.

CE sequence alignment was computed with Muscle alignment and tree (left) was calculated with UPGMA. Dendrogram based on correlation of gene deregulation (right) was calculated with

Pearson's correlation coefficient of Fold Change levels and bootstrap values were obtained with pvclust. Branches with bootstrap support < 70% are shown in grey. Central lines indicate shared clusters and cophenetic correlation between the dendrograms is shown in the bottom. The underlying data for this figure can be found at (dos Santos *et al.*, 2020d).

Effectors converge on deregulating the same metabolic pathways while others display unique patterns

Even though the transcripts affected by related effectors are different, in theory they could fall within the same metabolic pathway and therefore similarly alter the plant. To test this hypothesis, we searched for KEGG pathways over-represented in the up- and down-regulated genes in each transgenic line. “Biosynthesis of secondary metabolites” and “Metabolic pathways” were enriched among gene sets (either up-, red, or down-regulated, blue) of eight transgenic lines, while “MAPK signaling pathway” and “Plant-pathogen interaction” were enriched only among the down-regulated genes of six and five transgenic lines, respectively (**Figure 4**). We also found that “Starch and sucrose metabolism” was down-regulated in the transgenic lines Mlp123227 and Mlp124266, but up-regulated in the lines Mlp123218 and Mlp124497, whereas several transgenic lines showed impact on specialized metabolism. This was also visible in the enriched GO terms found on the WGCNA gene sets (**Table 2** and (dos Santos *et al.*, 2020d)). **File S1** shows heatmaps of 11 different metabolic pathways in which there were at least 10 genes deregulated across the experiment. The circadian rhythm pathway, although enriched only among the down-regulated genes of the lines Mlp124499, Mlp37347 and Mlp123531 and up-regulated genes in the Mlp124357 transgenic line, has several genes deregulated in all the transgenic lines studied. The plant-hormone signal transduction pathway is enriched among down-regulated genes in the transgenic lines Mlp37347, Mlp123531, and Mlp124497, and which we found several down-regulated genes (17, 23, and 17 DEGs, respectively) related to auxin response. From these results we conclude that CEs with similar sequences not only deregulate different genes but also alter different pathways.

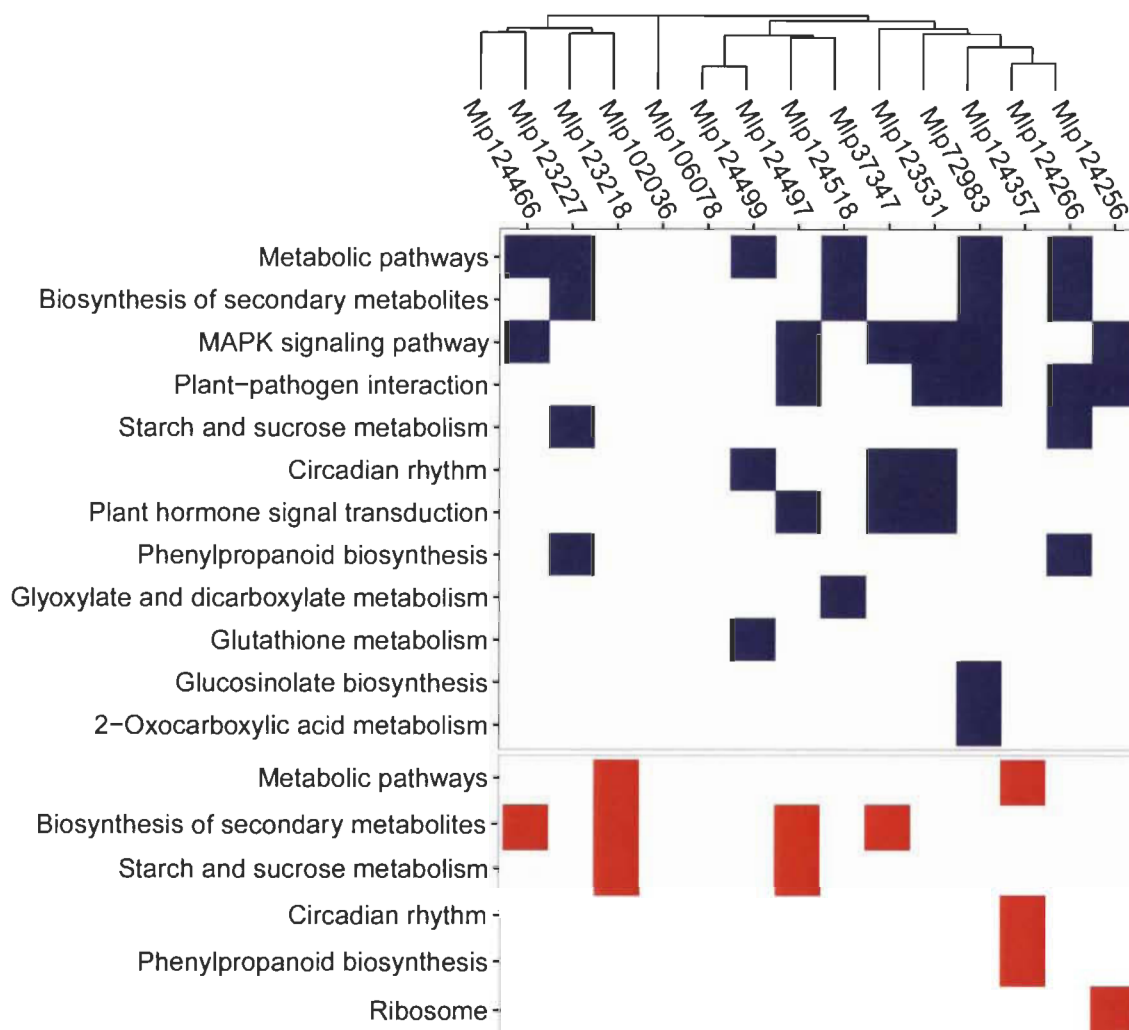


Figure 4. Effectors converge on deregulating the same metabolic pathways while others display unique patterns.

KEGG pathways over-represented among the sets of genes down- (blue) and up-regulated (red) in each transgenic line (columns) were calculated with KEGGprofile. Transgenic lines are ordered according to dendrogram of sequence similarity calculated with Muscle. The underlying data for this figure can be found at (dos Santos *et al.*, 2020d).

As both primary and specialized metabolisms were affected at the transcriptomic level and their levels can have an important role in the outcome of an infection, we proceeded with an untargeted analysis of the metabolome of these plants. We extracted metabolites with solutions containing 20% and 80% methanol and used ultra-high-resolution mass spectrometry in negative mode. A total of 5 192 masses were assigned across the experiment, ranging from 2 679 (Mlp123227) to 3 151 (Mlp124357) masses in each transgenic line (**Table S3**). When separated in biochemical categories, assigned formula belonged mostly to highly unsaturated and phenolic and aliphatic categories,

while peptides, sugars, condensed aromatics and polyphenolics were less important both in number of formulas and in relative abundance (**Figure 5**). Compared to the control, we found 680 assigned molecular formulas with a $|\log_2\text{-transformed Fold change}| > 2$ (**Figure 6A**), ranging from 69 metabolites in the line Mlp124466 (1.95% of the masses detected in this sample and/or in the control) to 353 in the line Mlp123227 (9.68% of the masses detected in this line and/or in the control, **Table S3**). In all transgenic lines, with exception of Mlp72983 and Mlp124256, there was over-representation of highly unsaturated and phenolic compounds among the down-regulated metabolites (accumulation level lower than in the control line) whereas up-regulated metabolites (accumulation level higher than in the control line) were enriched in peptides in all samples, except Mlp72983, Mlp106078 and Mlp124466 (**Figure 6B, Table S4**). As done with the transcriptomic data, we assessed whether the variation in the number of metabolites deregulated in each transgenic line could be explained by the level of expression of the transgene. For this, we plotted the number of deregulated metabolites per transgenic line (left Y-axis, blue) against the average expression level of the CEs in each transgenic line (X-axis, **Figure S2**). As the number of metabolites detected in each transgenic line varied (**Figure 5A**), we also plotted the ratio of deregulated metabolites:identified (detected either in the control or in the corresponding sample) metabolites in the right Y-axis (red). We found that the variation in transgene expression could explain neither the number ($R^2=0.0063$, $p\text{-Value}=0.7872$) nor the ratio of deregulated metabolites ($R^2=0.0033$, $p\text{-Value}=0.8444$), suggesting that the magnitude of the impact on the metabolome depends on the identity of the CE expressed in the plant rather than the strength of the CE expression.

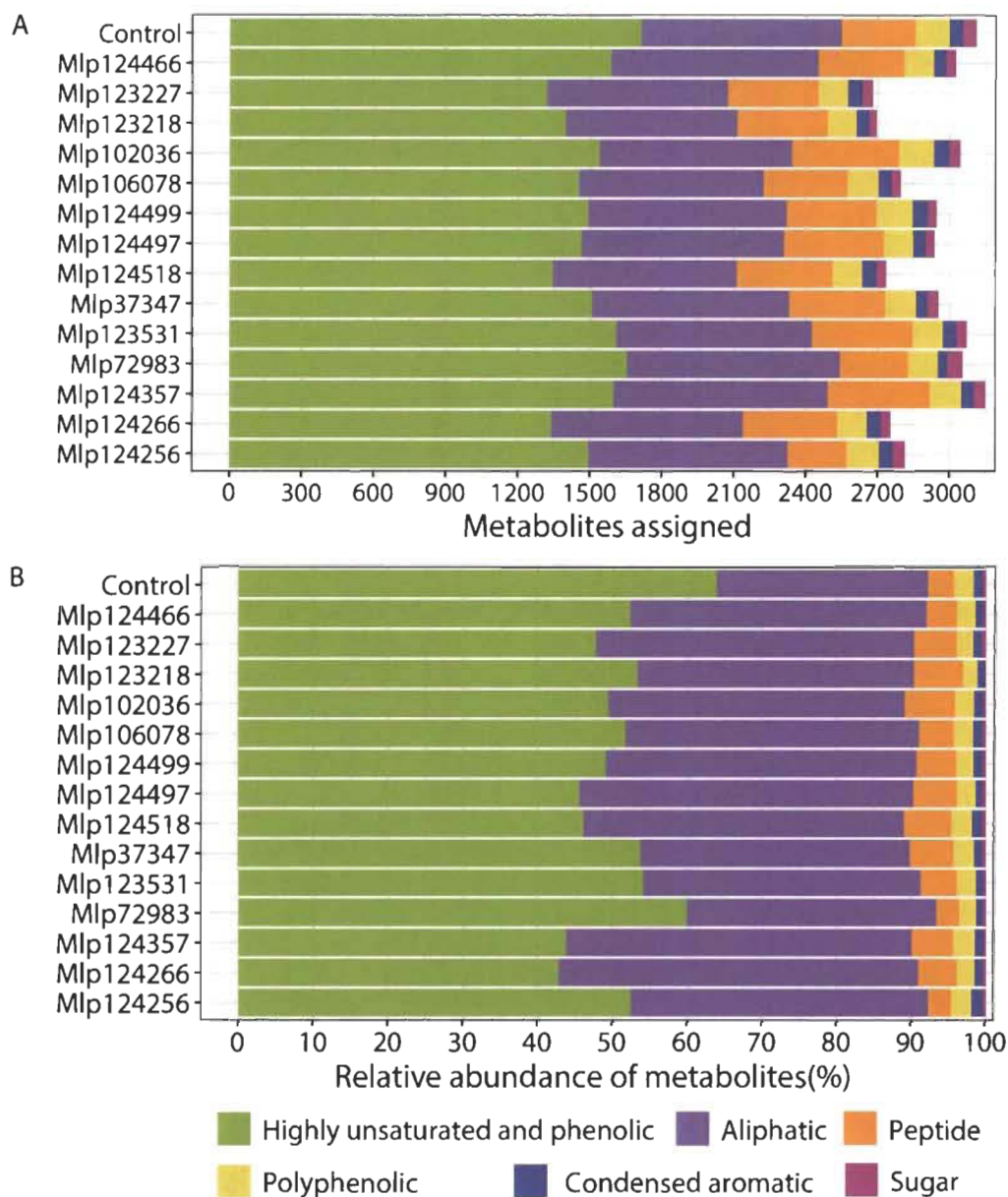


Figure 5. Metabolic composition of samples in number of formulas (A) and relative abundance of compounds (B).

Samples were analyzed in negative mode and estimated molecular formulas were separated in six categories: highly unsaturated and phenolic (green), aliphatic (purple), peptide (orange), polyphenolic (yellow), condensed aromatic (blue), and sugar (pink). The underlying data for this figure can be found at (dos Santos *et al.*, 2020d).

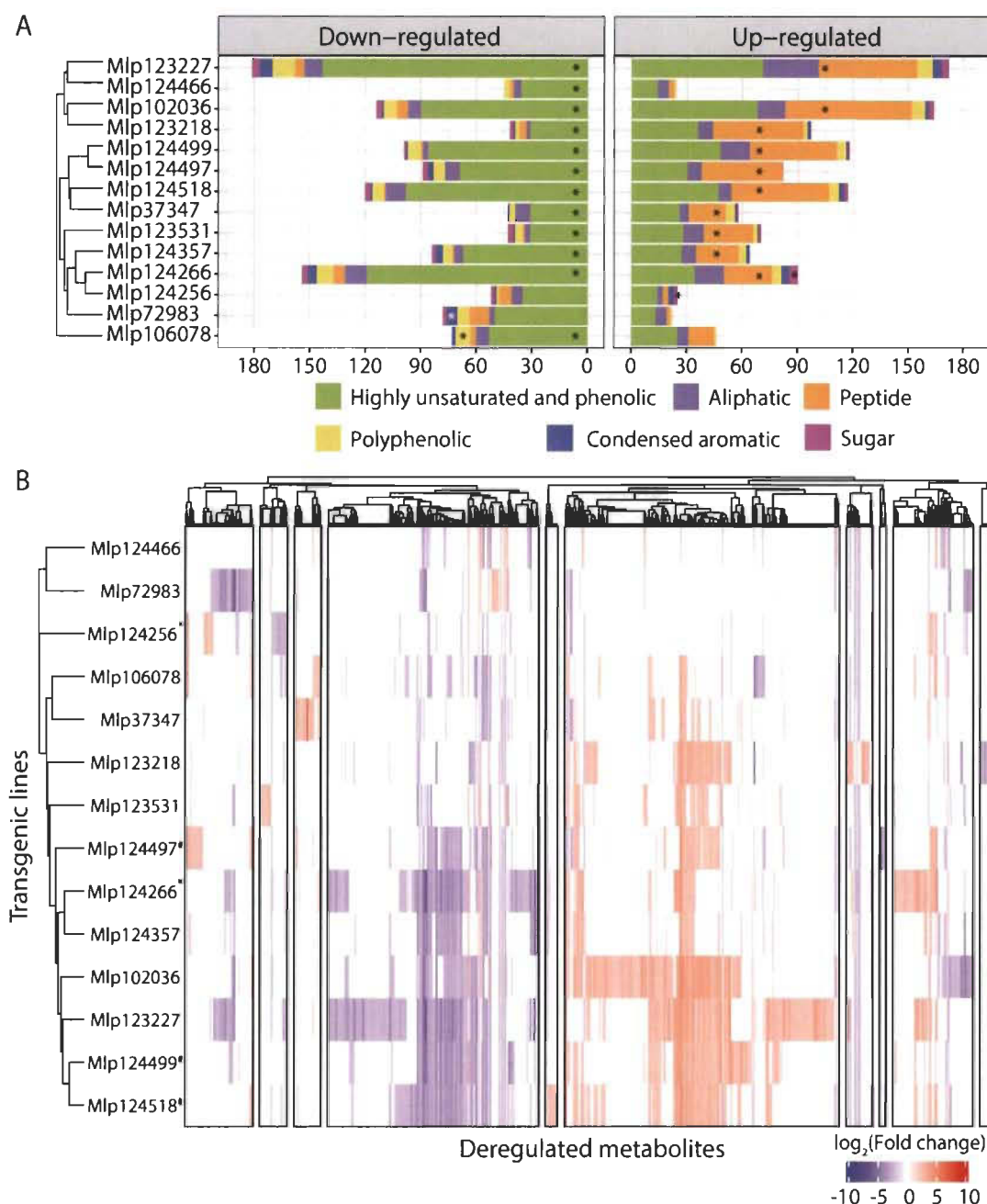


Figure 6. (A) Metabolites down-regulated (left) are enriched in highly unsaturated and phenolic compounds while peptides are over-represented among those up-regulated (right). Samples were analyzed in negative mode and relative abundance of metabolites in samples was compared to that in the control plants. Estimated molecular formulas were separated in six categories: highly unsaturated and phenolic (green), aliphatic (purple), peptide (orange), polyphenolic (yellow), condensed aromatic (blue), and sugar (pink). (B) Transgenic lines expressing candidate effectors with no similarity in amino acid sequence have correlated patterns of metabolite deregulation. Both metabolites and transgenic lines were clustered using Pearson's correlation. * indicates transgenic lines with CEs from the CPG5464 family; # indicates transgenic lines with CEs from the CPGH1 family. The underlying data for this figure can be found at (dos Santos *et al.*, 2020d).

In order to find shared patterns of metabolite deregulation across the transgenic lines studied, we used Pearson's correlation to group metabolites with correlated deregulation across the experiment and transgenic lines which deregulated the same metabolites. As observed with the gene deregulation, we found that transgenic lines expressing CEs without sequence similarity have correlated patterns of metabolite deregulation (**Figure 6B**). In the case of the CPGH1 family (CEs Mlp12497, Mlp124499, Mlp124518), lines Mlp124499 and Mlp124518 are correlated at 0.77 (Pearson's correlation), but their correlation with the line Mlp12497 is less strong (Mlp12497-Mlp124499: 0.59; Mlp124497-Mlp124518: 0.64). The two AvrP4 homologues, Mlp124256 and Mlp124266, have 46.3% of amino acid sequence similarity (dos Santos *et al.*, 2020d), but the correlation in metabolites deregulation patterns of the transgenic lines expressing these CEs is of 0.32. On the other hand, although Mlp124266 and Mlp124357 have 21.2% of amino acid sequence similarity (**Table S2**), multiple sequence alignment groups the AvrP4 homologues with the CE Mlp124357 (**Figure 7**) and their metabolite deregulation correlation is 0.69.

Remarkably, there was no correlation between the dendrograms gene and metabolite deregulation (cophenetic correlation of 0.1046 (**Figure 7**). When considering the number of genes and metabolites deregulated in each sample, the correlation was also low (Pearson's correlation = -0.1182). These results suggest these two omics approaches are needed to understand the magnitude of the impact of the CEs in the plant. Nevertheless, the possibility that the metabolic pathways deregulated at the metabolite level are the same as those deregulated in the gene level cannot be discarded.

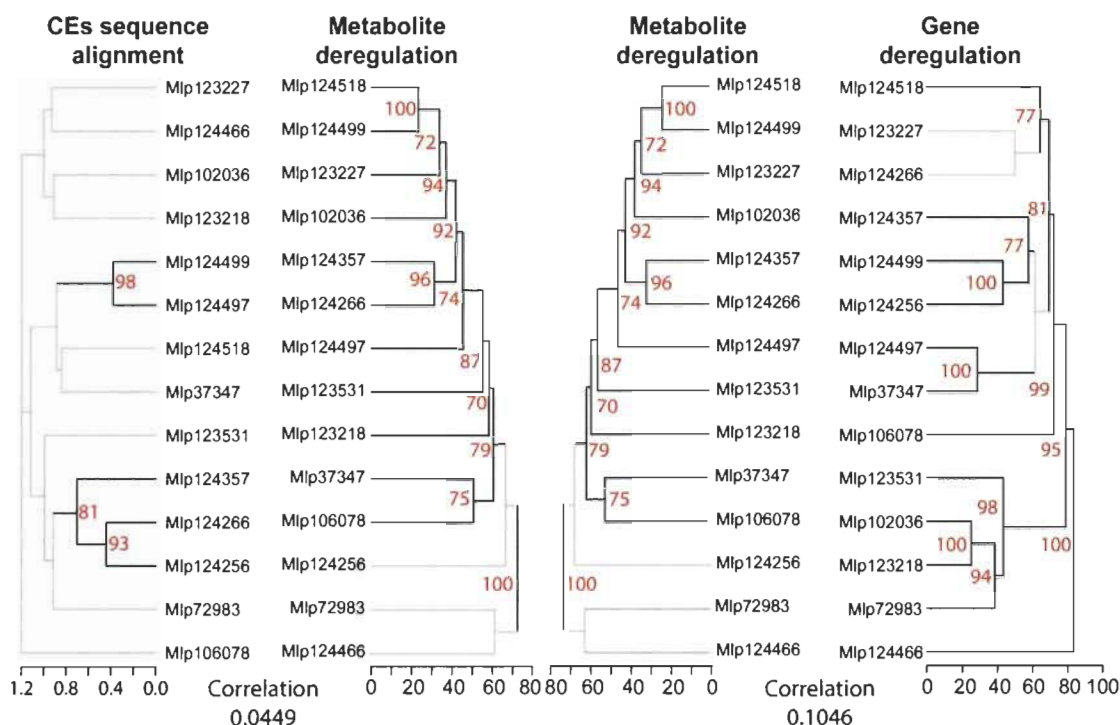


Figure 7. Hierarchical clustering based on metabolite deregulation groups effectors independently of amino acid sequence homology and gene deregulation patterns are not correlated to metabolite deregulation patterns in CE-expressing lines.

CE sequence alignment was computed with Muscle alignment and tree (left) was calculated with UPGMA. Dendrograms based on correlation of metabolite deregulation (center) or gene deregulation (right) were calculated with Pearson's correlation coefficient of Fold Change levels and bootstrap values were obtained with pvclust. Branches with bootstrap support < 70% are shown in grey. The underlying data for this figure can be found at [do \(dos Santos *et al.*, 2020d\)](https://doi.org/10.26434/chemrxiv-2020-00000).

To identify the molecular formula assigned in each sample and associate the metabolomic results with metabolic pathways, we searched for compounds with matching formula or matching m/z values in the KEGG database. From the 5 192 m/z detected across the experiment, 385 (7.42%) had a single match in KEGG database, while other 600 corresponded to multiple metabolites. When only considering the 680 deregulated metabolites, 54 (7.07%) matched a single metabolite and 82 (12.06%) matched multiple metabolites (dos Santos *et al.*, 2020d), leaving 546 unmatched.

Discussion

Effector biologists have tackled both the identification and the functional characterization of candidate effectors (CEs) (Selin *et al.*, 2016; Sperschneider *et al.*,

2015), as this is a key step towards a better understanding of plant-microbe interactions. In rust fungi, different approaches are used in the functional characterization of these proteins, including analysis of subcellular localization *in planta* (Gaouar *et al.*, 2016; Germain *et al.*, 2018; Madina *et al.*, 2020; Petre *et al.*, 2016; Petre *et al.*, 2015), infection assays in true host or in a model plant, and induction/repression of plant cell death (Germain *et al.*, 2018; Madina *et al.*, 2020; Ramachandran *et al.*, 2016; Schmidt, 2009). The transcriptome or metabolome of the host in responses to the pathogen are frequently evaluated (Azaiez *et al.*, 2009; Chandra *et al.*, 2016; Tao *et al.*, 2020; Tremblay *et al.*, 2010; Trujillo-Moya *et al.*, 2020; Zhang *et al.*, 2014), but the assessment of the role of individual CEs in these processes is not easily measured and seldom analyzed (Ahmed *et al.*, 2018; Meng *et al.*, 2018). Here we investigated 14 CEs from *Melampsora larici-populina* by evaluating their individual impact on the transcriptome and metabolome of stable transgenic *Arabidopsis* plants. By studying the impact of several individual CEs, we were able to compare patterns of gene and metabolite deregulation. Unexpectedly, we found that transgenic lines expressing CEs belonging to the same family did not have comparable patterns of gene or metabolite deregulation.

Previous studies in *M. larici-populina* have shown that genes encoding fungal effectors are expressed in waves in the telial host (Duplessis *et al.*, 2011b) and that members of the same family may be expressed during the infection of different hosts (Lorrain *et al.*, 2018a). This reflects the functional diversification of effectors, indicating that the fungus uses different sets of effectors for each stage of the infection and suggesting that effector families can have different functions, may target different host proteins or the same host protein that diverged in different hosts. The concurrent study of individual *M. larici-populina* CEs allows the comparison of their individual impact in the plant (Germain *et al.*, 2018). We found variability in the magnitude of the impact of each CE on the transcriptome (from 84 to 898 DEGs) and the metabolome (from 69 to 363 metabolites deregulated, **Figure 1** and **Figure 6**) of the transgenic plants, a variability which is not related to the level of expression of the transgenes (**Figure S1** and **Figure S2**). This suggests that the identities of the CEs are orienting the deregulations. By comparing the correlation of gene and metabolite deregulation patterns with the CEs sequence

similarity (**Figure 3** and **Figure 7**), we show that CEs belonging to the same family do not deregulate the transcriptome or the metabolome in a same way nor do they deregulate the same metabolic pathways (**Figure 4**). These results corroborate the infection assays from Germain and colleagues (Germain *et al.*, 2018). In their study, *Arabidopsis* plants, constitutively expressing *Mlp* CEs, were infected with *P. syringae* DC3000 or *H. arabidopsidis* Noco2. Mlp124497, Mlp124499 and Mlp124518 (family CPGH1) and Mlp124256 and Mlp124266 (family CPG5464) (Hacquard *et al.*, 2012), all increased *Arabidopsis* susceptibility to *H. arabidopsidis*. However, only Mlp124266, Mlp124497 and Mlp124499 made *Arabidopsis* more susceptible to *P. syringae*.

It has been suggested that proteins with higher sequence similarity have higher probability of having the same function (Joshi & Xu, 2007), thus small secreted proteins from many fungal and oomycete plant pathogens (Anderson *et al.*, 2017; Duplessis *et al.*, 2011a; Evangelisti *et al.*, 2017; Haas *et al.*, 2009; Saunders *et al.*, 2012) have been grouped in protein families to guide functional annotation and to help understand effector evolution. Nevertheless, recent studies have hypothesized that effectors from the same family may have different functions in the same host. This is the case for HopAF1 effectors from *P. savastanoi* (Castañeda-Ojeda *et al.*, 2017) and GALA effectors from *Ralstonia solanacearum* (Remigi *et al.*, 2011), which impact differently the plant defense. It is also the case for XopD effectors from plant pathogenic bacteria, which show different levels of SUMO protease activity and have different impacts in *Nicotiana* leaves (Kim *et al.*, 2011). This hypothesis is also supported by the evolution of the Tin2 effector in Ustilaginaceae. Tin2 from *Ustilago maydis* interacts with *Zea mays* TTK1 protein to stabilize it, leading to accumulation of anthocyanin. However, Tin2 from *Sporosorium reiliannum* interacts with *Zea mays* TTK2 and TTK3, inhibiting their activity (Tanaka *et al.*, 2019).

The CEs studied here deregulate diverse biochemical pathways in the plant (**Figure 4**). In relation to primary metabolism, genes in the “starch and sucrose metabolism” pathway were over-represented among up-regulated genes in the transgenic lines expressing the CEs Mlp123218 and Mlp124497, comparable to what is observed in

susceptible wheat infected with *Puccinia triticina* (Chandra *et al.*, 2016). On the other hand, the plants expressing Mlp123227 and Mlp124266 showed an enrichment of this pathway among down-regulated genes and the transgenic lines Mlp72983 and Mlp124266 had several genes down-regulated in this pathway as well (**File S1**), a pattern seen in resistant wheat infected with *P. triticina* (Chandra *et al.*, 2016). This difference in the direction of gene deregulation within the same pathway by different CEs may be an indication that deregulated genes have different functions. It can also suggest that these CEs are used in different stages of the infection. When considering pathways related to defense, the transcriptomic deregulations found in this study differ from previous reports of susceptible plants infected by rust fungi. While genes encoding Glutathione-S-transferase are down-regulated in at least one of 12 transgenic lines studied here (**File S1**), these genes are up-regulated in apple leaves infected with *Gymnosporangium yamadae* (Tao *et al.*, 2020). Moreover, Tremblay and colleagues (Tremblay *et al.*, 2010) reported up-regulation of genes in the “photosystem” and “nitrogen metabolism” pathways in susceptible *Glycine max* infected with *P. pachyrhizi*, whereas genes from these pathways were down-regulated in our transgenic lines.

There are several possible explanations for the differences between previous studies and our own. First, our results may be due to the long-term exposure of our plants to CEs, as they are stable transgenic lines, whereas during the infection rust fungi secrete effectors in waves (Duplessis *et al.*, 2011b), these proteins are not constitutively present in the host. It is also possible that results from Tao and colleagues (Tao *et al.*, 2020) and Tremblay and colleagues (Tremblay *et al.*, 2010) included the activation of PTI as well as the combinatory effect of multiple effectors, as they investigated plant response to the fungal infection, not to individual CEs. Our approach was to express CEs from *M. larici-populina* in a plant that cannot be infected by this fungus, thus should not recognize these proteins nor mount active defense responses against them (ETI). Nevertheless, as we evaluated the impact of each CE in the plant using one single transgenic line, it is not possible to know for sure if the impact on the transcriptome and metabolome is caused by the CE or is a secondary effect of the DNA insertion site in each of these transgenic lines. Yet, the probability that the insertion site impacted in the same manner the results of

all the 14 transgenic lines studied here is low, thus the results that consider the 14 transgenic lines are robust. Finally, although there are limitations in the use of heterologous systems, they allow faster functional characterization of CEs (Lorrain *et al.*, 2018b; Popa *et al.*, 2016) and they may be indispensable for high-throughput studies of CEs of obligate biotrophic pathogens or other microorganisms not amenable to genetic manipulation (Cunha *et al.*, 2014; Rice *et al.*, 2017).

Taken together, our results reinforces the hypothesis that the CEs studied here and functionally characterized by Germain and colleagues (Germain *et al.*, 2018) are *bona fide* effectors. Nevertheless, future studies interested in CEs evaluated here should analyze more independent transgenic lines. In addition, since our methodology for the metabolomic analysis is semi-quantitative and does not allow the distinction of metabolites with the same *m/z*, follow up studies should use chromatography in tandem with mass spectrometry and should analyze more replicates for the mass spectrometry. Our study also questions the validity of grouping CEs by sequence similarity. The importance of this approach for understanding the evolution of effectors is obvious (Duplessis *et al.*, 2011a), but basing functional characterization on sequence similarity may be misleading (Castañeda-Ojeda *et al.*, 2017; Remigi *et al.*, 2011; Tanaka *et al.*, 2019).

Materials and Methods

Plant growth conditions

Arabidopsis thaliana transgenic plants in Columbia-0 background expressing GFP alone (control) or fused to a candidate effector of the fungus *Melampsora larici-populina* (Mlp37347, Mlp72983, Mlp102036, Mlp106078, Mlp123218, Mlp123227, Mlp123531, Mlp124256, Mlp124266, Mlp124357, Mlp124466, Mlp124497, Mlp124499, Mlp124518) previously obtained in our laboratory (Germain *et al.*, 2018; Madina *et al.*, 2020), were grown at 22°C at 16h/8h light/dark cycles.

RNA extraction and transcriptome analysis

RNA was extracted from pooled aerial tissue of 2-week-old soil-grown plants, using three replicates per genotype, with the Plant Total RNA Mini Kit (Geneaid) using RB buffer following manufacturer's protocol. The samples were treated with DNase, then RNA quality was assessed using agarose gel electrophoresis. QC was performed using a 2100 Bioanalyzer (Agilent) and only samples having an RNA Integrity Number higher than 7 were kept for library preparation. Libraries were generated with the NeoPrep Library Prep System (Illumina) using the TruSeq Stranded mRNA Library Prep kit (Illumina) and 100 ng of total RNA as per the manufacturer's recommendations. The libraries were then sequenced with Illumina HiSeq 4000 Sequencer with paired-end reads of 100 nt at the Genome Quebec Innovation Centre (McGill University, Montreal, Canada).

The bioinformatic analyses were done with Compute Canada servers, the parameters used are presented in **Table S5**. We trimmed the reads using Trimmomatic (Bolger *et al.*, 2014) and we aligned the surviving paired reads to the genome of *A. thaliana* assembly TAIR10 with HISAT2 (Kim *et al.*, 2019). Unmapped reads were aligned to the sequences of the CEs, without signal peptide, attached to eGFP. We counted the reads assigned to each transcript with the R (v3.6) packages Rsamtools (v2.2.3 (Morgan *et al.*, 2020)), GenomicAlignments and GenomicFeatures (Lawrence *et al.*, 2013). The general information of the sequencing results and mapping data is presented in **Table S6**. Before comparing the samples, we used the CustomSelection package (dos Santos *et al.*, 2020a) to select as reference genes the top 5% genes with lowest coefficient of variation of TPM among the 45 samples (dos Santos *et al.*, 2020d). We assessed the variation between the replicates and the similarity of the samples with principal component analysis (**Figure S3**). Differential expression analysis was performed with DeSeq2 (Love *et al.*, 2014), using the un-normalized counts as input, and genes with $|\log_2 \text{Fold change}| \geq 2$ (p-Value ≤ 0.01), when comparing each CE-expressing lines to the control line, were considered as deregulated. We used clusterProfiler (Yu *et al.*, 2012) for GO term enrichment analysis and KEGGprofile (v1.24.0 (Zhao *et al.*, 2018)) for KEGG enrichment analysis. Sets of deregulated genes were computed using WGCNA

(Langfelder & Horvath, 2008). We calculated the similarity of gene deregulation of different transgenic lines with the R package pvclust v2.2-0 (Suzuki *et al.*, 2019), using Pearson's correlation and 5 000 bootstrap replications.

Metabolite extraction and metabolomics analysis

Metabolites were extracted from pooled aerial tissue of 2-week-old soil-grown plants, with four replicates per genotype. After pulverizing the tissues with a TissueLyser (30 cycles per second for 45 seconds repeated 3 times), we added 300 μ L of distilled water to it. From the mix of tissue and water, we used 100 μ L of tissue slurry for an extraction with 1 mL of 20% methanol and a separate 100 μ L for an extraction with 1 mL of 80% methanol. After agitation with the solvent, we pooled the samples of the same genotype and extraction together and filtered them using glass microfiber filters (Whatman GF/F CAT No. 1825-025). We evaporated the extracts with a speed vacuum at room temperature and chamber vacuum of 7.4 torr's and resolubilized them in 2 mL of distilled water. Then, we solid phase extracted 50 μ g of dissolved organic carbon (DOC) of each sample, using Agilent PPL cartridges, and eluted it in 1 mL of 100% methanol.

The mass spectrometry was performed in an Orbitrap LTQ-Velos calibrated and tuned to maximize the peak at 369.1 in Suwannee River Fulvic Acid (SRFA) reference material. The extracts were analysed by direct injection in negative mode at a resolution setting 100 000, with accumulation time set to a maximum of 500 ms and a target of 1×10^6 ions. Peaks were only considered for formula assignment if their intensity was higher than 10x the median noise baseline. We assigned formulas to masses using an in-house MATLAB script (Hawkes *et al.*, 2018) and we allowed assignments with mass error < 2 ppm. Briefly, formulas were considered over the ranges $C_{4-50}H_{4-100}O_{2-40}N_{0-2}$ under the conditions $O \leq C$; $0.3C \leq H \leq 2.2C$. For each sample, the intensity of the peaks was normalised so that the sum of the intensities equalled 10 000. Following analyses were performed using R software (v4.0). We used the molecular formulas to calculate the modified aromaticity index (AImod) of each metabolite (Koch & Dittmar, 2006) and the compound categories were defined as: condensed aromatic (AImod > 0.66), polyphenolic

($0.66 \geq \text{AI}_{\text{mod}} > 0.5$), highly unsaturated and phenolic ($\text{AI}_{\text{mod}} < 0.5$ and $\text{H/C} < 1.5$), aliphatic ($2 \geq \text{H/C} \geq 1.5$, $\text{N} = 0$), peptide ($2 \geq \text{H/C} \geq 1.5$, $\text{N} > 0$) or sugar ($\text{O/C} > 0.9$) (Kellerman *et al.*, 2018).

The results of the two extractions, with 20% and 80% methanol, were combined and the fold changes (FC) were calculated as $\log_2\left(\frac{0.5+M_x^y}{0.5+M_c^y}\right)$, where M_x^y is the relative abundance of the metabolite y in the CE-sample x and M_c^y is the relative abundance of the metabolite y in the control. For each sample, only metabolites with $|\text{FC}| > 2$ were considered to have relative abundance different to that of the control. Categories enriched among up- and down-regulated genes were found by applying Fisher's test. We calculated the similarity of metabolite deregulation of different transgenic lines with the R package *pvcust* v2.2-0 (Suzuki *et al.*, 2019), using Pearson's correlation and 5 000 bootstrap replications. Pairwise correlation of metabolite deregulation between specific transgenic lines was calculated with the function *cor* from the R package *stats*, using the method "pearson". We were not able to analyze the extraction with 80% methanol of the transgenic line Mlp123218, thus the results presented for this line are only of the extraction with 20% methanol and they are compared to the results of the Control for the same extraction for consistency.

We searched the molecular formulas, obtained with the in-house script, in KEGG database using the R package KEGGREST 1.24.0 (Tenenbaum, 2020) for identification of the metabolites detected. We also used Pathos (Leader *et al.*, 2011) to search for metabolites with the same m/z (settings: negative mode, all organisms, $-\text{H}^+$ as adduct and mass error at 3 ppm).

Sequence analysis and integration

Multiple sequence alignment of CE amino acid sequences without signal peptides was performed with the software MEGA X (Kumar *et al.*, 2018) using Muscle (Madeira *et al.*, 2019) default settings. Evolutionary history was inferred using UPGMA method and 1 000 bootstrap replicates. Comparisons of dendrograms from CE sequence

alignment, gene and metabolite deregulation correlation were done with the R package dendextend (Galili, 2015) by calculating the cophenetic correlation between two dendrograms. We performed pairwise sequence alignment of the 14 CEs using Needle (Madeira *et al.*, 2019), with default parameters.

Data availability

Transcriptomics: Raw reads and count matrices are available in NCBI GEO under the accession GSE158410 (dos Santos *et al.*, 2020b).

Metabolomics: Raw and mzXML files along with annotation of metabolites and their relative abundances in each sample are available at MetaboLights under the accession MTBLS2096 (dos Santos *et al.*, 2020c).

Data underlying figures, full list of enriched GO terms in the WGNCA gene sets, information on the deregulated metabolites and the list of selected reference genes is available at (dos Santos *et al.*, 2020d).

Acknowledgements

We thank Melodie B. Plourde and Benjamin Petre for critical review of the manuscript. Funding for the project was provided by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants to HG. The project in HG's laboratory was also partially funded by an institutional Research Chair and a Canada Research Chair held by HG and a Canada Research Chair held by IDP. KCGS was funded by a master's scholarship from the Fondation de l'Université du Québec à Trois-Rivières, an international PhD scholarship from the Fonds de Recherche du Québec sur la Nature et les Technologies (FRQNT) and a graduate fellowship from MITACS.

References

- Ahmed, M. B., Santos, K. C. G. d., Petre, B., Lorrain, C., Duplessis, S., Desgagne-Penix, I., & Germain, H. (2018). A rust fungal effector binds plant DNA and modulates transcription. *Nature Scientific Reports*, 8, 14718. doi: 10.1038/s41598-018-32825-0
- Aime, M. C., McTaggart, A. R., Mondo, S. J., & Duplessis, S. (2017). Phylogenetics and phylogenomics of rust fungi. *Advances in Genetics*, doi: 10.1016/bs.adgen.2017.09.011
- Anderson, J. P., Sperschneider, J., Win, J., Kidd, B., Yoshida, K., Hane, J., . . . Singh, K. B. (2017). Comparative secretome analysis of *Rhizoctonia solani* isolates with different host ranges reveals unique secretomes and cell death inducing effectors. *Nature Scientific Reports*, 7, 10410. doi: 10.1038/s41598-017-10405-y
- Azaiez, A., Boyle, B., Levée, V., & Séguin, A. (2009). Transcriptome profiling in hybrid poplar following interactions with *Melampsora* rust fungi. *Molecular Plant-Microbe Interactions*, 22(2), 190-200. doi: 10.1094/MPMI-22-2-0190
- Bentem, S. D. L. F. v., Vossen, J. H., Vries, K. J. d., Wees, S. v., Tameling, W. I. L., Dekker, H. L., . . . Cornelissen, B. J. C. (2005). Heat shock protein 90 and its co-chaperone protein phosphatase 5 interact with distinct regions of the tomato I-2 disease resistance protein. *The Plant Journal*, 43(2), 284-298. doi: 10.1111/j.1365-313X.2005.02450.x
- Bigeard, J., Colcombet, J., & Hirt, H. (2015). Signaling mechanisms in pattern-triggered immunity (PTI). *Molecular Plant*, 8(4), 521-539. doi: 10.1016/j.molp.2014.12.022
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. doi: 10.1093/bioinformatics/btu170
- Caillaud, M. C., Piquerez, S. J. M., Fabro, G., Steinbrenner, J., Ishaque, N., Beynon, J., & Jones, J. D. G. (2012). Subcellular localization of the *Hpa* RxLR effector repertoire identifies a tonoplast-associated protein HaRxL17 that confers enhanced plant susceptibility. *The Plant Journal*, 69(2), 252-265. doi: 10.1111/j.1365-313X.2011.04787.x
- Castañeda-Ojeda, M. P., López-Solanilla, E., & Ramos, C. (2017). Differential modulation of plant immune responses by diverse members of the *Pseudomonas savastanoi* pv. *savastanoi* HopAF type III effector family. *Molecular Plant Pathology*, 18(5), 625-634. doi: 10.1111/mpp.12420

- Chandra, S., Singh, D., Pathak, J., Kumari, S., Kumar, M., Poddar, R., . . . Mukhopadhyay, K. (2016). *De novo* assembled wheat transcriptomes delineate differentially expressed host genes in response to leaf rust infection. *PLoS One*, 11(2), e0148453. doi: 10.1371/journal.pone.0148453
- Chaudhari, P., Ahmed, B., Joly, D. L., & Germain, H. (2014). Effector biology during biotrophic invasion of plant cells. *Virulence*, 7(7), 703-709. doi: 10.4161/viru.29652
- Cunha, M. d., Milho, C., Almeida, F., Pais, S. V., Borges, V., Maurício, R., . . . Mota, L. J. (2014). Identification of type III secretion substrates of *Chlamydia trachomatis* using *Yersinia enterocolitica* as a heterologous system. *BMC Microbiology*, 14, 40. doi: 10.1186/1471-2180-14-40
- Dangl, J. L., Horvath, D. M., & Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science*, 341(6147), 6746-6751. doi: 10.1126/science.1236011
- Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., . . . Foster, G. D. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*, 13(4), 414-430. doi: 10.1111/j.1364-3703.2011.00783.x
- Dodds, P. N., Lawrence, G. J., Catanzariti, A.-M., Ayliffe, M. A., & Ellis, J. G. (2004). The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell*, 16(3), 755-768. doi: 10.1105/tpc.020040
- dos Santos, K. C. G., Desgagné-Pénix, I., & Germain, H. (2020a). Custom selected reference genes outperform pre-defined reference genes in transcriptomic analysis. *BMC Genomics*, 21, 35. doi: 10.1186/s12864-019-6426-2
- dos Santos, K. C. G., Pelletier, G., Séguin, A., Guillemette, F., Hawkes, J., Desgagné-Pénix, I., & Germain, H. (2020b) Differential alteration of plant functions by homologous fungal candidate effectors. Repéré le 2020/09/23 de <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158410>
- dos Santos, K. C. G., Pelletier, G., Séguin, A., Guillemette, F., Hawkes, J. A., Desgagné-Pénix, I., & Germain, H. (2020c) MTBLS2096: Impact of *Melampsora larici-populina*'s candidate effectors in *Arabidopsis* is unrelated to sequence similarity. Repéré le 2020/12/03 de <https://www.ebi.ac.uk/metabolights/MTBLS2096/descriptors>

- dos Santos, K. C. G., Pelletier, G., Séguin, A., Guillemette, F., Hawkes, J. A., Desgagné-Penix, I., & Germain, H. (2020d) Supplementary material: Differential alteration of plant functions by homologous fungal candidate effectors. Repéré le 2020/10/29 de <https://doi.org/10.6084/m9.figshare.13166501.v3>
- Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., . . . Martin, F. (2011a). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *PNAS*, *108*, 9166-9171. doi: 10.1073/pnas.1019315108
- Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., Martin, F., & Kohler, A. (2011b). *Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Molecular Plant-Microbe Interactions*, *24*(7), 808-818. doi: 10.1094/MPMI-01-11-0006
- Enright, A. J., Dongen, S. V., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, *30*(7), 1575-1584. doi: 10.1093/nar/30.7.1575
- Evangelisti, E., Gogleva, A., Hainaux, T., Doumane, M., Tulin, F., Quan, C., . . . Schormack, S. (2017). Time-resolved dual transcriptomics reveal early induced *Nicotiana benthamiana* root genes and conserved infection-promoting *Phytophthora palmivora* effectors. *BMC Biology*, *15*, 39. doi: 10.1186/s12915-017-0379-1
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*, *31*(22), 3718-3720. doi: 10.1093/bioinformatics/btv428
- Gan, P. H. P., Rafiqi, M., Ellis, J. G., Jones, D. A., Hardham, A. R., & Dodds, P. N. (2010). Lipid binding activities of flax rust AvrM and AvrL567 effectors. *Plant Signaling & Behavior*, *5*(10), 1272-1275. doi: 10.4161/psb.5.10.13013
- Gaouar, O., Morency, M.-J., Letanneur, C., Séguin, A., & Germain, H. (2016). The 124202 candidate effector of *Melampsora larici-populina* interacts with membranes in *Nicotiana* and *Arabidopsis*. *Canadian Journal of Plant Pathology*, doi: 10.1080/07060661.2016.1153523
- Germain, H., Joly, D. L., Mireault, C., Plourde, M. B., Letanneur, C., Stewart, D., . . . Séguin, A. (2018). Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Molecular Plant Pathology*, *19*(1), 191-200. doi: 10.1111/mp.12514

- Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., . . . Nusbaum, H. S. J. C. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature*, 461, 393-398. doi: 10.1038/nature08358
- Hacquard, S., Joly, D. L., Lin, Y. C., Tisserant, E., Feau, N., Delaruelle, C., . . . Duplessis, S. (2012). A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Molecular Plant-Microbe Interactions*, 25(3), 279-293. doi: 10.1094/MPMI-09-11-0238
- Hawkes, J. A., Patriarca, C., Sjöberg, P. J. R., Tranvik, L. J., & Bergquist, J. (2018). Extreme isomeric complexity of dissolved organic matter found across aquatic environments. *Limnology And Oceanography Letters*, 3(2018), 2021-2030. doi: 10.1002/lol2.10064
- Henry, G., Thonart, P., & Ongena, M. (2012). PAMPs, MAMPs, DAMPs and others: an update on the diversity of plant immunity elicitors. *Biotechnology, Agronomy, Society and Environment*, 16(2), 257-268. <https://www.pressesagro.be/base/index.php/base/article/view/611>
- Hogenhout, S. A., Hoorn, R. A. L. V. d., Terauchi, R., & Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. *Molecular Plant-Microbe Interactions*, 22(2), 115-122. doi: 10.1094/MPMI-22-2-0115
- Houterman, P. M., L., M., van Ooijen, G., De Vroomen, M. J., Cornelissen, B. J. C., Takken, F. L. W., & Rep, M. (2009). The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *The Plant Journal*, 58(6), 970-978. doi: 10.1111/j.1365-313X.2009.03838.x
- Jamir, Y., Guo, M., Oh, H. S., Petnicki-Ocwieja, T., Chen, S., Tang, X., . . . Alfano, J. R. (2004). Identification of *Pseudomonas syringae* type III effectors that can suppress programmed cell death in plants and yeast. *The Plant Journal*, 37(4), 554-565. doi: 10.1046/j.1365-313X.2003.01982.x
- Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 7323-7332. doi: 10.1038/nature05286
- Joshi, T., & Xu, D. (2007). Quantitative assessment of relationship between sequence similarity and function similarity. *BMC Genomics*, 8, 222. doi: 10.1186/1471-2164-8-222

- Kellerman, A. M., Guillemette, F., Podgorski, D. C., Aiken, G. R., Butler, K. D., & Spencer, R. G. M. (2018). Unifying concepts linking dissolved organic matter composition to persistence in aquatic ecosystems. *Environmental Science & Technology*, 52(5), 2538-2548. doi: 10.1021/acs.est.7b05513
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology*, 37, 907-915. doi: 10.1038/s41587-019-0201-4
- Kim, J.-G., Taylor, K. W., & Mudgett, M. B. (2011). Comparative analysis of the XopD type III secretion (T3S) effector family in plant pathogenic bacteria. *Molecular Plant Pathology*, 12(8), 715-730. doi: 10.1111/J.1364-3703.2011.00706.X
- Koch, B. P., & Dittmar, T. (2006). From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communication in Mass Spectrometry*, 20(5), 926-932. doi: 10.1002/rcm.2386
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35, 1547-1549. doi: 10.1093/molbev/msz312
- Kunjetti, S. G., Iyer, G., Johnson, E., Li, E., & Broglie, K. E. (2016). Identification of *Phakopsora pachyrhizi* candidate effectors with virulence activity in a distantly related pathosystem. *Frontiers in Plant Science*, 7, 269. doi: 10.3389/fpls.2016.00269
- Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559-571. doi: 10.1186/1471-2105-9-559
- Lawrence, G. J., Huber, M. L. W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., . . . Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, 9, e1003118. doi: 10.1371/journal.pcbi.1003118
- Leader, D. P., Burgess, K., Creek, D., & Barrett, M. P. (2011). Pathos: A web facility that uses metabolic maps to display experimental changes in metabolites identified by mass spectrometry. *Rapid Communication in Mass Spectrometry*, 25(2), 3422-3426. doi: 10.1002/rcm.5245
- Li, Z., Yin, Z., Fan, Y., Xu, M., Kang, Z., & Huang, L. (2015). Candidate effector proteins of the necrotrophic apple canker pathogen *Valsa mali* can suppress BAX-induced PCD. *Frontiers in Plant Science*, 6, 579. doi: 10.3389/fpls.2015.00579

- Lorrain, C., Hecker, A., & Duplessis, S. (2015). Effector-mining in the poplar rust fungus *Melampsora larici-populina*. *Frontiers in Plant Science*, 6, 1051. doi: 10.3389/fpls.2015.01051
- Lorrain, C., Marchal, C., Hacquard, S., Delaruelle, C., Péytowski, J., Petre, B., . . . Duplessis, S. (2018a). The rust fungus *Melampsora larici-populina* expresses a conserved genetic program and distinct sets of secreted protein genes during infection of its two host plants, larch and poplar. *Molecular Plant-Microbe Interactions*, 31(7), 695-706. doi: 10.1094/MPMI-12-17-0319-R
- Lorrain, C., Petre, B., & Duplessis, S. (2018b). Show me the way: rust effector targets in heterologous plant systems. *Current Opinion in Microbiology*, 46, 19-25. doi: 10.1016/j.mib.2018.01.016
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), 550. doi: 10.1186/s13059-014-0550-8
- Luna, E., Pastor, V., Robert, J., Flors, V., Mauch-Mani, B., & Ton, J. (2011). Callose deposition: A multifaceted plant defense response. *Molecular Plant-Microbe Interactions*, 24(2), 183-193. doi: 10.1094/MPMI-07-10-0149
- Lyu, X., Shen, C., Fu, Y., Xie, J., Jiang, D., Li, G., & Cheng, J. (2016). A small secreted virulence-related protein is essential for the necrotrophic interactions of *Sclerotinia sclerotiorum* with its host plants. *PLoS Pathogens*, 12(2), e1005435. doi: 10.1371/journal.ppat.1005435
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., . . . Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47(W1), W636-W641. doi: 10.1093/nar/gkz268
- Madina, M. H., Rahman, M. S., Huang, X., Zhang, Y., Zheng, H., & Germain, H. (2020). A poplar rust effector protein associates with protein disulfide isomerase and enhances plant susceptibility. *Biology*, 9(9), 294. doi: 10.3390/biology9090294
- Meng, Q., Gupta, R., Kwon, S. J., Wang, Y., Agrawal, G. K., Rakwal, R., . . . Kim, S. T. (2018). Transcriptomic analysis of *Oryza sativa* leaves reveals key changes in response to *Magnaporthe oryzae* MSP1. *The Plant Pathology Journal*, 34(4), 257-268. doi: 10.5423/PPJ.OA.01.2018.0008
- Petre, B., Saunders, D. G. O., Sklenar, J., Lorrain, C., Krasileva, K. V., Win, J., . . . Kamoun, S. (2016). Heterologous expression screens in *Nicotiana benthamiana* identify a candidate effector of the wheat yellow rust pathogen that associates with processing bodies. *PLoS One*, 11(2), e0149035. doi: 10.1371/journal.pone.0149035

- Petre, B., Saunders, D. G. O., Sklenar, J., Lorrain, C., Win, J., Duplessis, S., & Kamoun, S. (2015). Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. *Molecular Plant-Microbe Interactions*, 28(6), 689-700. doi: 10.1094/MPMI-01-15-0003-R
- Pitino, M., Armstrong, C. M., Cano, L. M., & Duan, Y. (2016). Transient expression of *Candidatus Liberibacter Asiaticus* effector induces cell death in *Nicotiana benthamiana*. *Frontiers in Plant Science*, 7, 982. doi: 10.3389/fpls.2016.00982
- Plett, J. M., Kemppainen, M., Kale, S. D., Kohler, A., Legué, V., Brun, A., . . . Martin, F. (2011). A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Current Biology*, 21(14), 1197-1203. doi: 10.1016/j.cub.2011.05.033
- Popa, C., Coll, N. S., Valls, M., & Sessa, G. (2016). Yeast as a heterologous model system to uncover type III effector function. *PLoS Pathogens*, 12 doi: 10.1371/journal.ppat.1005360
- Ramachandran, S. R., Yin, C., Kud, J., Tanaka, K., Mahoney, A. K., Xiao, F., & Hulbert, S. H. (2016). Effectors from wheat rust fungi suppress multiple plant defense responses. *Phytopathology*, 107(1), 75-83. doi: 10.1094/PHYTO-02-16-0083-R
- Ravensdale, M., Nemri, A., Thrall, P. H., Ellis, J. G., & Dodds, P. N. (2010). Co-evolutionary interactions between host resistance and pathogen effector genes in flax rust disease. *Molecular Plant Pathology*, 12, 93-102. doi: 10.1111/j.1364-3703.2010.00657.x
- Remigi, P., Anisimova, M., Guidot, A., Genin, S., & Peeters, N. (2011). Functional diversification of the GALA type III effector family contributes to *Ralstonia solanacearum* adaptation on different plant hosts. *New Phytologist*, 192(4), 976-987. doi: 10.1111/j.1469-8137.2011.03854.x
- Rice, D. W., Sheehan, K. B., & Newton, I. L. G. (2017). Large-scale identification of *Wolbachia pipientis* effectors. *Genome Biology and Evolution*, 9(7), 1925-1937. doi: 10.1093/gbe/evx139
- Saunders, D. G. O., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S., & Raffaele, S. (2012). Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS One*, 7(1), e29847. doi: 10.1371/journal.pone.0029847
- Schmidt, S. M. (2009). *Identification and functional characterization of powdery mildew effectors* (Thèse de Dissertation inédite). Universität zu Köln, Köln

- Selin, C., Kievit, T. R. d., Belmonte, M. F., & Fernando, W. G. D. (2016). Elucidating the role of effectors in plant-fungal interactions: Progress and challenges. *Frontiers in Microbiology*, doi: 10.3389/fmicb.2016.00600
- Sperschneider, J., Dodds, P. N., Gardiner, D. M., Manners, J. M., Singh, K. B., & Taylor, J. M. (2015). Advances and challenges in computational prediction of effectors from plant pathogenic fungi. *PLoS Pathogens*, doi: 10.1371/journal.ppat.1004806
- Tanaka, S., Schweizer, G., Rössel, N., Thines, M., & Kahmann, R. (2019). Neofunctionalization of the secreted Tin2 effector in the fungal pathogen *Ustilago maydis*. *Nature Microbiology*, 4, 251-257. doi: 10.1038/s41564-018-0304-6
- Tao, S., Auer, L., Morin, E., Liang, Y.-M., & Duplessis, S. (2020). Transcriptome analysis of apple leaves infected by the rust fungus *Gymnosporangium yamadae* at two sporulation stages reveals detoxication and secondary metabolite host responses and fungal pathogenesis related genes. *Molecular Plant-Microbe Interactions*, 33(3), 444-461. doi: 10.1094/MPMI-07-19-0208-R
- Tremblay, A., Hosseini, P., Alkharouf, N. W., Li, S., & Matthews, B. F. (2010). Transcriptome analysis of a compatible response by *Glycine max* to *Phakopsora pachyrhizi* infection. *Plant Science*, 179(3), 183-193. doi: 10.1016/j.plantsci.2010.04.011
- Trujillo-Moya, C., Ganthaler, A., Stöggel, W., Kranner, I., Schüler, S., Ertl, R., . . . Mayr, S. (2020). RNA-Seq and secondary metabolite analyses reveal a putative defence-transcriptome in Norway spruce (*Picea abies*) against needle bladder rust (*Chrysomyxa rhododendri*) infection. *BMC Genomics*, 54, 336. doi: 10.1186/s12864-020-6587-z
- Yu, G., Wang, L., Han, Y., & He, Q. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS : A journal of integrative biology*, 16(5), 284-287. doi: 10.1089/omi.2011.0118
- Zhang, H., Yang, Y., Wang, C., Liu, M., Li, H., Fu, Y., . . . Ji, W. (2014). Large-scale transcriptome comparison reveals distinct gene activations in wheat responding to stripe rust and powdery mildew. *BMC Genomics*, 15, 898. doi: 10.1186/1471-2164-15-898

Supplementary material

Supplementary Table S1 is not shown here, but available in xls format at: https://github.com/KarenGoncalves/Dos_Santos_2020b/blob/main/Dos_Santos_Chapter_3_Supplementary_table_3-1.xlsx

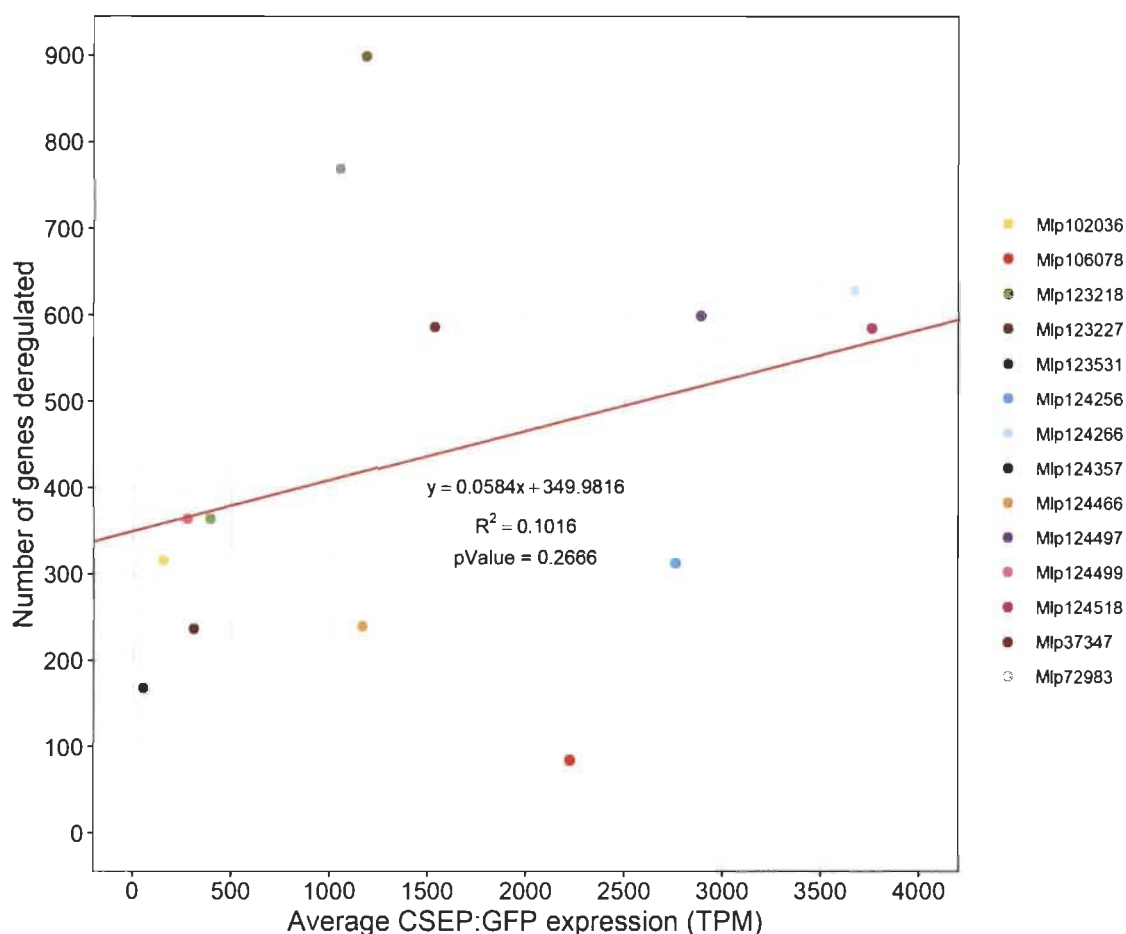


Figure S1. Magnitude of impact of CE on the plant's transcriptome is independent of its level of expression.

Reads not mapped to *Arabidopsis* genome were aligned to the transgene sequences (CE:GFP fusion) and average expression (in transcripts per million) across replicates of each transgenic line was calculated. Linear regression was performed using the number of genes deregulated in each transgenic line as the dependent variable and the average expression of the CE as the independent variable. The underlying data for this figure can be found at (dos Santos *et al.*, 2020).

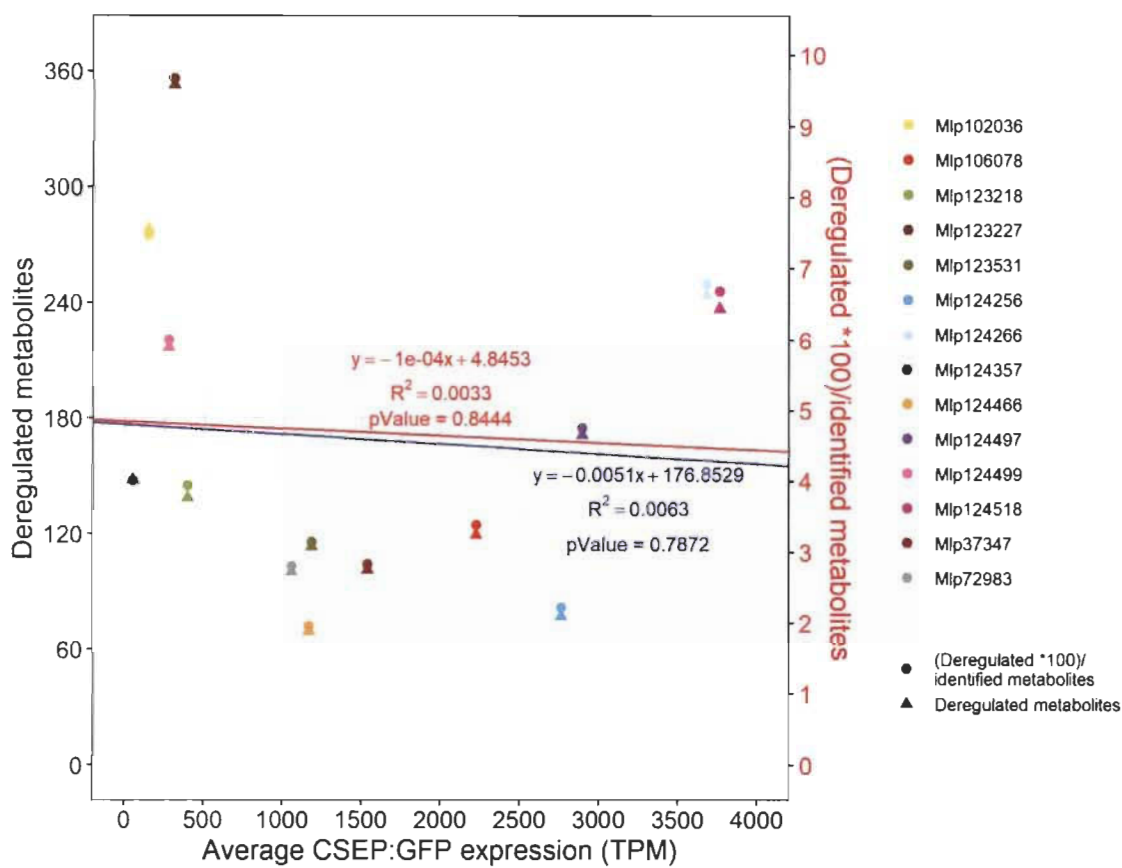


Figure S2. Magnitude of impact of CE on the plant's metabolome is independent of its level of expression, considering either the absolute number of deregulated metabolites (triangles, linear regression results in blue) or the ratio of metabolites deregulated by those identified (circles, linear regression results in red).

Reads not mapped to *Arabidopsis* genome were aligned to the transgene sequences (CE:GFP fusion) and average expression (in transcripts per million) across replicates of each transgenic line was calculated. Two separate linear regressions were performed using the number of metabolites deregulated and the ratio between metabolites deregulated by those detected in each transgenic line as the dependent variables and the average expression of the CE as the independent variable in both cases. The underlying data for this figure can be found at (dos Santos *et al.*, 2020).

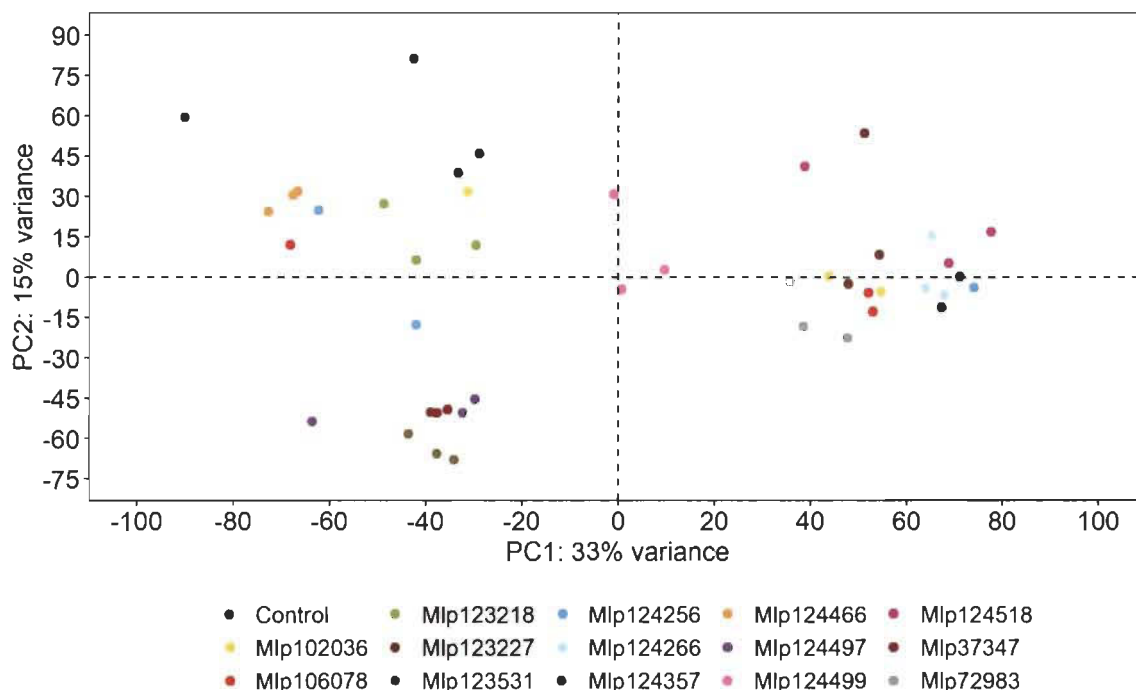
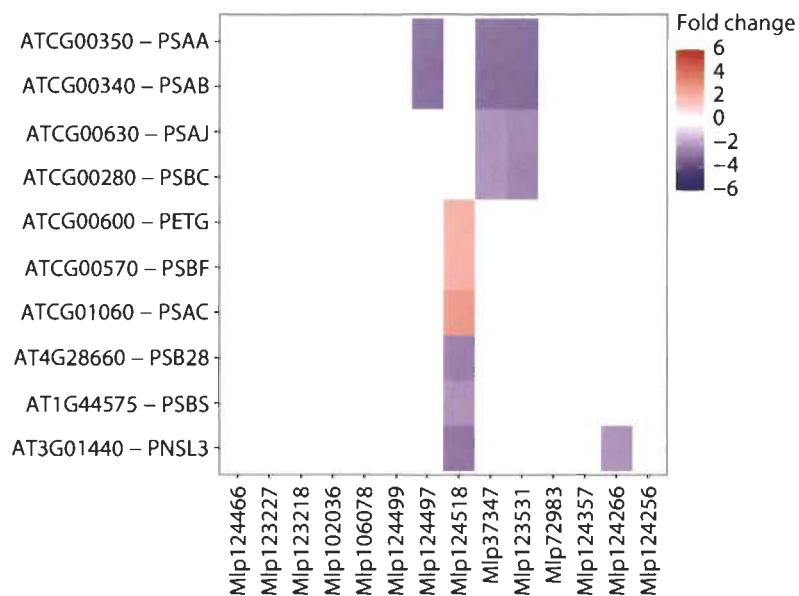


Figure S3. Principal component analysis of the replicates of 14 transgenic lines expressing candidate effectors from *Melampsora larici-populina* attached to GFP and a control line expressing only GFP (black dots).

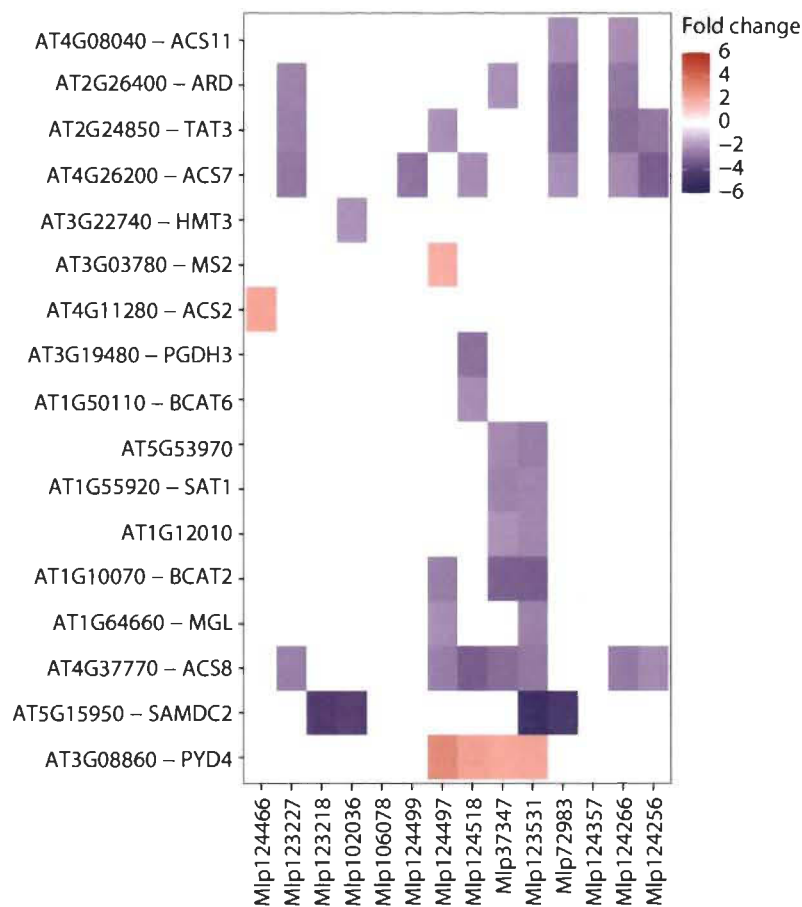
Replicates of the same transgenic lines are close together, indicating the homogeneity of the sample, with exception of one replicate of each of the following transgenic lines: Mlp102036 (yellow), Mlp106078 (red), Mlp124256 (sky blue) and Mlp124357 (dark green). The underlying data for this figure can be found at (dos Santos *et al.*, 2020).

File S1. Heatmaps of deregulated genes separated by KEGG pathway.

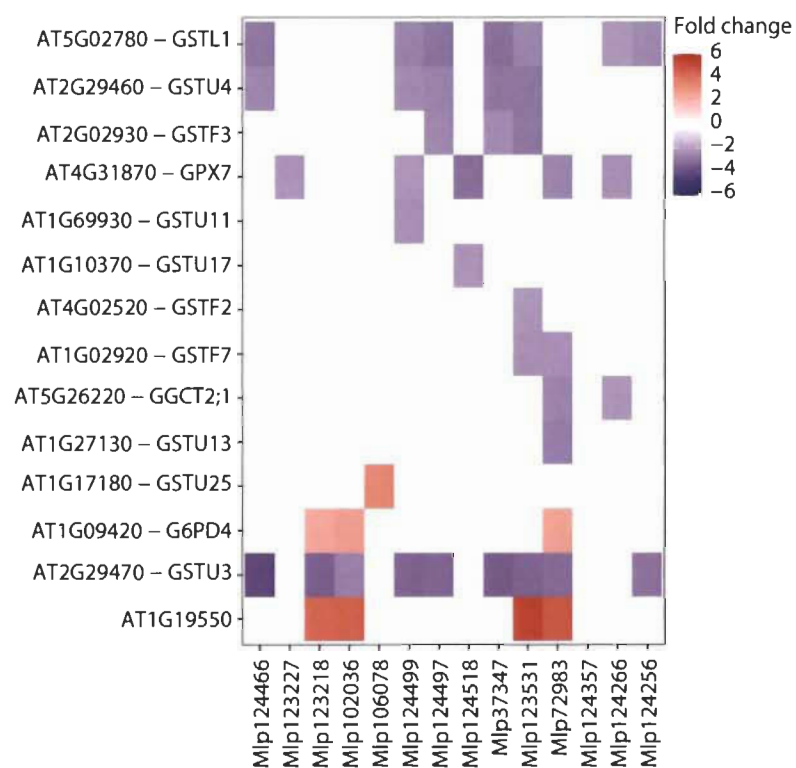
Pathways with at least 10 genes deregulated across the experiment were selected for display of genes deregulated in each transgenic line. The underlying data for this file can be found at dos Santos *et al.* (dos Santos *et al.*, 2020).



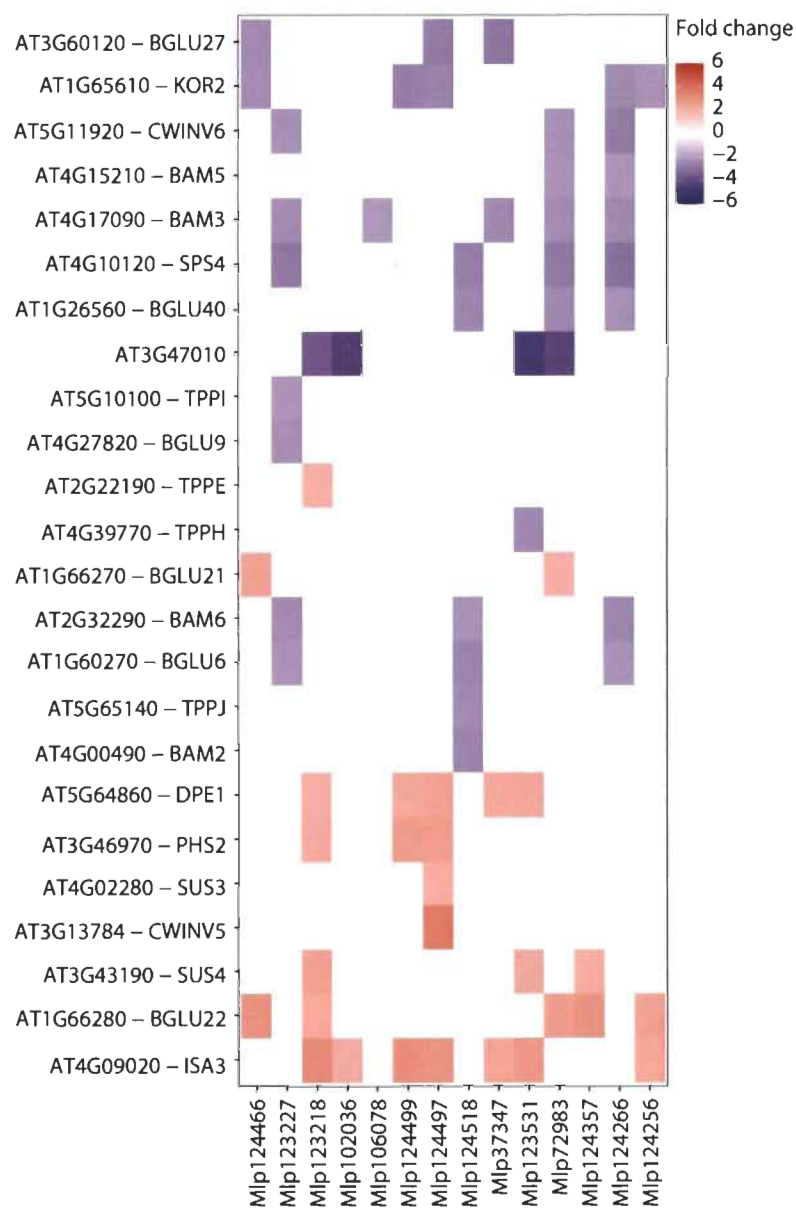
File S1. Figure 1. Photosynthesis.



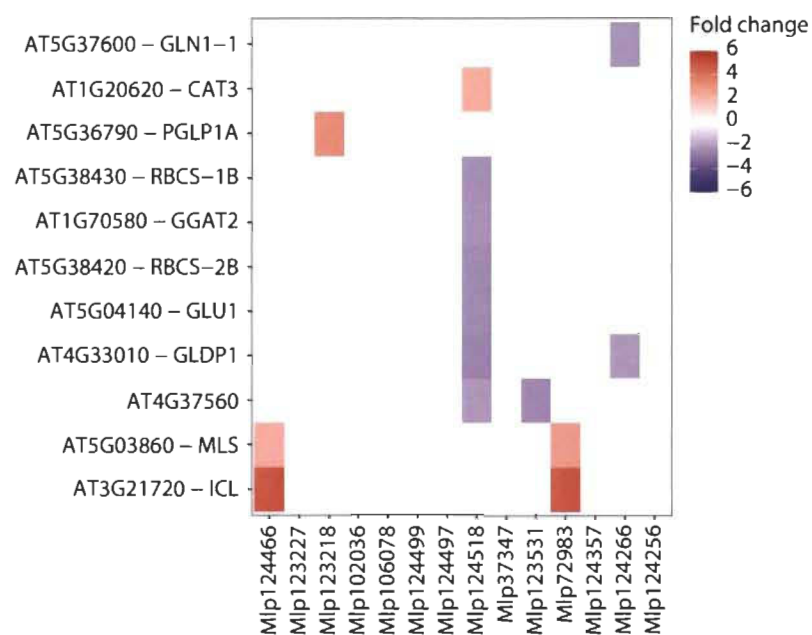
File S1. Figure 2. Cysteine and methionine metabolism.



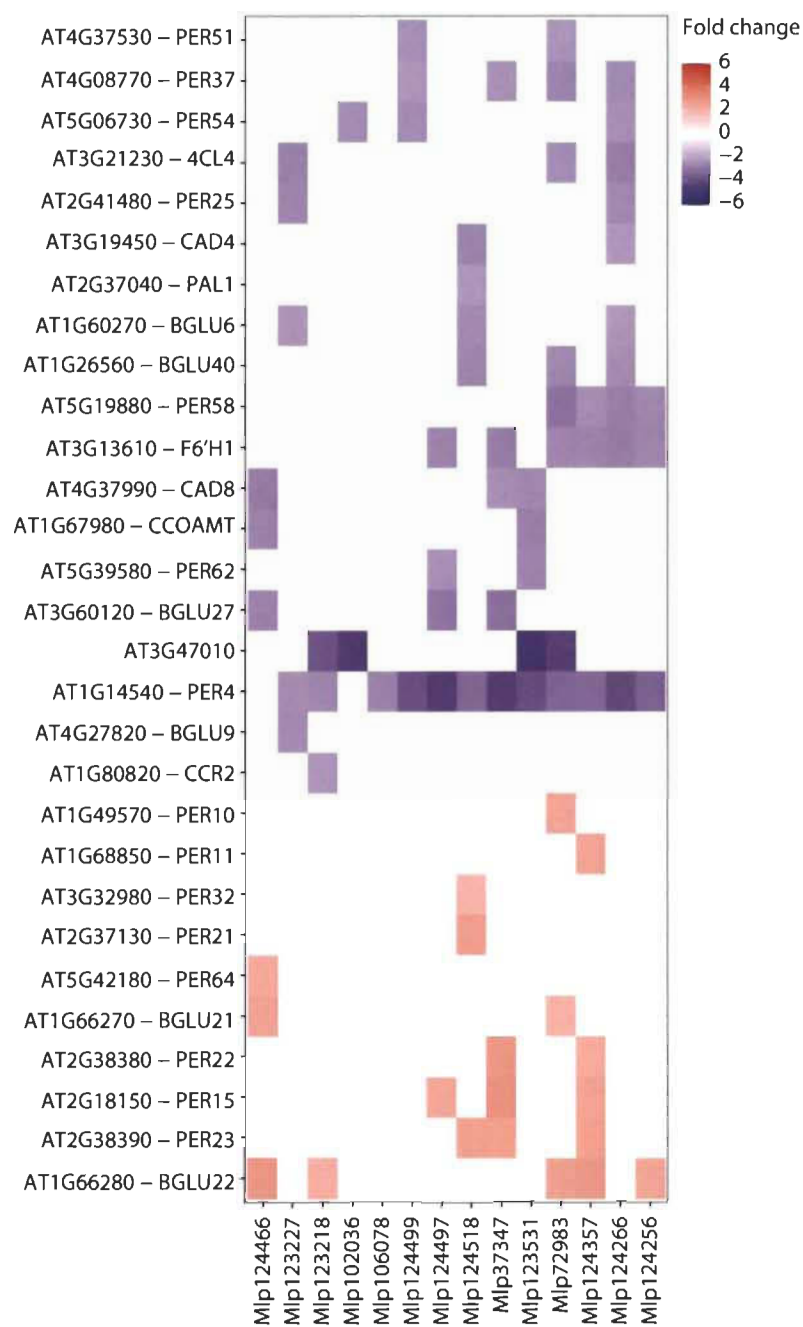
File S1. Figure 3. Glutathione metabolism.



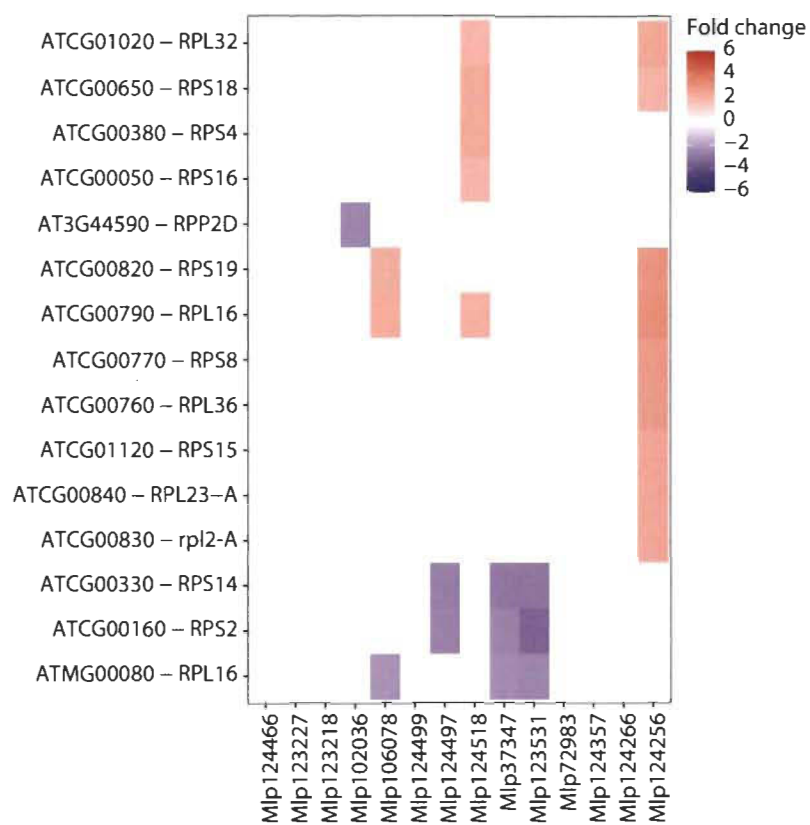
File S1. Figure 4. Starch and sucrose metabolism.



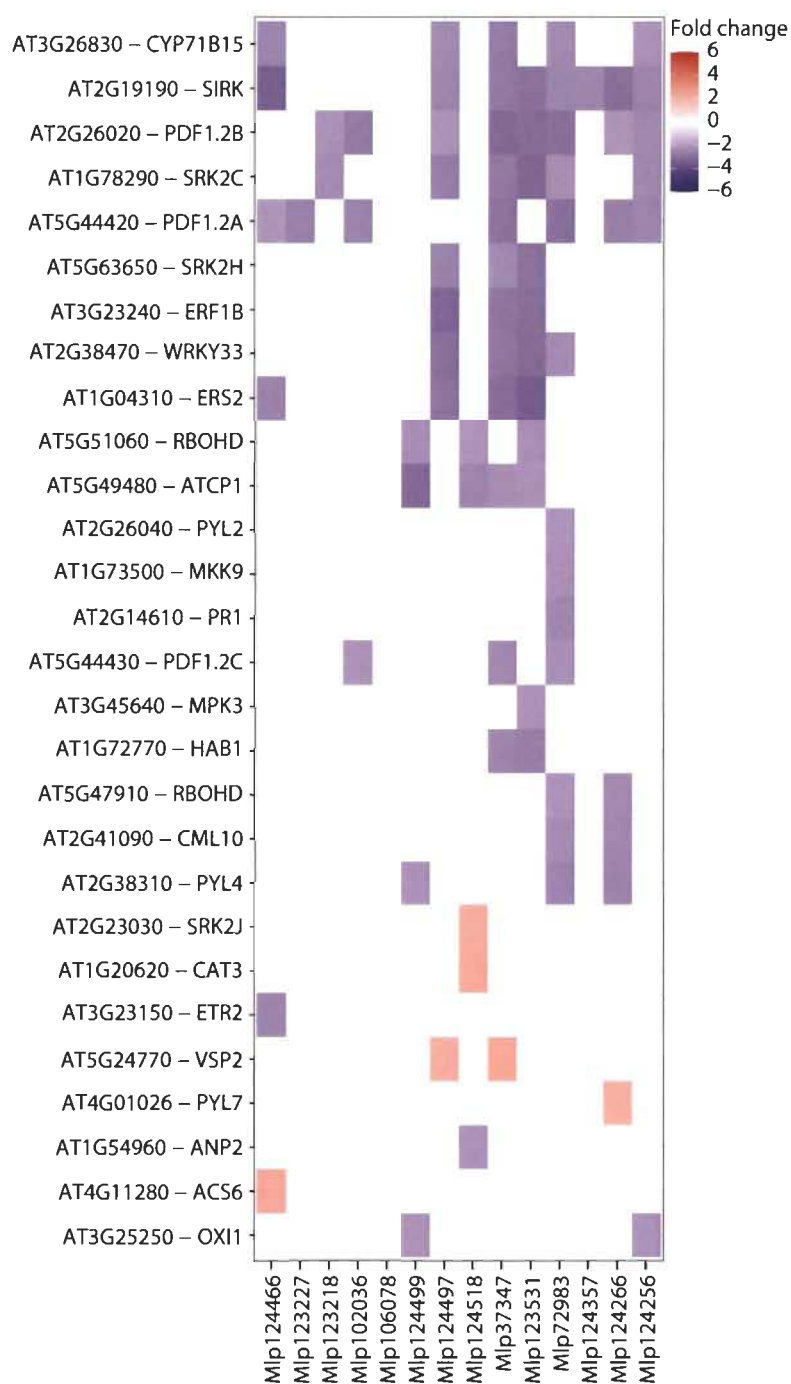
File S1. Figure 5. Glyoxylate and dicarboxylate metabolism.



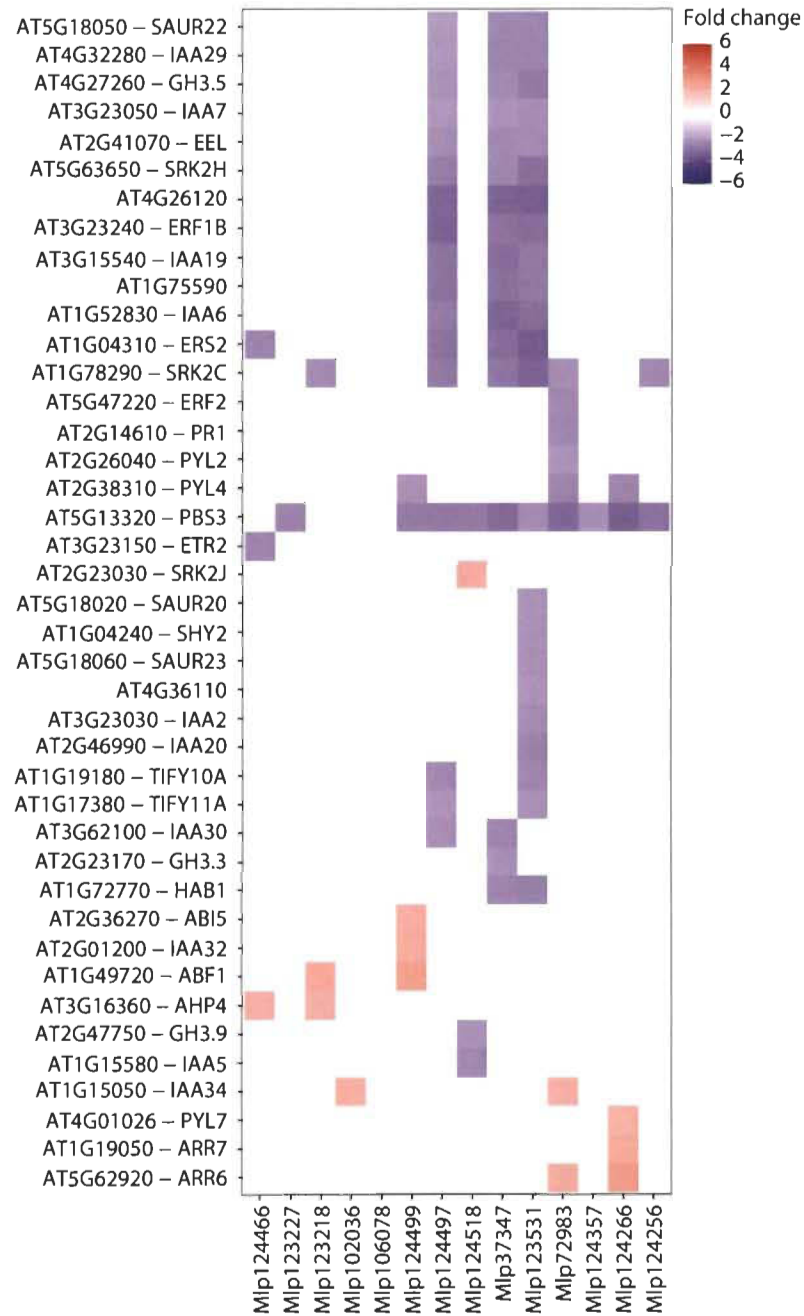
File S1. Figure 6. Phenylpropanoid biosynthesis.



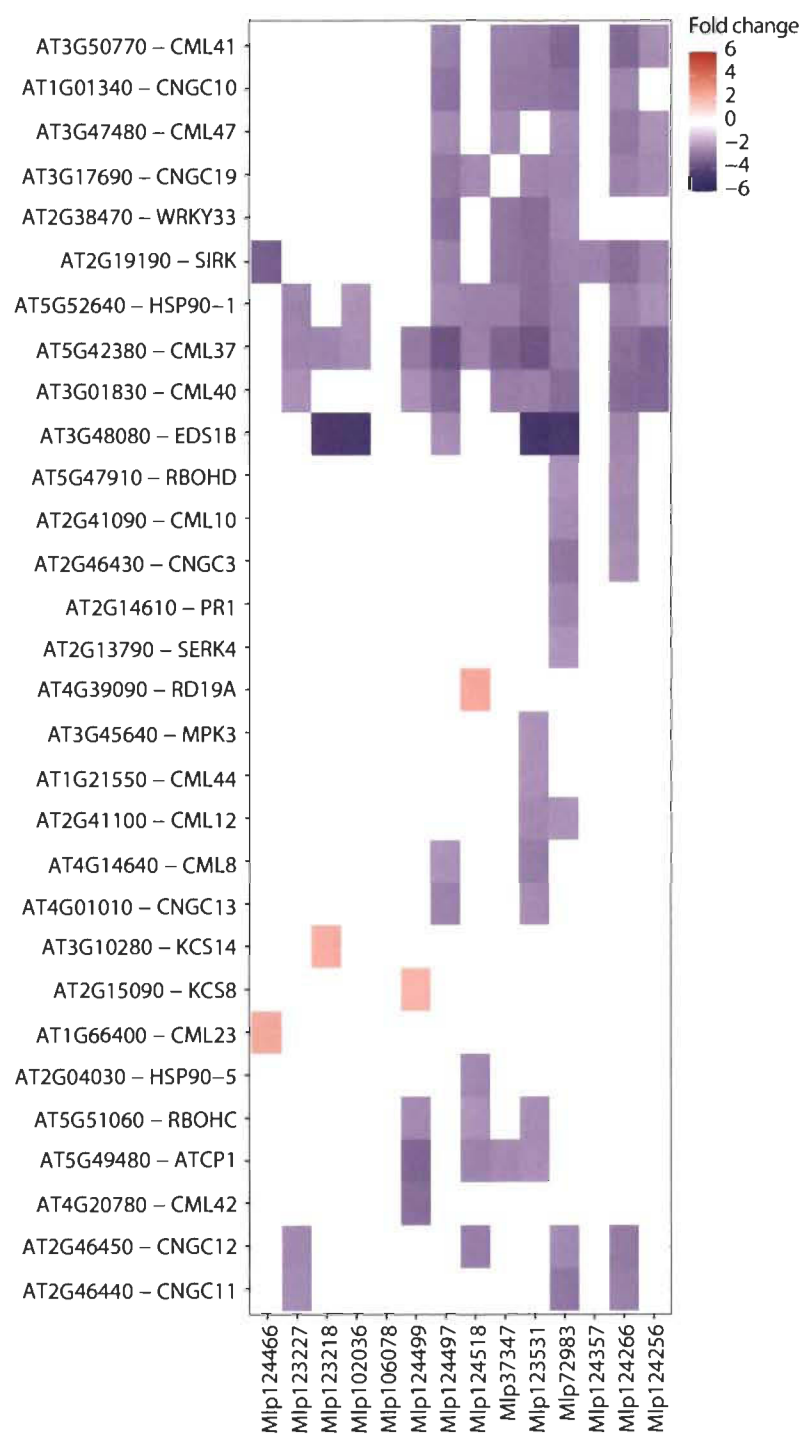
File S1. Figure 7. Ribosome.



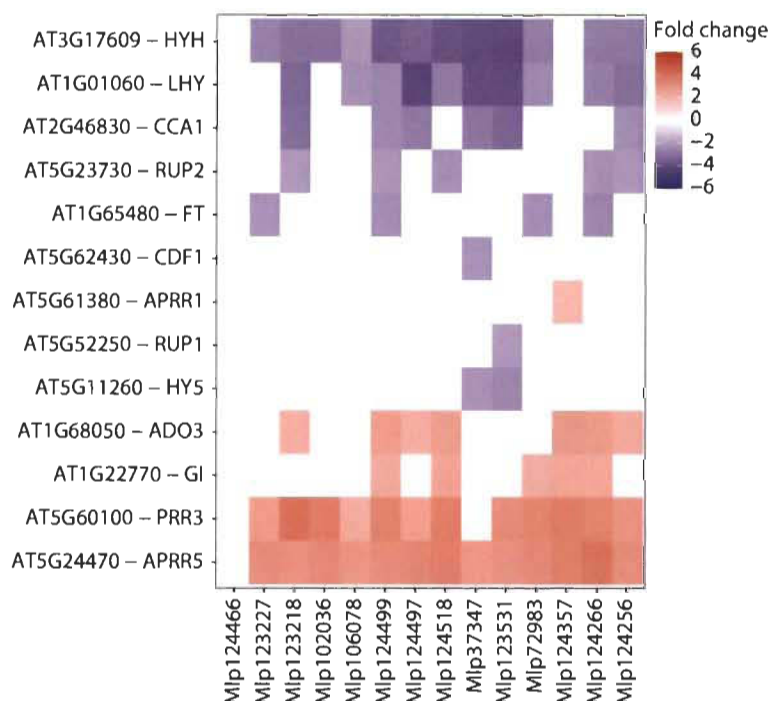
File S1. Figure 8. MAPK signaling pathway.



File S1. Figure 9. Plant-hormone signal transduction.



File S1. Figure 10. Plant-pathogen interaction.



File S1. Figure 11. Circadian rhythm.

Table S1. List of deregulated genes across the experiment with log2-transformed fold changes (FC) and false discovery rates (FDR) for each transgenic line.

Table S2. Percentage of identity and similarity, presented as “ID (SIM)”, calculated with pairwise sequence alignment of CEs using Needle.

	Mlp37347	Mlp72983	Mlp102036	Mlp106078	Mlp123218
Mlp37347	100 (100)	3.1 (4.9)	14.1 (28.9)	8.8 (13.8)	1.6 (2.0)
Mlp72983		100 (100)	12.2 (14.1)	13.8 (17.3)	22.2 (31.7)
Mlp102036			100 (100)	11 (14.2)	8.1 (12.9)
Mlp106078				100 (100)	12.4 (16.1)
Mlp123218					100 (100)
Mlp123227					
Mlp123531					
Mlp124256					
Mlp124266					
Mlp124357					
Mlp124466					
Mlp124497					
Mlp124499					
Mlp124518					

Table S2. Continuation.

	Mlp123227	Mlp123531	Mlp124256	Mlp124266	Mlp124357
Mlp37347	8.1 (12.4)	8.8 (13.1)	2.9 (5.9)	10.7 (16.1)	9.3 (11.0)
Mlp72983	12.0 (16.8)	9.4 (11.3)	3.9 (5.6)	9.6 (12.5)	7.5 (11.3)
Mlp102036	0.6 (1.1)	9.7 (16.1)	7.6 (13.6)	7.0 (14.9)	7.9 (11.8)
Mlp106078	6.7 (8.4)	6.8 (9.3)	9.2 (14.8)	10.1 (17.1)	7.1 (8.3)
Mlp123218	3.9 (6.7)	7.1 (14.8)	4.6 (7.8)	11.6 (17.6)	9.3 (12.5)
Mlp123227	100 (100)	1.1 (1.1)	1.3 (1.3)	0.6 (1.2)	1.2 (3.7)
Mlp123531		100 (100)	8.0 (8.8)	8.9 (14.3)	9.8 (9.8)
Mlp124256			100 (100)	38.2 (45.6)	22.7 (27.3)
Mlp124266				100 (100)	19.2 (21.2)
Mlp124357					100 (100)
Mlp124466					
Mlp124497					
Mlp124499					
Mlp124518					

Table S2. Continuation.

	Mlp124466	Mlp124497	Mlp124499	Mlp124518
Mlp37347	10.2 (18.0)	9.8 (11.2)	2.5 (7.0)	14.0 (16.8)
Mlp72983	3.6 (4.0)	6.6 (8.0)	6.0 (11.5)	4.4 (9.3)
Mlp102036	8.7 (15.5)	8.8 (13.3)	6.7 (13.5)	8.1 (9.9)
Mlp106078	2.8 (6.2)	7.5 (11.3)	10.3 (13.5)	5.9 (8.1)
Mlp123218	5.1 (10.2)	8.1 (12.6)	6.3 (12.1)	7.9 (14.2)
Mlp123227	1.4 (2.1)	0.7 (0.7)	2.1 (2.1)	6.5 (13.0)
Mlp123531	12.4 (17.5)	10.8 (17.2)	4.5 (5.5)	4.2 (5.9)
Mlp124256	10.5 (14.0)	15.7 (22.9)	14.8 (17.3)	13.9 (24.1)
Mlp124266	0.9 (3.7)	17.1 (28.9)	13.0 (24.7)	0.9 (2.6)
Mlp124357	4.3 (4.3)	19.8 (34.6)	14.1 (17.6)	13.3 (2.0)
Mlp124466	100 (100)	12.8 (17.9)	21.5 (33.8)	12.0 (16.0)
Mlp124497		100 (100)	41.8 (54.5)	20.0 (38.3)
Mlp124499			100 (100)	35.7 (53.6)
Mlp124518				100 (100)

Table S3. Summary of metabolomic analysis in negative mode of extractions with 20% and 80% methanol combined.

Assigned, CHO, CHON and Mean mass refer exclusively to the sample in question, while the amount of deregulated formulas considers those *m/z* detected in the sample or in the Control.

Sample	Assigned	CHO	CHON	Mean mass	Highly unsaturated and phenolic
Mlp37347	2954	1800	1154	451.517	1509
Mlp72983	3056	2171	885	461.914	1655
Mlp102036	3045	1893	1152	442.952	1539
Mlp106078	2796	1691	1105	445.203	1454
Mlp123218*	2695	1710	985	431.274	1398
Mlp123227	2679	1663	1016	439.461	1320
Mlp123531	3073	1946	1127	450.879	1610
Mlp124256	2815	1887	928	453.082	1495
Mlp124266	2754	1679	1075	435.403	1340
Mlp124357	3151	1983	1168	448.153	1599
Mlp124466	3026	1969	1057	450.488	1587
Mlp124497	2937	1801	1136	449.426	1464
Mlp124499	2946	1878	1068	443.459	1493
Mlp124518	2735	1651	1084	437.807	1345
Control	3113	2097	1016	458.505	1712

Table S3. Continuation

Sample	Aliphatic	Peptides	Polyphenolics	Condensed aromatics	Sugars
Mlp37347	822	402	127	47	47
Mlp72983	887	285	125	40	64
Mlp102036	802	448	145	64	47
Mlp106078	769	351	128	54	40
Mlp123218*	715	379	118	53	32
Mlp123227	754	378	122	59	46
Mlp123531	815	422	125	57	44
Mlp124256	829	248	134	57	52
Mlp124266	798	395	122	58	41
Mlp124357	892	428	131	53	48
Mlp124466	863	361	122	50	43
Mlp124497	844	418	122	52	37

Sample	Aliphatic	Peptides	Polyphenolics	Condensed aromatics	Sugars
Mlp124499	827	375	150	63	38
Mlp124518	767	400	122	61	40
Control	835	309	143	56	58

Table S3. Continuation

Sample	Identified	Deregulated (%)	Up-regulated (%)	Down-regulated (%)
Mlp37347	3576	101 (2.82)	58 (1.62)	43 (1.20)
Mlp72983	3575	100 (2.80)	22 (0.62)	78 (2.18)
Mlp102036	3719	278 (7.48)	164 (4.41)	114 (3.07)
Mlp106078	3522	119 (3.38)	46 (1.31)	73 (2.07)
Mlp123218*	3515	139 (3.95)	97 (2.76)	42 (1.19)
Mlp123227	3645	353 (9.68)	172 (4.72)	181 (4.97)
Mlp123531	3599	113 (3.14)	70 (1.94)	43 (1.19)
Mlp124256	3470	77 (2.22)	25 (0.72)	52 (1.50)
Mlp124266	3595	244 (6.79)	90 (2.50)	154 (4.28)
Mlp124357	3689	148 (4.01)	64 (1.73)	84 (2.28)
Mlp124466	3530	69 (1.95)	24 (0.68)	45 (1.27)
Mlp124497	3601	171 (4.75)	82 (2.28)	89 (2.47)
Mlp124499	3613	217 (6.01)	118 (3.27)	99 (2.74)
Mlp124518	3547	237 (6.68)	117 (3.30)	120 (3.38)
Control	3113	-	-	-

Table S4. Metabolites assigned and deregulated in each sample separated by category.

Identified metabolites are *m/z* detected either in the sample or in the control. The percentages were calculated by dividing the number of formulas assigned or deregulated in the sample in each category by the number of formulas identified in that sample and multiplying by 100.

Sample	Compound category	Identified (%)	Assigned	Deregulated (%)	Up (%)	Down (%)
Control	Polyphenolic	143	-	-	-	-
	Highly unsaturated and phenolic	1712	-	-	-	-
	Peptide-like	309	-	-	-	-
	Aliphatic	835	-	-	-	-
	Sugar	58	-	-	-	-
	Condensed aromatic	56	-	-	-	-

Sample	Compound category	Identified (%)	Assigned	Deregulated (%)	Up (%)	Down (%)
Mlp37347	Polyphenolic	1947 (54.45)	1509	57 (1.59)	26 (0.73)	31 (0.87)
	Highly unsaturated and phenolic	437 (12.22)	402	20 (0.56)	20 (0.56)	0 (0.00)
	Peptide-like	61 (1.71)	47	2 (0.06)	1 (0.03)	1 (0.03)
	Aliphatic	166 (4.64)	127	8 (0.22)	5 (0.14)	3 (0.08)
	Sugar	67 (1.87)	47	1 (0.03)	1 (0.03)	0 (0.00)
	Condensed aromatic	898 (25.11)	822	13 (0.36)	5 (0.14)	8 (0.22)
Mlp72983	Polyphenolic	1966 (54.99)	1655	63 (1.76)	13 (0.36)	50 (1.40)
	Highly unsaturated and phenolic	366 (10.24)	285	13 (0.36)	2 (0.06)	11 (0.31)
	Peptide-like	62 (1.73)	40	6 (0.17)	0 (0.00)	6 (0.17)
	Aliphatic	166 (4.64)	125	7 (0.20)	1 (0.03)	6 (0.17)
	Sugar	80 (2.24)	64	2 (0.06)	0 (0.00)	2 (0.06)
	Condensed aromatic	935 (26.15)	887	9 (0.25)	6 (0.17)	3 (0.08)
Mlp102036	Polyphenolic	2004 (53.89)	1539	158 (4.25)	68 (1.83)	90 (2.42)
	Highly unsaturated and phenolic	495 (13.31)	448	75 (2.02)	69 (1.86)	6 (0.16)
	Peptide-like	75 (2.02)	64	3 (0.08)	2 (0.05)	1 (0.03)
	Aliphatic	178 (4.79)	145	14 (0.38)	7 (0.19)	7 (0.19)
	Sugar	70 (1.88)	47	6 (0.16)	3 (0.08)	3 (0.08)
	Condensed aromatic	897 (24.12)	802	22 (0.59)	15 (0.40)	7 (0.19)
Mlp106078	Polyphenolic	1942 (55.14)	1454	78 (2.21)	25 (0.71)	53 (1.50)
	Highly unsaturated and phenolic	400 (11.36)	351	17 (0.48)	14 (0.40)	3 (0.09)
	Peptide-like	66 (1.87)	54	2 (0.06)	0 (0.00)	2 (0.06)
	Aliphatic	172 (4.88)	128	9 (0.26)	1 (0.03)	8 (0.23)
	Sugar	64 (1.82)	40	0 (0.00)	0 (0.00)	0 (0.00)
	Condensed aromatic	878 (24.93)	769	13 (0.37)	6 (0.17)	7 (0.20)
Mlp123218	Polyphenolic	1895 (53.91)	1398	67 (1.91)	36 (1.02)	31 (0.88)
	Highly unsaturated and phenolic	444 (12.63)	379	53 (1.51)	49 (1.39)	4 (0.11)
	Peptide-like	69 (1.96)	53	3 (0.09)	2 (0.06)	1 (0.03)
	Aliphatic	166 (4.72)	118	4 (0.11)	2 (0.06)	2 (0.06)
	Sugar	62 (1.76)	32	2 (0.06)	0 (0.00)	2 (0.06)
	Condensed aromatic	879 (25.01)	715	10 (0.28)	8 (0.23)	2 (0.06)

Sample	Compound category	Identified (%)	Assigned	Deregulated (%)	Up (%)	Down (%)
Mlp123227	Polyphenolic	1974 (54.16)	1320	214 (5.87)	71 (1.95)	143 (3.92)
	Highly unsaturated and phenolic	439 (12.04)	378	59 (1.62)	54 (1.48)	5 (0.14)
	Peptide-like	77 (2.11)	59	12 (0.33)	5 (0.14)	7 (0.19)
	Aliphatic	182 (4.99)	122	20 (0.55)	8 (0.22)	12 (0.33)
	Sugar	76 (2.09)	46	8 (0.22)	4 (0.11)	4 (0.11)
	Condensed aromatic	897 (24.61)	754	40 (1.10)	30 (0.82)	10 (0.27)
Mlp123531	Polyphenolic	1951 (54.21)	1610	59 (1.64)	28 (0.78)	31 (0.86)
	Highly unsaturated and phenolic	449 (12.48)	422	28 (0.78)	27 (0.75)	1 (0.03)
	Peptide-like	68 (1.89)	57	2 (0.06)	1 (0.03)	1 (0.03)
	Aliphatic	162 (4.50)	125	6 (0.17)	2 (0.06)	4 (0.11)
	Sugar	68 (1.89)	44	4 (0.11)	1 (0.03)	3 (0.08)
	Condensed aromatic	901 (25.03)	815	14 (0.39)	11 (0.31)	3 (0.08)
Mlp124256	Polyphenolic	1916 (55.22)	1495	49 (1.41)	14 (0.40)	35 (1.01)
	Highly unsaturated and phenolic	340 (9.80)	248	10 (0.29)	3 (0.09)	7 (0.20)
	Peptide-like	68 (1.96)	57	4 (0.12)	3 (0.09)	1 (0.03)
	Aliphatic	171 (4.93)	134	1 (0.03)	0 (0.00)	1 (0.03)
	Sugar	77 (2.22)	52	4 (0.12)	2 (0.06)	2 (0.06)
	Condensed aromatic	898 (25.88)	829	9 (0.26)	3 (0.09)	6 (0.17)
Mlp124266	Polyphenolic	1936 (53.85)	1340	153 (4.26)	34 (0.95)	119 (3.31)
	Highly unsaturated and phenolic	438 (12.18)	395	32 (0.89)	26 (0.72)	6 (0.17)
	Peptide-like	74 (2.06)	58	9 (0.25)	4 (0.11)	5 (0.14)
	Aliphatic	168 (4.67)	122	14 (0.39)	5 (0.14)	9 (0.25)
	Sugar	71 (1.97)	41	8 (0.22)	5 (0.14)	3 (0.08)
	Condensed aromatic	908 (25.26)	798	28 (0.78)	16 (0.45)	12 (0.33)
Mlp124357	Polyphenolic	1983 (53.75)	1599	94 (2.55)	27 (0.73)	67 (1.82)
	Highly unsaturated and phenolic	459 (12.44)	428	24 (0.65)	23 (0.62)	1 (0.03)
	Peptide-like	67 (1.82)	53	6 (0.16)	2 (0.05)	4 (0.11)
	Aliphatic	169 (4.58)	131	9 (0.24)	4 (0.11)	5 (0.14)
	Sugar	71 (1.92)	48	2 (0.05)	0 (0.00)	2 (0.05)
	Condensed aromatic	940 (25.48)	892	13 (0.35)	8 (0.22)	5 (0.14)

Sample	Compound category	Identified (%)	Assigned	Deregulated (%)	Up (%)	Down (%)
Mlp124466	Polyphenolic	1920 (54.39)	1587	50 (1.42)	14 (0.40)	36 (1.02)
	Highly unsaturated and phenolic	402 (11.39)	361	5 (0.14)	3 (0.08)	2 (0.06)
	Peptide-like	62 (1.76)	50	0 (0.00)	0 (0.00)	0 (0.00)
	Aliphatic	164 (4.65)	122	4 (0.11)	1 (0.03)	3 (0.08)
	Sugar	66 (1.87)	43	0 (0.00)	0 (0.00)	0 (0.00)
	Condensed aromatic	916 (25.95)	863	10 (0.28)	6 (0.17)	4 (0.11)
Mlp124497	Polyphenolic	1951 (54.18)	1464	99 (2.75)	30 (0.83)	69 (1.92)
	Highly unsaturated and phenolic	448 (12.44)	418	44 (1.22)	44 (1.22)	0 (0.00)
	Peptide-like	64 (1.78)	52	3 (0.08)	0 (0.00)	3 (0.08)
	Aliphatic	162 (4.50)	122	6 (0.17)	0 (0.00)	6 (0.17)
	Sugar	67 (1.86)	37	3 (0.08)	0 (0.00)	3 (0.08)
	Condensed aromatic	909 (25.24)	844	16 (0.44)	8 (0.22)	8 (0.22)
Mlp124499	Polyphenolic	1960 (54.25)	1493	134 (3.71)	48 (1.33)	86 (2.38)
	Highly unsaturated and phenolic	423 (11.71)	375	48 (1.33)	47 (1.30)	1 (0.03)
	Peptide-like	71 (1.97)	63	2 (0.06)	1 (0.03)	1 (0.03)
	Aliphatic	185 (5.12)	150	12 (0.33)	5 (0.14)	7 (0.19)
	Sugar	65 (1.80)	38	2 (0.06)	1 (0.03)	1 (0.03)
	Condensed aromatic	909 (25.16)	827	19 (0.53)	16 (0.44)	3 (0.08)
Mlp124518	Polyphenolic	1913 (53.93)	1345	145 (4.09)	47 (1.33)	98 (2.76)
	Highly unsaturated and phenolic	438 (12.35)	400	54 (1.52)	53 (1.49)	1 (0.03)
	Peptide-like	72 (2.03)	61	4 (0.11)	3 (0.08)	1 (0.03)
	Aliphatic	171 (4.82)	122	11 (0.31)	5 (0.14)	6 (0.17)
	Sugar	68 (1.92)	40	5 (0.14)	2 (0.06)	3 (0.08)
	Condensed aromatic	885 (24.95)	767	18 (0.51)	7 (0.20)	11 (0.31)

Table S5. Parameters used for bioinformatic analyses.

Analysis step	Software	Parameters
Read trimming and filtering	Trimmomatic	phred 33 LEADING:3 3TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:25
Read alignment	HISAT2	default parameters and mate inner distance according to the replicate
Read count	GenomicFeatures	With makeTxDbFromBiomart: biomart = "plants_mart" dataset = "athaliana_eg_gene" id_prefix = "ensembl_" host = "plants.ensembl.org" taxonomyId = 3702 transcriptsBy("gene") summarizeOverlaps: mode = Union
	GenomicAlignments	singleEnd = F ignore.strand = FALSE fragments = T
Filtering out weakly expressed genes	CustomSeletion	mean(TPM) < mean(DAFS cutoff)
Variation between replicates and samples	R 4.0.0	plotPCA: ntop = All <i>Arabidopsis</i> genes expressed in samples estimateSizeFactors using controlGenes
Differential expression analysis	DESeq2	DESeq: betaPrior = T log2(Fold change) ≥ 2 adjusted pValue ≤ 0.01 enrichGO: universe = All <i>Arabidopsis</i> genes expressed in samples OrgDb = "org.At.tair.db" ont = "BP"
	clusterProfiler	keyType = "TAIR" pAdjustMethod = "BH" pvalueCutoff = 0.01 qvalueCutoff = 0.05 readable = TRUE
GO enrichment		simplify
KEGG enrichment	KEGGprofile	find_enriched_pathway: species = "ath" download_latest = TRUE

Analysis step	Software	Parameters
Co-expression network analysis	WGCNA	blockwiseModules: log2-transformed TPM values of deregulated genes power = 14 TOMType = “unsigned” minModuleSize = 30 reassignThreshold = 0 mergeCutHeight = 0.3 numericLabels = TRUE pamRespectsDendro = FALSE

Table S6. Sequencing results and alignment summary.

Sample name	Number of Reads	Average Quality	Mate inner distance	Duplicate (%)	Surviving	Aligned	Aligned (%)
CNT-1	12 959 539	37	50	39.57	11 420 407	6 578 769	58
CNT-2	17 628 091	39	60	24.15	16 441 057	15 326 908	93
CNT-3	44 449 775	39	84	32.81	40 740 349	38 447 039	94
102036-1	14 725 402	38	115	21.91	13 749 600	12 482 030	91
102036-2	16 097 041	38	110	24.53	14 863 654	13 794 120	93
102036-3	18 644 518	38	110	19.00	17 175 540	14 121 101	82
106078-1	14 806 842	38	70	27.15	13 707 505	12 657 991	92
106078-2	16 097 885	38	70	23.93	14 896 296	13 620 408	91
106078-3	57 392 974	39	55	35.94	52 226 039	49 146 796	94
123218-1	14 941 915	38	60	20.13	13 911 559	12 766 811	92
123218-2	19 202 251	38	60	24.98	17 740 615	16 007 510	90
123218-3	14 573 522	38	65	27.47	13 336 684	10 353 032	78
123227-1	14 506 123	38	60	23.68	13 559 794	12 650 378	93
123227-2	13 690 077	38	70	21.88	12 740 323	11 674 516	92
123227-3	14 221 348	38	65	25.24	13 044 995	10 662 871	82
123531-1	30 255 033	39	70	53.11	28 677 354	26 311 924	92
123531-2	29 847 359	39	60	46.49	28 400 730	26 243 184	92
123531-3	33 185 737	39	65	30.72	31 504 015	29 142 533	93
124256-1	13 219 309	38	110	22.30	12 270 185	11 319 954	92
124256-2	15 023 134	38	60	23.12	13 721 513	11 218 900	82
124256-3	53 119 639	39	60	38.35	48 711 081	45 705 676	94
124266-1	14 512 508	37	60	21.17	13 288 961	12 184 913	92
124266-2	14 343 047	37	100	23.84	13 084 917	11 752 213	90
124266-3	16 738 628	37	70	21.41	15 203 780	13 696 801	90
124357-1	17 594 958	38	80	22.41	16 313 595	15 140 354	93
124357-2	15 759 938	38	115	21.54	14 538 561	13 287 744	91
124357-3	17 205 907	38	105	19.78	15 823 808	13 959 229	88
124466-1	15 994 909	38	150	21.84	14 896 010	13 790 242	93
124466-2	17 877 587	38	105	20.43	16 470 499	14 898 728	90

Sample name	Number of Reads	Average Quality	Mate inner distance	Duplicate (%)	Surviving	Aligned	Aligned (%)
124466-3	12 451 485	39	70	19.33	11 616 591	10 726 259	92
124497-1	15 410 965	39	70	21.36	14 559 104	13 634 149	94
124497-2	25 299 826	39	60	29.76	23 916 582	22 416 808	94
124497-3	14 250 049	39	70	24.61	12 724 994	11 533 107	91
124499-1	12 422 097	38	60	22.33	11 615 353	10 818 003	93
124499-2	14 079 283	38	60	21.04	13 059 459	11 939 187	91
124499-3	16 445 351	38	100	27.87	14 910 023	13 122 045	88
124518-1	14 485 964	36	110	23.32	13 094 223	11 731 759	90
124518-2	12 911 956	36	105	23.39	11 553 534	9 944 892	86
124518-3	11 347 886	34	45	21.35	9 346 542	8 246 506	88
37347-1	27 784 075	39	70	28.79	26 345 883	25 136 138	95
37347-2	23 439 580	38	65	31.35	21 914 868	20 722 973	95
37347-3	14 033 172	39	65	23.16	13 268 922	12 674 770	96
72983-1	23 393 778	38	55	21.00	21 669 105	19 896 426	92
72983-2	15 230 634	38	55	24.76	13 982 126	12 497 170	89
72983-3	14 014 784	38	64	24.18	12 928 472	11 095 184	86

CHAPTER IV

CONCLUSION

4.1 Conclusion

Obligate biotrophic plant pathogens are difficult organisms to study because their lifestyle imposes limitations to genetic manipulations. This is the case for rust fungi. However, for poplar rust, *M. larici-populina*, there are resources available: the genomes of both the pathogen and its telial host, poplar, are assembled and annotated. These allow evolutionary analyses and facilitate the interpretation of functional analyses in heterologous systems. *M. larici-populina*'s genome has 19 550 predicted protein-coding genes, of which 1 184 encode small secreted proteins. These are enriched among the most up-regulated genes during poplar infection, compared to urediniospores, indicating they might play a role in the infection process (Duplessis *et al.*, 2011a). From these proteins, CEs were prioritized for functional studies in heterologous systems (Ahmed *et al.*, 2018; Germain *et al.*, 2018; Madina *et al.*, 2020; Petre *et al.*, 2016), revealing that they accumulate in specific cellular compartments, interact with plant proteins or DNA and impact *Arabidopsis* susceptibility to bacterial and/or oomycete infection (Ahmed *et al.*, 2018; Germain *et al.*, 2018; Madina *et al.*, 2020; Petre *et al.*, 2015).

In the current study, we further analysed, in *Arabidopsis*, the impact of 14 of these prioritized CEs from *M. larici-populina* (presented in Table 1 – Chapter III) using transcriptomics and metabolomics. We used the algorithm DESeq2 for differential expression analysis, which is suggested for experiments with five or less replicates per sample (Costa-Silva *et al.*, 2017; Quinn *et al.*, 2018), with reference genes for the calculation of size factors. However, it was not possible to know *a priori* which genes were impacted by the selected CEs. We selected several housekeeping genes, such as Actin and Tubulin commonly used as references in RT-qPCR analyses, but these showed high level of variability in expression in our samples (**Figure 1** – Chapter II). To solve

this issue, we developed a R package for selection of reference genes using read counts from RNAseq studies (dos Santos *et al.*, 2020a). This package is user-friendly and can be particularly beneficial for transcriptomic studies on non-model organisms, given that it does not require prior knowledge on specific genes. We used this R package to calculate the average expression (in TPM) of *Arabidopsis* genes across all samples, to eliminate genes weakly expressed and to select the genes with lowest covariance of TPM levels as references.

The differential expression analysis of the CE-expressing *Arabidopsis* plants revealed that each CE impacted the plants in different levels, from 84 differentially expressed genes (DEGs), in the transgenic line Mlp106078, to 898 DEGs, in the line Mlp123531. For the metabolomic analysis of these transgenic plants, it was not possible to know *a priori* which metabolites could be impacted by the selected CEs, thus we used an untargeted approach. Although we employed ultra-high-resolution mass spectrometry for the detection of metabolites, the lack of information in public databases prevented us from identifying 4 207 out of the 5 192 metabolites detected (81.03%). For this reason, we did not evaluate the metabolic pathways deregulated by the CEs. However, the number of compounds in each sample and the patterns of metabolite deregulation in our transgenic lines do not change with the identity of the molecules detected. We found that the CEs had a differential impact in the metabolite concentrations in the plants: while plants expressing the CE Mlp124466 showed differential accumulation of 68 compounds, 352 were deregulated in the line Mlp123227.

This unique concurrent study of multiple CEs allowed us to compare their impact in the plant. For two CE families, we studied more than one member, namely for the family CPG5464 (AvrP4 homologues) we studied the CEs Mlp124256 and Mlp124266, and for the family CPGH1 we studied the CEs Mlp124497, Mlp124499 and Mlp124518. We found that the members of the same family did not deregulate the same genes (**Figure 2** – Chapter III) nor the same metabolites (**Figure 6B** – Chapter III, with the exception of Mlp124499 and Mlp124518). However, metabolite and gene deregulation patterns from plants expressing these CEs are similar to those from plants expressing CEs

belonging to different families (for gene deregulation: Mlp124497-Mlp37347, Mlp124499-Mlp124256, and Mlp123227-Mlp124266; for metabolite deregulation: Mlp124266-Mlp124357). In other words: CEs of the same family deregulate different genes and metabolites, and vice-versa.

Still, we cannot rule out the possibility that our results for each CE were impacted by the insertion site of the transgenes, since we only evaluated one transgenic line for each CE. Nevertheless, given that this work was conducted with 14 CE-expressing lines, it is unlikely that the insertion site impacted all of them in the same manner, supporting the results that consider all transgenic lines. Ideally, both the transcriptomic and metabolomic analyses would have been conducted with three independent transgenic lines for each CE, which would serve as biological replicates for the impact of the CEs, and three independent replicates for the individual transgenic lines. Yet, we can corroborate the results of individual CEs: RT-qPCR analyses of selected deregulated genes with two transgenic lines for each CE will be performed in the near future.

Another point to address is our metabolomic data. We pooled several seedlings into each replicate. After metabolite extraction, four replicates of each transgenic line were pooled into a single sample. Thus, the results presented for individual transgenic lines are derived from multiple plants. However, the lack of replication for the quantification of metabolites allowed only a comparison of metabolic profiles, but not the calculation of the statistical significance of the results. In addition, only 385 compounds were putatively annotated (matching a single known metabolite), 81.03% of the compounds we detected remained unidentified due to the absence of matching data in public databases. Other 600 compounds (11.56% of the metabolites detected) were ambiguously annotated, reaching a maximum of 45 matches for the m/z 179.056111 (molecular formula $C_6H_{12}O_6$). Although we could have performed chromatography coupled to mass spectrometry, this would have solely allowed the potential unambiguous annotation of these compounds whose m/z matched multiple known metabolites.

Here we presented a macroscopic comparison of the effector-triggered susceptibility caused by 14 CEs from *M. larici-populina* in *Arabidopsis* plants, through constitutive expression of the selected transgenes directly *in planta*. This heterologous system allows the high throughput study of effectors/CEs, however it has several caveats. First, *M. larici-populina* infects *Larix* spp. (gymnosperm) and *Populus* spp. (angiosperm, order Malpighiales) (Hacquard *et al.*, 2011), the impact of this fungus' proteins in *Arabidopsis* (angiosperm, order Brassicales) may be different from what they cause in the true hosts because of the evolutionary divergency between these plants. However, the CEs selected for this study show induced expression during poplar infection (Duplessis *et al.*, 2011a; Hacquard *et al.*, 2012), suggesting they play a role in the infection of this host. In addition, both poplar and *Arabidopsis* are eudicots of the Rosid clade (Zeng *et al.*, 2017; Zhu *et al.*, 2007). Second, effectors are not present in the plant constantly, they are secreted by the pathogen at specific moments of the infection (Duplessis *et al.*, 2011b; Hacquard *et al.*, 2012). The expression of CEs under the control of a constitutive promoter (in this case, the Cauliflower Mosaic Virus 35S promoter) may provoke plant alterations (for example, developmental changes) that do not take place in the timespan the plant is in contact with the CE in the true system. Finally, in our transgenic lines the CEs are expressed in all plant cells, but during the infection effectors are not ubiquitous: they are present in cells in contact with the pathogen (Hacquard *et al.*, 2012) or close to the pathogen (Djamei *et al.*, 2011a). These issues need to be considered when translating the results presented here to poplar-poplar rust and larch-poplar rust systems.

4.2 Perspectives

To understand how pathogens interact with their hosts and to shed light into what makes plants susceptible it is necessary to study CEs. We found that paralogous CEs had different impact in the plant, while CEs with low sequence similarity converged into the deregulation of the same genes and pathways. It is possible that CEs from the same family underwent functional diversification, as shown for Tin2 from *U. maydis* and *Sporosorium reilianum* (Tanaka *et al.*, 2019), GALA effectors from *R. solanacearum* (Remigi *et al.*, 2011), and HopAF from *P. savastanoi* (Castañeda-Ojeda *et al.*, 2017). On the other hand,

CEs from different families with similar function could have similar three-dimensional structure, which is the case of several CEs from *Blumeria graminis*: although they share sequence identity of about 20%, their predicted structure is conserved, similar to the structure of the ribonuclease T1 from *Aspergillus* sp. (Pedersen *et al.*, 2012). It is also possible that the CEs from the same family interact with similar host targets that play a role in different pathways, while CEs from different families may interact with different host targets in the same pathway (Win *et al.*, 2012). These avenues should be the subject of future studies with rust effectors. A more in-depth evaluation of the phenomenon uncovered here should compare the protein structure of several members of multiple CE families and their impact in the host.

We also showed in this work that the CEs prioritized impact many plant pathways, including the deregulation of important genes related to plant defense and hormone signaling (File S1 – Chapter III). This supports the hypothesis that these proteins are *bona fide* effectors. Of the 14 CEs from this work, two have been selected for in-depth analysis of their function: Mlp37347 (unpublished) and Mlp124357 (Madina *et al.*, 2020). In addition, the CEs Mlp124266 (CPG5464) and Mlp124499 (CPGH1) were the subject of a structural genomics study, however the three-dimensional structure of Mlp124499 could not be elucidated (de Guillen *et al.*, 2019). Mlp124266 has a structure similar to the knottin cycloviolacin O2, an antibiotic protein from *Viola odorata*, but the differences between the two proteins do not suggest that Mlp124266 has a similar antibiotic activity (de Guillen *et al.*, 2019). It remains to be evaluated if Mlp124256 (CPG5464) and other members of this family are also structurally similar to knottins, and more specifically to the membrane-interacting cycloviolacin O2.

Two other CEs deserve to be further analysed based on our data. Mlp123531 and Mlp72983 were the CEs which most deregulated the transcriptome in this study (**Figure 1** – Chapter III), including the down-regulation of genes related to photosynthesis, ethylene biosynthesis, nitrogen metabolism, ribosome, MAPK signaling, plant-pathogen interaction and hormone signaling (**File S1** – Chapter III). Whether these proteins impact the transcriptome directly, through interaction with DNA or with transcription factors,

or indirectly, remains to be elucidated. Future studies with these CEs should corroborate in the true host the results obtained in *Arabidopsis/N. benthamiana* (subcellular localization, and transcriptome and metabolome deregulations). In addition, questions to be answered in the future about these CEs include:

- These CEs down-regulated key pathogen-responsive genes in the absence of pathogen or stress condition, including PR1, RBOHD, PDF1.2A, PDF1.2B and AT4G26120 (homologous to NPR1, Mlp72983), RBOHC and MPK3 (Mlp123531) and WRKY33, several calmodulin-like proteins, PDF1.2B and EDS1B (both CEs). How these proteins impact the plant during stress conditions, namely pattern-triggered immunity?
- Are the transcriptomic changes caused by these CEs related to the deregulation of defense-related hormones, such as SA, JA, ET and ABA?

4.3 Final conclusions

Here we show that the similarity of effectors' sequences and of their impact in the plant are not correlated, and this finding is supported by few previous studies which investigated single effector families, showing that, in both *P. savastanoi* (Castañeda-Ojeda *et al.*, 2017) and in *R. solanacearum* (Remigi *et al.*, 2011), effectors from the same family affected plant immunity differently. This is significant because, in plant pathology, CEs are grouped into families to facilitate their study, the characteristics of whole families are considered when selecting priority candidates for functional analyses (Hacquard *et al.*, 2012; Saunders *et al.*, 2012), and high-throughput studies have selected one member per effector family for functional characterization (Germain *et al.*, 2018; Petre *et al.*, 2015). Thus, our results indicate that sequence similarity may misguide efforts for functional characterization CEs.

REFERENCES

- Abel, S., Nguyen, M. D., Chow, W., & Theologis, A. (1995). ACS4, a primary indoleacetic acid-responsive gene encoding 1-aminocyclopropane-1-carboxylate synthase in *Arabidopsis thaliana*. Structural characterization, expression in *Escherichia coli*, and expression characteristics in response to auxin. *Journal of Biological Chemistry*, 270(32), 19093-19099. doi: 10.1074/jbc.270.32.19093
- Adhikari, T. B., Bai, J., Meinhardt, S. W., Gurung, S., Myrfield, M., Patel, J., . . . Rasmussen, J. B. (2009). Tsn1-mediated host responses to ToxA from *Pyrenophora tritici-repentis*. *Molecular Plant-Microbe Interactions*, 22(9), 1056-1068. doi: 10.1094/MPMI-22-9-1056
- Aguilar, G. B., Pedersen, C., & Thordal-Christensen, H. (2016). Identification of eight effector candidate genes involved in early aggressiveness of the barley powdery mildew fungus. *Plant Pathology*, 65, 953-958. doi: 10.1111/ppa.12476
- Ahmed, M. B., Santos, K. C. G. d., Petre, B., Lorrain, C., Duplessis, S., Desgagne-Penix, I., & Germain, H. (2018). A rust fungal effector binds plant DNA and modulates transcription. *Nature Scientific Reports*, 8, 14718. doi: 10.1038/s41598-018-32825-0
- Ai, G., Yang, K., Ye, W., Tian, Y., Du, Y., Zhu, H., . . . Dou, D. (2020). Prediction and characterization of RXLR effectors in *Pythium* species. *Molecular Plant-Microbe Interactions*, 33(8), 1046-1058. doi: 10.1094/MPMI-01-20-0010-R
- Aime, M. C., McTaggart, A. R., Mondo, S. J., & Duplessis, S. (2017). Phylogenetics and phylogenomics of rust fungi. Dans Townsend, J. P. , & Wang, Z. *Fungal Phylogenetics and Phylogenomics* (pp. 267-307). United States: Academic Press.
- Angot, A., Peeters, N., Lechner, E., Vailleau, F., Baud, C., Gentzbittel, L., . . . Genin, S. (2006). *Ralstonia solanacearum* requires F-box-like domain-containing type III effectors to promote disease on several host plants. *PNAS*, 103(39), 14620-14625. doi: 10.1073/pnas.0509393103
- Armitage, A. D., Lysøe, E., Nellist, C. F., Lewis, L. A., Cano, L. M., Harrison, R. J., & Brurberg, M. B. (2018). Bioinformatic characterisation of the effector repertoire of the strawberry pathogen *Phytophthora cactorum*. *PLoS One*, 13(10), e0202305. doi: 10.1371/journal.pone.0202305

- Baka, Z. A., Larous, L., & Lösel, D. M. (1995). Distribution of ATPase activity at the host-pathogen interfaces of rust infections. *Physiological and Molecular Plant Pathology*, 47, 67-82. doi: 10.1006/pmpp.2000.0264
- Bent, A. F., Innes, R. W., Ecker, J. R., & Staskawicz, B. J. (1992). Disease development in ethylene-insensitive *Arabidopsis thaliana* infected with virulent and avirulent *Pseudomonas* and *Xanthomonas* pathogens. *Molecular Plant-Microbe Interactions*, 5(5), 372-378. doi: 10.1094/mpmi-5-372
- Bentham, A. R., De la Concepcion, J. C., Mukhi, N., Zdrzalek, R., Draeger, M., Gorenkin, D., . . . Banfield, M. J. (2020). A molecular roadmap to the plant immune system. *Journal of Biological Chemistry*, 295(44), 14916-14935. doi: 10.1074/jbc.REV120.010852
- Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., . . . 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(3), 200-211. doi: 10.1016/j.chom.2011.02.009
- Buendia, L., Girardin, A., Wang, T., Cottret, L., & Lefebvre, B. (2018). LysM receptor-like kinase and LysM receptor-like protein families: An update on phylogeny and functional characterization. *Frontiers in Plant Science*, 9(October), 1531. doi: 10.3389/fpls.2018.01531
- Büttner, D. (2016). Behind the lines—actions of bacterial type III effector proteins in plant cells. *FEMS Microbiology Reviews*, 40(6), 894-937. doi: 10.1093/femsre/fuw026
- Büttner, D., & He, S. Y. (2009). Type III protein secretion in plant pathogenic bacteria. *Plant Physiology*, 150(4), 1656-1664. doi: 10.1104/pp.109.139089
- Carreón-Anguiano, K. G., Islas-Flores, I., Vega-Arreguín, J., Sáenz-Carbonell, L., & Canto-Canché, B. (2020). EffHunter: A tool for prediction of effector protein candidates in fungal proteomic databases. *Biomolecules*, 10(5), 712. doi: 10.3390/biom10050712
- Castañeda-Ojeda, M. P., López-Solanilla, E., & Ramos, C. (2017). Differential modulation of plant immune responses by diverse members of the *Pseudomonas savastanoi* pv. *savastanoi* HopAF type III effector family. *Molecular Plant Pathology*, 18(5), 625-634. doi: 10.1111/mpm.12420
- Catanzariti, A., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A., & Ellis, J. G. (2006). Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *The Plant Cell*, 18(January), 243-256. doi: 10.1105/tpc.105.035980.1

- Cesari, S. (2018). Multiple strategies for pathogen perception by plant immune receptors. *New Phytologist*, 219, 17-24. doi: 10.1111/nph.14877
- Chang, J. H., Rathjen, J. P., Bernal, A. J., Staskawicz, B. J., & Michelmore, R. W. (2000). avrPto enhances growth and necrosis caused by *Pseudomonas syringae* pv. *tomato* in tomato lines lacking either Pto or Prf. *Molecular Plant-Microbe Interactions*, 13(5), 568-571. doi: 10.1094/MPMI.2000.13.5.568
- Chen, L. Q., Hou, B. H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X. Q., . . . Frommer, W. B. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*, 468(7323), 7527-7532. doi: 10.1038/nature09606
- Chen, Z., Kloeck, A. P., Cuzick, A., Moeder, W., Tang, D., Innes, R. W., . . . Kunkel, B. N. (2004). The *Pseudomonas syringae* type III effector AvrRpt2 functions downstream or independently of SA to promote virulence on *Arabidopsis thaliana*. *The Plant Journal*, 37(4), 494-504. doi: 10.1111/j.1365-313x.2003.01984.x
- Christie, P. J., Whitaker, N., & González-Rivera, C. (2014). Mechanism and structure of the bacterial type IV secretion systems. *Biochimica et Biophysica Acta (BBA) – Molecular Cell Research*, 1843(8), 1578-1591. doi: 10.1016/j.bbamcr.2013.12.019
- Clark, C. A., & Lorbeer, J. W. (1975). The role of phenols in *Botrytis* brown stain of onion. *Phytopathology*, 65, 338-341. doi: 10.1094/Phyto-65-338
- Cohn, M., Bart, R. S., Shybut, M., Dahlbeck, D., Gomez, M., Morbitzer, R., . . . Staskawicz, B. J. (2014). *Xanthomonas axonopodis* virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar transporter in cassava. *Molecular Plant-Microbe Interactions*, 27(11), 1186-1198. doi: 10.1094/MPMI-06-14-0161-R
- Collemare, J., O'Connell, R., & Lebrun, M. H. (2019). Nonproteinaceous effectors: the terra incognita of plant–fungal interactions. *New Phytologist*, 223, 590-596. doi: 10.1111/nph.15785
- Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One*, 12(12), e0190152. doi: 10.1371/journal.pone.0190152
- Couto, D., & Zipfel, C. (2016). Regulation of pattern recognition receptor signalling in plants. *Nature Reviews Immunology*, 16, 537-552. doi: 10.1038/nri.2016.77

- Covacci, A., & Rappuoli, R. (1993). Pertussis toxin export requires accessory genes located downstream from the pertussis toxin operon. *Molecular Microbiology*, 8, 429-434. doi: 10.1111/j.1365-2958.1993.tb01587.x
- Cui, F., Wu, S., Sun, W., Coaker, G., Kunkel, B., He, P., & Shan, L. (2013). The *Pseudomonas syringae* type III effector AvrRpt2 promotes pathogen virulence via stimulating *Arabidopsis* auxin/indole acetic acid protein turnover. *Plant Physiology*, 162(2), 1018-1029. doi: 10.1104/pp.113.219659
- Cui, H., Wang, Y., Xue, L., Chu, J., Yan, C., Fu, J., . . . Zhou, J. M. (2010). *Pseudomonas syringae* effector protein AvrB perturbs *Arabidopsis* hormone signaling by activating MAP kinase 4. *Cell Host and Microbe*, 7(2), 164-175. doi: 10.1016/j.chom.2010.01.009
- Cunnac, S., Lindeberg, M., & Collmer, A. (2009). *Pseudomonas syringae* type III secretion system effectors: repertoires in search of functions. *Current Opinion in Microbiology*, 12(1), 53-60. doi: 10.1016/j.mib.2008.12.003
- de Guillen, K., Lorrain, C., Tsan, P., Barthe, P., Petre, B., Saveleva, N., . . . Hecker, A. (2019). Structural genomics applied to the rust fungus *Melampsora larici-populina* reveals two candidate effector proteins adopting cystine knot and NTF2-like protein folds. *Scientific Reports*, 9(1), 18084. doi: 10.1038/s41598-019-53816-9
- De Wit, P. J. G. M., Mehrabi, R., Van Den Burg, H. A., & Stergiopoulos, I. (2009). Fungal effector proteins: past, present and future. *Molecular Plant Pathology*, 10, 735-747. doi: 10.1111/j.1364-3703.2009.00591.x
- Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., . . . Foster, G. D. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*, 13(4), 414-430. doi: 10.1111/j.1364-3703.2011.00783.x
- Deslandes, L., Olivier, J., Peeters, N., Feng, D. X., Khounlotham, M., Boucher, C., . . . Marco, Y. (2003). Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *PNAS*, 100(13), 8024-8029. doi: 10.1073/pnas.1230660100
- Dixon, R. A., Harrison, M. J., & Lamb, C. J. (1994). Early events in the activation of plant defense responses. *Annual Review of Phytopathology*, 32, 451-479. doi: 10.1146/annurev.py.32.090194.002403
- Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J., . . . Kahmann, R. (2011a). Metabolic priming by a secreted fungal effector. *Nature*, 478(7369), 7395-7398. doi: 10.1038/nature10454

- Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J., . . . Kahmann, R. (2011b). Metabolic priming by a secreted fungal effector. *Nature*, 478398. doi: 10.1038/nature10454
- Dodds, P. N., Lawrence, G. J., Catanzariti, A.-M., Teh, T., Wang, C.-I. A., Ayliffe, M. A., . . . Ellis, J. G. (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *PNAS*, 103(23), 8888-8893. doi: 10.1073/pnas.0602577103
- Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development*, 35, 57-65. doi: 10.1016/j.gde.2015.09.001
- dos Santos, K. C. G., Desgagné-Pénix, I., & Germain, H. (2020a). Custom selected reference genes outperform pre-defined reference genes in transcriptomic analysis. *BMC Genomics*, 21, 35. doi: 10.1186/s12864-019-6426-2
- dos Santos, K. C. G., Pelletier, G., Séguin, A., Guillemette, F., Hawkes, J., Desgagné-Pénix, I., & Germain, H. (2020b). Differential alteration of plant functions by homologous fungal candidate effectors. *bioRxiv*, doi: 10.1101/2020.10.30.363010
- Dou, D., Kale, S. D., Wang, X., Jiang, R. H. Y., Bruce, N. A., Arredondo, F. D., . . . Tyle, B. M. (2008). RXLR-mediated entry of *Phytophthora sojae* effector Avr1b into soybean cells does not require pathogen-encoded machinery. *Plant Cell*, 20(7), 1930-1947. doi: 10.1105/tpc.107.056093
- Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., . . . Martin, F. (2011a). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *PNAS*, 108, 9166-9171. doi: 10.1073/pnas.1019315108
- Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., Martin, F., & Kohler, A. (2011b). *Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Molecular Plant-Microbe Interactions*, 24(7), 808-818. doi: 10.1094/MPMI-01-11-0006
- Ellis, J. G., & Dodds, P. N. (2011). Showdown at the RXLR motif: Serious differences of opinion in how effector proteins from filamentous eukaryotic pathogens enter plant cells. *PNAS*, 108(35), 14381-14382. doi: 10.1073/pnas.1111668108
- Elmer, W. H. (2001). The economically important diseases of asparagus in the United States. *Plant Health Progress*, 2doi: 10.1094/PHP-2001-0521-01-RV

- Fan, J., Crooks, C., Creissen, G., Hill, L., Fairhurst, S., Doerner, P., & Lamb, C. (2011). *Pseudomonas* sax genes overcome aliphatic isothiocyanate-mediated non-host resistance in *Arabidopsis*. *Science*, 331(6021), 1185-1188. doi: 10.1126/science.1199707
- Fiorin, G. L., Sánchez-Vallet, A., Thomazella, D. P. d. T., Prado, P. F. V., Nascimento, L. C., Figueira, A. V. d. O., . . . Teixeira, P. J. P. L. (2018). Suppression of plant immunity by fungal chitinase-like effectors. *Current Biology*, 28(18), 3023-3030. doi: 10.1016/j.cub.2018.07.055
- Fritz-Laylin, L. K., Krishnamurthy, N., Tör, M., Sjölander, K. V., & Jones, J. D. G. (2005). Phylogenomic analysis of the Receptor-Like Proteins of rice and *Arabidopsis*. *Plant Physiology*, 138(2), 611-623. doi: 10.1104/pp.104.054452
- Fronzes, R., Christie, P. J., & Waksman, G. (2009). The structural biology of type IV secretion systems. *Nature Reviews Microbiology*, 7, 703-714. doi: 10.1038/nrmicro2218
- Fudal, I., Balesdent, M. H., & Rouxel, T. (2018). Effector biology in fungal pathogens of nonmodel crop plants. *Trends in Plant Science*, 23(9), 753-755. doi: 10.1016/j.tplants.2018.07.002
- Garnica, D. P., Nemri, A., Upadhyaya, N. M., Rathjen, J. P., & Dodds, P. N. (2014). The ins and outs of rust haustoria. *PLoS Pathogens*, 10(9), e1004329. doi: 10.1371/journal.ppat.1004329
- Geng, X., Jin, L., Shimada, M., Kim, M. G., & Mackey, D. (2014). The phytotoxin coronatine is a multifunctional component of the virulence armament of *Pseudomonas syringae*. *Planta*, 240(6), 1149-1165. doi: 10.1007/s00425-014-2151-x
- Germain, H., Joly, D. L., Mireault, C., Plourde, M. B., Letanneur, C., Stewart, D., . . . Séguin, A. (2018). Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Molecular Plant Pathology*, 19(1), 191-200. doi: 10.1111/mpp.12514
- Germain, H., & Seguin, A. (2011). Innate immunity: has poplar made its BED? *New Phytologist*, 189(3), 678-687. doi: 10.1111/j.1469-8137.2010.03544.x
- Gilbert, B. M., & Wolpert, T. J. Characterization of the LOV1-mediated, Victorin-induced, cell-death response with virus-induced gene silencing. *Molecular Plant-Microbe Interactions*, 26(8), 903-917. doi: 10.1094/MPMI-01-13-0014-R

- Green, E. R., & Mecsas, J. (2016). Bacterial secretion systems – An overview. *Microbiology Spectrum*, 4(1), VMBF-0012-2015. doi: 10.1128/microbiolspec.VMBF-0012-2015
- Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., . . . Yin, Z. (2005). R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature*, 435, 1122-1125. doi: 10.1038/nature03630
- Hacquard, S., Joly, D. L., Lin, Y. C., Tisserant, E., Feau, N., Delaruelle, C., . . . Duplessis, S. (2012). A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Molecular Plant-Microbe Interactions*, 25(3), 279-293. doi: 10.1094/MPMI-09-11-0238
- Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., & Duplessis, S. (2011). The poplar-poplar rust interaction: insights from genomics and transcriptomics. *Journal of Pathogens*, 2011, 716041. doi: 10.4061/2011/716041
- Hahn, M. (2000). The rust fungi: cytology, physiology and molecular biology of infection. Dans Kronstad, J. W. *Fungal pathology* (pp.267–306). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Han, Y., Song, L., Peng, C., Liu, X., Liu, L., Zhang, Y., . . . Lu, G.-d. (2019). A *Magnaporthe* chitinase interacts with a rice jacalin-related lectin to promote host colonization. *Plant Physiology*, 179(4), 1416-1430. doi: 10.1104/pp.18.01594
- He, Y., Zhou, J., Shan, L., & Meng, X. (2018). Plant cell surface receptor-mediated signaling - a common theme amid diversity. *Journal of Cell Science*, 131(2), jcs209353. doi: 10.1242/jcs.209353
- Heath, M. C. (1976). Ultrastructural and functional similarity of the haustorial neckband of rust fungi and the Casparian strip of vascular plants. *Canadian Journal of Botany*, 54, 2484-2489. doi: 10.1139/b76-266
- Heath, M. C. (1990). In vitro formation of haustoria of the cowpea rust fungus *Uromyces vignae* in the absence of a living plant cell. II. Electron microscopy. *Canadian Journal of Botany*, 68, 278-287. doi: 10.1139/b90-038
- Hogenhout, S. A., Hoorn, R. A. L. V. d., Terauchi, R., & Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. *Molecular Plant-Microbe Interactions*, 22(2), 115-122. doi: 10.1094/MPMI-22-2-0115

- Hoorn, R. A. L. V. d., Wit, P. J. G. M. D., & Joosten, M. H. A. J. (2002). Balancing selection favors guarding resistance proteins. *Trends in Plant Science*, 7(2), 67-71. doi: 10.1016/S1360-1385(01)02188-4
- Houterman, P. M., Ma, L., Van Ooijen, G., De Vroomen, M. J., Cornelissen, B. J., Takken, F. L., & Rep, M. (2009). The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *The Plant Journal*, 58978. doi: 10.1111/j.1365-313X.2009.03838.x
- Huh, S. U., Cevik, V., Ding, P., Duxbury, Z., Ma, Y., Tomlinson, L., . . . Jones, J. D. G. (2017). Protein-protein interactions in the RPS4/RRS1 immune receptor complex. *PLoS Pathogens*, 13(5), e1006376. doi: 10.1371/journal.ppat.1006376
- Hurley, B., Lee, D., Mott, A., Wilton, M., Liu, J., Liu, Y. C., . . . Desveaux, D. (2014). The *Pseudomonas syringae* type III effector HopF2 suppresses *Arabidopsis* stomatal immunity. *PLoS One*, 9(12), e114921. doi: 10.1371/journal.pone.0114921
- Irving, H. R., Gehring, C. A., & Parish, R. W. (1992). Changes in cytosolic pH and calcium of guard cells precede stomatal movements. *PNAS*, 89(5), 1790-1794. doi: 10.1073/pnas.89.5.1790
- Jamieson, P. A., Shan, L., & He, P. (2018). Plant cell surface molecular cypher: Receptor-like proteins and their roles in immunity and development. *Plant Science*, 274, 242-251. doi: 10.1016/j.plantsci.2018.05.030
- Joly, D. L., Feau, N., Tanguay, P., & Hamelin, R. C. (2010). Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (*Melampsora* spp.). *BMC Genomics*, 11, 422. doi: 10.1186/1471-2164-11-422
- Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 7323-7332. doi: 10.1038/nature05286
- Jones, J. T., Haegeman, A., Danchin, E. G. J., Gaur, H. S., Helder, J., Jones, M. G. K., . . . Perry, R. N. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology*, 14, 946-961. doi: 10.1111/mpp.12057
- Kale, S. D., Gu, B., Capelluto, D. G. S., Dou, D., Feldman, E., Rumore, A., . . . Tyler, B. M. (2010). External lipid PI3P mediates entry of eukaryotic pathogen effectors into plant and animal host cells. *Cell*, 142, 284-295. doi: 10.1016/j.cell.2010.06.008

- Kamoun, S. (2007). Groovy times: Filamentous pathogen effectors revealed. *Current Opinion in Plant Biology*, 10(4), 358-365. doi: 10.1016/j.pbi.2007.04.017
- Kamoun, S., Furzer, O., Jones, J. D. G., Judelson, H. S., Ali, G. S., Dalio, R. J. D., . . . Govers, F. (2015). The top 10 oomycete pathogens in molecular plant pathology. *Molecular Plant Pathology*, 16, 413-434. doi: 10.1111/mpp.12190
- Kang, X., Kirui, A., Widanage, M. C. D., Mentink-Vigier, F., Cosgrove, D. J., & Wang, T. (2019). Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR. *Nature Communications*, 10, 347. doi: 10.1038/s41467-018-08252-0
- Kanja, C., & Hammond-Kosack, K. E. (2020). Proteinaceous effector discovery and characterization in filamentous plant pathogens. *Molecular Plant Pathology*, 21, 1353-1376. doi: 10.1111/mpp.12980
- Khang, C. H., Berruyer, R., Giraldo, M. C., Kankanala, P., Park, S.-Y., Czymmek, K., . . . Valent, B. (2010). Translocation of *Magnaporthe oryzae* effectors into rice cells and their subsequent cell-to-cell movement. *Plant Cell*, 22(4), 1388-1403. doi: 10.1105/tpc.109.069666
- Kim, S., Kim, C. Y., Park, S. Y., Kim, K. T., Jeon, J., Chung, H., . . . Lee, Y. H. (2020). Two nuclear effectors of the rice blast fungus modulate host immunity via transcriptional reprogramming. *Nature Communications*, 11, 5845. doi: 10.1038/s41467-020-19624-w
- Kohler, A., Rinaldi, C., Duplessis, S., Baucher, M., Geelen, D., Duchaussoy, F., . . . Martin, F. (2008). Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Molecular Biology*, 66(6), 619-636. doi: 10.1007/s11103-008-9293-9
- Kroj, T., Chanclud, E., Michel-Romiti, C., Grand, X., & Morel, J. B. (2016). Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytologist*, 210(2), 618-626. doi: 10.1111/nph.13869
- Leisner, S. M., & Schoelz, J. E. (2018). Joining the crowd: Integrating plant virus proteins into the larger world of pathogen effectors. *Annual Review of Phytopathology*, 56, 89-110. doi: 10.1146/annurev-phyto-080417-050151
- Liu, C., Pedersen, C., Schultz-Larsen, T., Aguilar, G. B., Madriz-Ordeñana, K., Hovmøller, M. S., & Thordal-Christensen, H. (2016). The stripe rust fungal effector PEC6 suppresses pattern-triggered immunity in a host species-independent manner and interacts with adenosine kinases. *New Phytologist*, 14034. doi: 10.1111/nph.14034

- Liu, T., Song, T., Zhang, X., Yuan, H., Su, L., Li, W., . . . Dou, D. (2014). Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis. *Nature Communications*, 5, 4686-4695. doi: 10.1038/ncomms5686
- Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., . . . Kahmann, R. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, 66, 513-545. doi: 10.1146/annurev-arplant-043014-114623
- Loix, C., Huybrechts, M., Vangronsveld, J., Gielen, M., Keunen, E., & Cuypers, A. (2017). Reciprocal interactions between cadmium-induced cell wall responses and oxidative stress in plants. *Frontiers in Plant Science*, 8, 1867. doi: 10.3389/fpls.2017.01867
- Lorang, J. M., Sweat, T. A., & Wolpert, T. J. (2007). Plant disease susceptibility conferred by a “resistance” gene. *PNAS*, 104(37), 14861-14866. doi: 10.1073/pnas.0702572104
- Lorrain, C., Goncalves Dos Santos, K. C., Germain, H., Hecker, A., & Duplessis, S. (2019). Advances in understanding obligate biotrophy in rust fungi. *New Phytologist*, 222(3), 1190-1206. doi: 10.1111/nph.15641
- Lorrain, C., Petre, B., & Duplessis, S. (2018a). Show me the way: rust effector targets in heterologous plant systems. *Current Opinion in Microbiology*, 46, 19-25. doi: 10.1016/j.mib.2018.01.016
- Lorrain, C., Petre, B., & Duplessis, S. (2018b). Show me the way: rust effector targets in heterologous plant systems. *Current Opinion in Microbiology*, 46, 19-25. doi: 10.1016/j.mib.2018.01.016
- Luna, E., Pastor, V., Robert, J., Flors, V., Mauch-Mani, B., & Ton, J. (2011). Callose deposition: A multifaceted plant defense response. *Molecular Plant-Microbe Interactions*, 24(2), 183-193. doi: 10.1094/MPMI-07-10-0149
- Ma, L., Djavaheri, M., Wang, H., Larkan, N. J., Haddadi, P., Beynon, E., . . . Borhan, M. H. (2018). *Leptosphaeria maculans* effector protein AvrLm1 modulates plant immunity by enhancing MAP Kinase 9 phosphorylation. *iScience*, 3, 177-191. doi: 10.1016/j.isci.2018.04.015
- Ma, Z., Song, T., Zhu, L., Ye, W., Wang, Y., Shao, Y., . . . Wang, Y. (2015). A *Phytophthora sojae* glycoside hydrolase 12 protein is a major virulence factor during soybean infection and is recognized as a PAMP. *Plant Cell*, 27(7), 2057-2072. doi: 10.1105/tpc.15.00390

- Ma, Z., Zhu, L., Song, T., Wang, Y., Zhang, Q., Xia, Y., . . . Wang, Y. (2017). A paralogous decoy protects *Phytophthora sojae* apoplastic effector PsXEG1 from a host inhibitor. *Science*, 355(6326), 6710-6714. doi: 10.1126/science.aai7919
- Mackey, D., Holt III, B. F., Wiig, A., & Dangl, J. L. (2002). RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell*, 108(6), 743-754. doi: 10.1016/S0092-8674(02)00661-X
- Madina, M. H., Rahman, M. S., Huang, X., Zhang, Y., Zheng, H., & Germain, H. (2020). A poplar rust effector protein associates with protein disulfide isomerase and enhances plant susceptibility. *Biology*, 9(9), 294. doi: 10.3390/biology9090294
- Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., . . . Foster, G. D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, 13, 614-629. doi: 10.1111/j.1364-3703.2012.00804.x
- Marchal, C., Zhang, J., Zhang, P., Fenwick, P., Steuernagel, B., Adamski, N. M., . . . Uauy, C. (2018). BED-domain-containing immune receptors confer diverse resistance spectra to yellow rust. *Nature Plants*, 4(9), 662-668. doi: 10.1038/s41477-018-0236-4
- Marshall, R., Kombrink, A., Motteram, J., Loza-Reyes, E., Lucas, J., Hammond-Kosack, K. E., . . . Rudd, J. J. (2011). Analysis of two in planta expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiology*, 156(2), 756-769. doi: 10.1104/pp.111.176347
- Martínez-Cruz, J., Romero, D., Torre, F. N. d. l., Fernández-Ortuño, D., Torés, J. A., Vicente, A. d., & Pérez-García, A. (2018). The Functional Characterization of *Podosphaera xanthii* Candidate Effector Genes Reveals Novel Target Functions for Fungal Pathogenicity. *Molecular Plant-Microbe Interactions*, 31(9), 914-931. doi: 10.1094/MPMI-12-17-0318-R
- McDonald, M. C., Taranto, A. P., Hill, E., Schwessinger, B., Liu, Z., Simpfendorfer, S., . . . Solomon, P. S. (2019). Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. *mBio*, 10(2), e01515-01519. doi: 10.1128/mbio.e01515-19
- Mejias, J., Truong, N. M., Abad, P., Favery, B., & Quentin, M. (2019). Plant proteins and processes targeted by parasitic nematode effectors. *Frontiers in Plant Science*, 10, 970. doi: 10.3389/fpls.2019.00970

- Mendgen, K. (1975). Ultrastructural demonstration of different peroxidase activities during the bean rust infection process. *Physiological Plant Pathology*, 6(3), 275-278. doi: 10.1016/0048-4059(75)90082-X
- Mentlak, T. A., Kombrink, A., Shinya, T., Ryder, L. S., Otomo, I., Saitoh, H., . . . Talbot, N. J. (2012). Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell*, 24(1), 322-335. doi: 10.1105/tpc.111.092957
- Motteram, J., Kufner, I., Deller, S., Brunner, F., Hammond-Kosack, K. E., Nummerger, T., & Rudd, J. J. (2009). Molecular characterization and functional analysis of MgNLP, the sole NPP1 domain-containing protein, from the fungal wheat leaf pathogen *Mycosphaerella graminicola*. *Molecular Plant-Microbe Interactions*, 22(7), 790-799. doi: 10.1094/MPMI-22-7-0790
- Mucyn, T. S., Clemente, A., Andriotis, V. M., Balmuth, A. L., Oldroyd, G. E., Staskawicz, B. J., & Rathjen, J. P. (2006). The tomato NBARC-LRR protein Prf interacts with Pto kinase in vivo to regulate specific plant immunity. *Plant Cell*, 18(10), 2792-2806. doi: 10.1105/tpc.106.044016
- Mugford, S. T., Barclay, E., Drurey, C., Findlay, K. C., & Hogenhout, S. A. (2016). An immuno-suppressive aphid saliva protein is delivered into the cytosol of plant mesophyll cells during feeding. *Molecular Plant-Microbe Interactions*, 29(11), 854-861. doi: 10.1094/MPMI-08-16-0168-R
- Mukhtar, M. S., Carvunis, A.-R., Dreze, M., Eppele, P., Steinbrenner, J., Moore, J., . . . Dangl, J. L. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science*, 333(6042), 6596-6601. doi: 10.1126/science.1203659
- Narusaka, M., Shirasu, K., Noutoshi, Y., Kubo, Y., Shiraishi, T., Iwabuchi, M., & Narusaka, Y. (2009). RRS1 and RPS4 provide a dual resistance-gene system against fungal and bacterial pathogens. *The Plant Journal*, 60(2), 218-226. doi: 10.1111/j.1365-3113X.2009.03949.x
- Nelson, M. S., Chun, C. L., & Sadowsky, M. J. (2017). Type IV effector proteins involved in the *Medicago-Sinorhizobium* symbiosis. *Molecular Plant-Microbe Interactions*, 30(1), 28-34. doi: 10.1094/MPMI-10-16-0211-R
- Ngou, B. P. M., Ahn, H.-K., Ding, P., & Jones, J. D. (2020). Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *bioRxiv*, doi: 10.1101/2020.04.10.034173

- Nicaise, V., Roux, M., & Zipfel, C. (2009). Recent advances in PAMP-triggered immunity against bacteria: pattern recognition receptors watch over and raise the alarm. *Plant Physiology*, 150(4), 1638-1647. doi: 10.1104/pp.109.139709
- Ninio, S., & Roy, C. R. (2007). Effector proteins translocated by *Legionella pneumophila*: strength in numbers. *Trends in Microbiology*, 15(8), 372-380. doi: 10.1016/j.tim.2007.06.006
- Nissan, G., Manulis-Sasson, S., Weinthal, D., Mor, H., Sessa, G., & Barash, I. (2006). The type III effectors HsvG and HsvB of gall-forming *Pantoea agglomerans* determine host specificity and function as transcriptional activators. *Molecular Microbiology*, 61, 1118-1131. doi: 10.1111/j.1365-2958.2006.05301.x
- Nurnberger, T., & Kemmerling, B. (2009). Pathogen-associated molecular patterns (PAMP) and PAMP-triggered immunity. Dans Parker, J. *Molecular aspects of plant disease resistance* New Dehli, India: Wiley-Blackwell.
- Oh, S.-K., Young, C., Lee, M., Oliva, R., Bozkurt, T. O., Cano, L. M., . . . Kamoun, S. (2009). *In planta* expression screens of *Phytophthora infestans* RXLR effectors reveal diverse phenotypes, including activation of the *Solanum bulbocastanum* disease resistance protein Rpi-blb2. *Plant Cell*, 21(9), 2928-2947. doi: 10.1105/tpc.109.068247
- Oliveira, C. M., Auad, A. M., Mendes, S. M., & Frizzas, M. R. (2014). Crop losses and the economic impact of insect pests on Brazilian agriculture. *Crop Protection*, 56, 50-54. doi: 10.1016/j.cropro.2013.10.022
- Ono, Y. (2002). Life cycle and nuclear behavior in three rust fungi (Uredinales). *Mycoscience*, 43, 37-45. doi: 10.1007/s102670200007
- Parmelee, J. A., & Malloch, D. (1972). *Puccinia hystrium* on *Tragopogon*: A new north american rust record. *Mycologia*, 64(4), 922-924. doi: 10.1080/00275514.1972.12019345
- Pedersen, C., Ver Loren van Themaat, E., McGuffin, L. J., Abbott, J. C., Burgis, T. A., Barton, G., . . . Spanu, P. D. (2012). Structure and evolution of barley powdery mildew effector candidates. *BMC Genomics*, 13, 694. doi: 10.1186/1471-2164-13-694
- Perrot-Rechenmann, C. (2010). Cellular responses to auxin: division versus expansion. *Cold Spring Harbour Perspectives in Biology*, 2(5), a001446. doi: 10.1101/cshperspect.a001446

- Petre, B., Joly, D. L., & Duplessis, S. (2014). Effector proteins of rust fungi. *Frontiers in Plant Science*, 5(7), 416. doi: 10.3389/fpls.2014.00416
- Petre, B., Saunders, D. G. O., Sklenar, J., Lorrain, C., Krasileva, K. V., Win, J., . . . Kamoun, S. (2016). Heterologous expression screens in *Nicotiana benthamiana* identify a candidate effector of the wheat yellow rust pathogen that associates with processing bodies. *PLoS One*, 11(2), e0149035. doi: 10.1371/journal.pone.0149035
- Petre, B., Saunders, D. G. O., Sklenar, J., Lorrain, C., Win, J., Duplessis, S., & Kamoun, S. (2015). Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. *Molecular Plant-Microbe Interactions*, 28(6), 689-700. doi: 10.1094/MPMI-01-15-0003-R
- Pretorius, Z. A., Singh, R. P., Wagoire, W. W., & Payne, T. S. (2000). Detection of virulence to wheat stem rust resistance gene Sr31 in *Puccinia graminis* f. sp. *tritici* in Uganda. *Plant Disease*, 84(2), 203. doi: 10.1094/PDIS.2000.84.2.203B
- Quinn, T. P., Crowley, T. M., & Richardson, M. F. (2018). Benchmarking differential expression analysis tools for RNA-Seq: normalization-based vs. log-ratio transformation-based methods. *BMC Bioinformatics*, 19(1), 274. doi: 10.1186/s12859-018-2261-8
- Ray, S. K., Macoy, D. M., Kim, W. Y., Lee, S. Y., & Kim, M. G. (2019). Role of RIN4 in regulating PAMP-Ttriggered immunity and effector-triggered immunity: Current status and future perspectives. *Molecules and Cells*, 42(7), 503-511. doi: 10.14348/molcells.2019.2433
- Redkar, A., Hoser, R., Schilling, L., Zechmann, B., Krzymowska, M., Walbot, V., & Doehlemann, G. (2015). A secreted effector protein of *Ustilago maydis* guides maize leaf cells to form tumors. *Plant Cell*, 27(4), 1332-1351. doi: 10.1105/tpc.114.131086
- Remigi, P., Anisimova, M., Guidot, A., Genin, S., & Peeters, N. (2011). Functional diversification of the GALA type III effector family contributes to *Ralstonia solanacearum* adaptation on different plant hosts. *New Phytologist*, 192, 976-987. doi: 10.1111/j.1469-8137.2011.03854.x
- Rep, M., Meijer, M., Houterman, P. M., van der Does, H. C., & Cornelissen, B. J. C. (2005). *Fusarium oxysporum* evades I-3-mediated resistance without altering the matching avirulence gene. *Molecular Plant-Microbe Interactions*, 18(1), 15-23. doi: 10.1094/MPMI-18-0015

- Robin, G. P., Kleemann, J., Neumann, U., Cabre, L., Dallery, J. F., Lapalu, N., & O'Connell, R. J. (2018). Subcellular localization screening of *Colletotrichum higginsianum* effector candidates identifies fungal proteins targeted to plant peroxisomes, Golgi bodies, and microtubules. *Frontiers in Plant Science*, 9, 562. doi: 10.3389/fpls.2018.00562
- Rocafort, M., Fudal, I., & Mesarich, C. H. (2020). Apoplastic effector proteins of plant-associated fungi and oomycetes. *Current Opinion in Plant Biology*, 56, 9-19. doi: 10.1016/j.pbi.2020.02.004
- Rodriguez, P. A., & Bos, J. I. B. (2013). Toward understanding the role of aphid effectors in plant infestation. *Molecular Plant-Microbe Interactions*, 26(1), 25-30. doi: 10.1094/MPMI-05-12-0119-FI
- Rojas-Estevez, P., Urbina-Gómez, D. A., Ayala-Usma, D. A., Guayazan-Palacios, N., Mideros, M. F., Bernal, A. J., . . . Restrepo, S. (2020). Effector repertoire of *Phytophthora betacei*: In search of possible virulence factors responsible for its host specificity. *Frontiers in Genetics*, 11, 579. doi: 10.3389/fgene.2020.00579
- Rose, J. K. C., Ham, K.-S., Darvill, A. G., & Albersheim, P. (2002). Molecular cloning and characterization of glucanase inhibitor proteins. *Plant Cell*, 14(6), 1329-1345. doi: 10.1105/tpc.002253
- Salama, N. K. G. (2009). *Regulation of a biennial host plant population by an autoecious, demicyclic rust fungus: Puccinia hysterium on Tragopogon pratensis in the Park Grass Experiment* (Thèse de Dissertation inédite). University of London, Ascot
- Sanchez-Vallet, A. (2018). The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens. *Annual Review of Phytopathology*, 56, 21-40. doi: 10.1146/annurev-phyto-080516-035303
- Sarris, P. F., Cevik, V., Dagdas, G., Jones, J. D. G., & Krasileva, K. V. (2016). Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biology*, 14, 8. doi: 10.1186/s12915-016-0228-7
- Sarris, P. F., Duxbury, Z., Huh, S. U., Ma, Y., Segonzac, C., Sklenar, J., . . . Jones, J. D. G. (2015). A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell*, 161(5), 1089-1100. doi: 10.1016/j.cell.2015.04.024
- Saunders, D. G. O., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S., & Raffaele, S. (2012). Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS One*, 7(1), e29847. doi: 10.1371/journal.pone.0029847

- Scholthof, K.-B. G., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., . . . Foster, G. D. (2011). Top 10 plant viruses in molecular plant pathology. *Molecular Plant Pathology*, 12, 938-954. doi: 10.1111/j.1364-3703.2011.00752.x
- Schornack, S., Damme, M. v., Bozkurt, T. O., Cano, L. M., Smoker, M., Thines, M., . . . Huitema, E. (2010). Ancient class of translocated oomycete effectors targets the host nucleus. *PNAS*, 107(40), 17421-17426. doi: 10.1073/pnas.1008491107
- Segal, E. D., Cha, J., Lo, J., Falkow, S., & Tompkins, L. S. (1999). Altered states: Involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *PNAS*, 96(25), 14559-14564. doi: 10.1073/pnas.96.25.14559
- Singh, R. P., Hodson, D. P., Huerta-Espino, J., Jin, Y., Bhavani, S., Njau, P., . . . Govindan, V. (2011). The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annual Review of Phytopathology*, 49, 465-481. doi: 10.1146/annurev-phyto-072910-095423
- Singh, R. P., Hodson, D. P., Huerta-Espino, J., Jin, Y., Njau, P., Wanyera, R., . . . Ward, R. W. (2008). Will stem rust destroy the world's wheat crop? *Advances in Agronomy*, 98, 271-309. doi: 10.1016/S0065-2113(08)00205-8
- Snelders, N. C., Rovenich, H., Petti, G. C., Rocafort, M., Vorholt, J. A., Mesters, J. R., . . . Thomma, B. P. H. J. (2020). A plant pathogen utilizes effector proteins for microbiome manipulation [PREPRINT]. *bioRxiv*, doi: 10.1101/2020.01.30.926725
- Stam, R., Jupe, J., Howden, A. J. M., Morris, J. A., Boevink, P. C., Hedley, P. E., & Huitema, E. (2013). Identification and characterisation CRN effectors in *Phytophthora capsici* shows modularity and functional diversity. *PLoS One*, 8(3), e59517. doi: 10.1371/journal.pone.0059517
- Stassen, J. H. M., & Van den Ackerveken, G. (2011). How do oomycete effectors interfere with plant life? *Current Opinion in Plant Biology*, 14(4), 407-414. doi: 10.1016/j.pbi.2011.05.002
- Stergiopoulos, I., & de Wit, P. J. G. M. (2009). Fungal effector proteins. *Annual Review of Phytopathology*, 47, 233-263. doi: 10.1146/annurev.phyto.112408.132637
- Strange, R. N. (1993). *Plant disease control: Towards environmentally acceptable methods* Boston, MA, USA: Springer.

- Strange, R. N., & Scott, P. R. (2005). Plant disease: A threat to global food security. *Annual Review of Phytopathology*, 43, 83-116. doi: 10.1146/annurev.phyto.43.113004.133839
- Sun, F., Kale, S. D., Azurmendi, H. F., Li, D., Tyler, B. M., & Capelluto, D. G. S. (2013). Structural basis for interactions of the *Phytophthora sojae* RxLR effector Avh5 with phosphatidylinositol 3-phosphate and for host cell entry *Molecular Plant-Microbe Interactions*, 26(3), 330-344. doi: 10.1094/MPMI-07-12-0184-R
- Szurek, B., Rossier, O., Hause, G., & Bonas, U. (2002). Type III-dependent translocation of the *Xanthomonas* AvrBs3 protein into the plant cell. *Molecular Microbiology*, 46, 13-23. doi: 10.1046/j.1365-2958.2002.03139.x
- Tabima, J. F., & Grünwald, N. J. (2019). effectR: an expandable r package to predict candidate RxLR and CRN effectors in oomycetes using motif searches. *Molecular Plant-Microbe Interactions*, 32(9), 1067-1076. doi: 10.1094/MPMI-10-18-0279-TA
- Tanaka, S., Brefort, T., Neidig, N., Djamei, A., Kahnt, J., Vermerris, W., . . . Kahmann, R. (2014). A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. *eLife*, 3, e01355. doi: 10.7554/eLife.01355
- Tanaka, S., Djamei, A., Presti, L. L., Schipper, K., Winterberg, S., Amati, S., . . . Kahmann, R. (2015). Experimental approaches to investigate effector translocation into host cells in the *Ustilago maydis*/maize pathosystem. *Experimental Journal of Cell Biology*, 94(7-9), 349-358. doi: 10.1016/j.ejcb.2015.06.007
- Tanaka, S., Schweizer, G., Rössel, N., Thines, M., & Kahmann, R. (2019). Neofunctionalization of the secreted Tin2 effector in the fungal pathogen *Ustilago maydis*. *Nature microbiology*, 4(2019), 2251-2257. doi: 10.1038/s41564-018-0304-6
- Tang, L., Yang, G., Ma, M., Liu, X., Li, B., Xie, J., . . . Cheng, J. (2020). An effector of a necrotrophic fungal pathogen targets the calcium-sensing receptor in chloroplasts to inhibit host resistance. *Molecular Plant Pathology*, 21(5), 686-701. doi: 10.1111/mpp.12922
- Tian, H., MacKenzie, C. I., Rodriguez-Moreno, L., van den Berg, G. C. M., Chen, H., Rudd, J. J., . . . Thomma, B. P. H. J. (2020). Three LysM effectors of *Zymoseptoria tritici* collectively disarm chitin-triggered plant immunity [PREPRINT]. *bioRxiv*, doi: 10.1101/2020.06.24.169789

- Tounekti, T., Hernández, I., & Munné-Bosch, S. (2013). Salicylic acid biosynthesis and role in modulating terpenoid and flavonoid metabolism in plant responses to abiotic stress. Dans Hayat, S., Ahmad, A. , & Alyemeni, M. *Salicylic acid* (pp. 141-162). Dordrecht: Springer.
- Tsuda, K., & Katagiri, F. (2010). Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current Opinion in Plant Biology*, 13, 459-465. doi: 10.1016/j.pbi.2010.04.006
- Tsuda, K., Mine, A., Bethke, G., Igarashi, D., Botanga, C. J., Tsuda, Y., . . . Katagiri, F. (2013). Dual regulation of gene expression mediated by extended MAPK activation and salicylic acid contributes to robust innate immunity in *Arabidopsis thaliana*. *PLoS Genetics*, 9(12), e1004015. doi: 10.1371/journal.pgen.1004015
- Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J., & Katagiri, F. (2009). Network properties of robust immunity in plants. *PLoS Genetics*, 5(12), 16. doi: 10.1371/journal.pgen.1000772
- Tyler, B. M. (2017). The fog of war: How network buffering protects plants' defense secrets from pathogens. *PLoS Genetics*, 13(5), e1006713. doi: 10.1371/journal.pgen.1006713
- Tyler, B. M., & Rouxel, T. (2013). Effectors of fungi and oomycetes: Their virulence and avirulence functions and translocation from pathogen to host cells. Dans Sessa, G. *Molecular plant immunity* (pp.123-167). New Delhi, India: John Wiley & Sons.
- van der Hooft, R., & Kamoun, S. (2008). From guard to decoy: A new model for perception of plant pathogen effectors. *Plant Cell*, 20(8), 2009-2017. doi: 10.1105/tpc.108.060194
- van der Hoorn, R. A. L., & Kamoun, S. (2008). From guard to decoy: a new model for perception of plant pathogen effectors. *Plant Cell*, 20(8), 2009-2017. doi: 10.1105/tpc.108.060194
- Van Esse, H. P., Bolton, M. D., Stergiopoulos, I., de Wit, P. J. G. M., & Thomma, B. P. H. J. (2007). The chitin-binding *Cladosporium fulvum* effector protein Avr4 is a virulence factor. *Molecular Plant-Microbe Interactions*, 20(9), 1092-1101. doi: 10.1094/MPMI-20-9-1092
- VanEtten, H. D., Mansfield, J. W., Bailey, J. A., & Farmer, E. E. (1994). Two Classes of Plant Antibiotics: Phytoalexins versus "Phytoanticipins". *Plant Cell*, 6, 1191-1192. doi: 10.1105/tpc.6.9.1191

- Vargas, W. A., Sanz-Martin, J. M., Rech, G. E., Armijos-Jaramillo, V. D., Rivera, L. P., Echeverria, M. M., . . . Sukno, S. A. (2016). A fungal effector with host nuclear localization and DNA-binding properties is required for maize anthracnose development. *Molecular Plant-Microbe Interactions*, 29(2), 83-95. doi: 10.1094/MPMI-09-15-0209-R
- Villajuana-Bonequi, M., Matei, A., Ernst, C., Hallab, A., Usadel, B., & Doehlemann, G. (2019). Cell type specific transcriptional reprogramming of maize leaves during *Ustilago maydis* induced tumor formation. *Scientific Reports*, 9(1), 10227. doi: 10.1038/s41598-019-46734-3
- Volk, H., Marton, K., Flajšman, M., Radišek, S., Tian, H., Hein, I., . . . Berne, S. (2019). Chitin-binding protein of *Verticillium nonalfalfae* disguises fungus from plant chitinases and suppresses chitin-triggered host immunity. *Molecular Plant-Microbe Interactions*, 32(10), 1378-1390. doi: 10.1094/MPMI-03-19-0079-R
- Wang, M., Weiberg, A., & Jin, H. (2015). Pathogen small RNAs: a new class of effectors for pathogen attacks. *Molecular Plant Pathology*, 2(3), 218-223. doi: 10.1111/mpp.12233
- Wang, X., Jiang, N., Liu, J., Wende Liu, & Wang, G.-L. (2014). The role of effectors and host immunity in plant–necrotrophic fungal interactions. *Virulence*, 5(7), 722-732. doi: 10.4161/viru.29798
- Wang, Y., Li, J., Hou, S., Wang, X., Li, Y., Ren, D., . . . Zhou, J. M. (2010). A *Pseudomonas syringae* ADP-ribosyltransferase inhibits *Arabidopsis* mitogen-activated protein kinase kinases. *Plant Cell*, 22(6), 2033-2044. doi: 10.1105/tpc.110.075697
- Wawra, S., Agacan, M., Boddey, J. A., Davidson, I., Gachon, C. M., Zanda, M., . . . van West, P. (2012a). Avirulence protein 3a (AVR3a) from the potato pathogen *Phytophthora infestans* forms homodimers through its predicted translocation region and does not specifically bind phospholipids. *Journal of Biological Chemistry*, 287(45), 38101-38109. doi: 10.1074/jbc.M112.395129
- Wawra, S., Belmonte, R., Löbach, L., Saraiva, M., Willems, A., & West, P. (2012b). Secretion, delivery and function of oomycete effector proteins. *Current Opinion in Microbiology*, 15(6), 685-691. doi: 10.1016/j.mib.2012.10.008
- Wiethölter, N., Horn, S., Reisinger, K., Beike, U., & Moerschbacher, B. M. (2003). In vitro differentiation of haustorial mother cells of the wheat stem rust fungus, *Puccinia graminis* f. sp. *tritici*, triggered by the synergistic action of chemical and physical signals. *Fungal Genetics and Biology*, 38(3), 320-326. doi: 10.1016/S1087-1845(02)00539-X

- Wilton, M., Subramaniam, R., Elmore, J., Felsensteiner, C., Coaker, G., & Desveaux, D. (2010). The type III effector HopF2Pto targets *Arabidopsis* RIN4 protein to promote *Pseudomonas syringae* virulence. *PNAS*, *107*(5), 2349-2354. doi: 10.1073/pnas.0904739107
- Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, D. G. O., Yoshida, K., Dong, S., . . . Kamoun, S. (2012). Effector Biology of Plant-Associated Organisms: Concepts and Perspectives. *Cold Spring Harbor Symposia on Quantitative Biology*, *77*, 235-247. doi: 10.1101/sqb.2012.77.015933
- Wu, D., von Roepenack-Lahaye, E., Buntru, M., de Lange, O., Schandry, N., Perez-Quintero, A. L., . . . Lahaye, T. (2019). A plant pathogen type III effector protein subverts translational regulation to boost host polyamine levels. *Cell Host & Microbe*, *26*(5), 638-649. doi: 10.1016/j.chom.2019.09.014
- Wu, S., Lu, D., Kabbage, M., Wei, H. L., Swingle, B., Records, A. R., . . . Shan, L. (2011). Bacterial effector HopF2 suppresses *Arabidopsis* innate immunity at the plasma membrane. *Molecular Plant-Microbe Interactions*, *24*(5), 585-593. doi: 10.1094/MPMI-07-10-0150
- Xiang, T., Zong, N., Zou, Y., Wu, Y., Zhang, J., Xing, W., . . . Zhou, J.-M. (2008). *Pseudomonas syringae* effector AvrPto blocks innate immunity by targeting receptor kinases. *Current Biology*, *18*(1), 74-80. doi: 10.1016/j.cub.2007.12.020
- Xu, Q., Tang, C., Wang, X., Sun, S., Zhao, J., Kang, Z., & Wang, X. (2019). An effector protein of the wheat stripe rust fungus targets chloroplasts and suppresses chloroplast function. *Nature Communications*, *10*, 5571. doi: 10.1038/s41467-019-13487-6
- Yaeno, T., Li, H., Chaparro-Garcia, A., Schornack, S., Koshiba, S., Watanabe, S., . . . 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. *PNAS*, *108*(35), 14682-14687. doi: 10.1073/pnas.1106002108
- Yamaguchi, Y., Pearce, G., & Ryan, C. A. (2006). The cell surface leucine-rich repeat receptor for AtPep1, an endogenous peptide elicitor in *Arabidopsis*, is functional in transgenic tobacco cells. *PNAS*, *103*(26), 10104-10109. doi: 10.1073/pnas.0603729103
- Yang, B., Sugio, A., & White, F. F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *PNAS*, *103*(27), 10503-10508. doi: 10.1073/pnas.0604088103

- Yeats, T. H., & Rose, J. K. C. (2013). The formation and function of plant cuticles. *Plant Physiology*, 163(1), 5-20. doi: 10.1104/pp.113.222737
- Yuan, M., Jiang, Z., Bi, G., Nomura, K., Liu, M., He, S. Y., . . . Xin, X.-F. (2020). Pattern-recognition receptors are required for NLR-mediated plant immunity. *bioRxiv*, doi: 10.1101/2020.04.10.031294
- Zeng, L., Zhang, N., Zhang, Q., Endress, P. K., Huang, J., & Ma, H. (2017). Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytologist*, 214(3), 1338-1354. doi: 10.1111/nph.14503
- Zhang, W., Zhao, F., Jiang, L., Chen, C., Wu, L., & Liu, Z. (2018). Different pathogen defense strategies in *Arabidopsis*: More than pathogen recognition. *Cells*, 7(12), 252. doi: 10.3390/cells7120252
- Zhao, T., Rui, L., Li, J., Nishimura, M. T., Vogel, J. P., Liu, N., . . . Tang, D. (2015). A truncated NLR protein, TIR-NBS2, is required for activated defense responses in the *exo70B1* mutant. *PLoS Genet*, 11(1), e1004945. doi: 10.1371/journal.pgen.1004945
- Zhu, J., Oger, P. M., Schrammeijer, B., Hooykaas, P. J. J., Farrand, S. K., & Winans, S. C. (2000). The Bases of crown gall tumorigenesis. *Journal of Bacteriology*, 182(14), 3885-3895. doi: 10.1128/JB.182.14.3885-3895.2000
- Zhu, X. Y., Chase, M. W., Qiu, Y. L., Kong, H. Z., Dilcher, D. L., Li, J. H., & Chen, Z. D. (2007). Mitochondrial *matR* sequences help to resolve deep phylogenetic relationships in rosids. *BMC Evolutionary Biology*, 7, 217. doi: 10.1186/1471-2148-7-217
- Zong, N., Xiang, T., Zou, Y., Chai, J., & Zhou, J. M. (2008). Blocking and triggering of plant immunity by *Pseudomonas syringae* effector AvrPto. *Plant Signaling & Behavior*, 3(8), 583-585. doi: 10.4161/psb.3.8.5741
- Zuppini, A., Navazio, L., Sella, L., Castiglioni, C., Favaron, F., & Mariani, P. (2005). An endopolygalacturonase from *Sclerotinia sclerotiorum* induces calcium-mediated signaling and programmed cell death in soybean cells. *Molecular Plant-Microbe Interactions*, 18(8), 849-855. doi: 10.1094/MPMI-18-0849

ANNEX A

ADVANCES IN UNDERSTANDING OBLIGATE BIOTROPHY IN RUST FUNGI

Cécile Lorrain, Karen Cristine Gonçalves dos Santos, Hugo Germain,
Arnaud Hecker, Sébastien Duplessis

New Phytologist (2019), 222: 1190-1206. doi:10.1111/nph.15641

*Tansley review*

Advances in understanding obligate biotrophy in rust fungi

Author for correspondence:






Sébastien Duplessis

Tel: +33 383 39 40 13

Email: sebastien.duplessis@inra.fr

Received: 21 July 2018

Accepted: 13 November 2018

Cécile Lorrain¹ , Karen Cristine Gonçalves dos Santos² , Hugo Germain² ,
Arnaud Hecker³  and Sébastien Duplessis¹ ¹INRA Centre Grand Est - Nancy, UMR 1136 INRA/Université de Lorraine Interactions Arbres/Microorganismes, Champenoux54280, France; ²Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, Trois-Rivières, QC G9A5H7, Canada; ³Université de Lorraine, UMR 1136 Université de Lorraine/INRA Interactions Arbres/Microorganismes, Vandœuvre-

lès-Nancy, France

Contents

Summary	1190	V. Rusts in the genomics era: the ever-expanding list of candidate effector genes	1195
I. Introduction	1190	VI. Functional characterization of rust effectors	1197
II. Rust fungi: a diverse and serious threat to agriculture	1191	VII. Putting rusts to sleep: Pucciniales research outlooks	1201
III. The different facets of rust life cycles and unresolved questions about their evolution	1191	Acknowledgements	1202
IV. The biology of rust infection	1192	References	1202

Summary

New Phytologist (2019) 222: 1190–1206
doi: 10.1111/nph.15641**Key words:** effectoromic screens, fungal genomics, heterologous systems, pathogenesis, Pucciniales, rust effectors, virulence.

Rust fungi (Pucciniales) are the largest group of plant pathogens and represent one of the most devastating threats to agricultural crops worldwide. Despite the economic importance of these highly specialized pathogens, many aspects of their biology remain obscure, largely because rust fungi are obligate biotrophs. The rise of genomics and advances in high-throughput sequencing technology have presented new options for identifying candidate effector genes involved in pathogenicity mechanisms of rust fungi. Transcriptome analysis and integrated bioinformatics tools have led to the identification of key genetic determinants of host susceptibility to infection by rusts. Thousands of genes encoding secreted proteins highly expressed during host infection have been reported for different rust species, which represents significant potential towards understanding rust effector function. Recent high-throughput *in planta* expression screen approaches (effectoromics) have pushed the field ahead even further towards predicting high-priority effectors and identifying avirulence genes. These new insights into rust effector biology promise to inform future research and spur the development of effective and sustainable strategies for managing rust diseases.

I. Introduction

Obligate biotrophy is widespread across plant pathogens. Plant pathogenic fungal biotrophs extend to powdery mildews, smuts and rusts and share common features with biotrophic oomycetes (Kemen & Jones, 2012). As obligate biotrophs, rust fungi can only feed, grow and reproduce on their living host(s) and they differentiate specific infection structures called

haustoria that are necessary to establish intimate interactions inside infected host tissues (Voegelé & Mendgen, 2011; Kemen *et al.*, 2015; Dracatos *et al.*, 2018). Rust fungi are presumed to have coevolved with their host plants, with which they exhibit a high specificity of interaction (Aime *et al.*, 2017). Absolute dependence on the host essentially means that rust fungi cannot be cultured on artificial media, which significantly complicates their manipulation under laboratory

conditions. To date, the complete life cycle of rust fungi has never been achieved outside their natural hosts and habitats.

Understanding the basis of obligate biotrophy and how evolution has shaped this particular trophic mode remains unresolved. Recent advances in genomics have provided some important cues about missing functions or over- and under-represented gene categories that may support biotrophic features (Spanu, 2012). However, comparison of general genomic features such as genome size, repeat content or the number of predicted genes has failed to reveal a single evolutionary trajectory towards this trophic mode, instead supporting convergent evolution (Kemen *et al.*, 2015). A common feature is the presence of a large secretome with thousands of predicted genes encoding secreted proteins in rust fungi (Lo Presti *et al.*, 2015; Aime *et al.*, 2017). Considering the current models in molecular phytopathology (Jones & Dangl, 2006; Wu *et al.*, 2018), these secreted proteins probably contain effectors that play a key role in establishing a successful infection. The identification of thousands of putative rust fungal effector candidates has pushed the field forwards towards understanding their possible role in manipulating host immunity and physiology (Figueroa *et al.*, 2016).

We present here an overview of recent advances that have facilitated dissection of the molecular mechanisms underlying pathogenesis of rust fungi as biotrophic plant pathogens. We first report on the peculiar life cycle of rust fungi and the current knowledge of the biology of infection. We then illustrate how the advent of functional genomics has played a pivotal role in advancing our knowledge of processes at play during plant infection by rust fungi. Indeed, in the past 5 years alone, functional studies of rust fungi have accelerated through the systematic application of high-throughput effector screens for predicting high-priority effector candidates and identifying avirulence genes. Our overarching goal here is to provide an overview of recent advances and current progress, which ultimately may help to identify upcoming challenges in rust pathogen research.

II. Rust fungi: a diverse and serious threat to agriculture

The Pucciniales (rust fungi) represent one of the largest fungal orders, with > 8000 species already described (Aime *et al.*, 2014). These intriguing parasites have the ability to infect a wide variety of host plants, from ferns to monocots and gymnosperms to angiosperms, which suggests an ancestral adaptation to the biotrophic lifestyle (Aime *et al.*, 2014). Worldwide, rust fungi are considered among the most serious threats to both agricultural crops (e.g. wheat (*Triticum aestivum*), soybean (*Glycine max*) or coffee), and tree species used for wood production and bioenergy (e.g. poplar, eucalypt or pines) (Fig. 1; Dean *et al.*, 2012; Fischer *et al.*, 2012; Snieszko *et al.*, 2012). To fend off rust diseases, the plant breeding industry primarily prioritized developing disease-resistant plants. Over the years, rust fungi have adapted to completely overcome plant resistances (Ellis *et al.*, 2014). Among wheat rust fungi, the stem rust fungus *Puccinia graminis* f. sp. *tritici* has devastated wheat crops throughout the history of agriculture, and it is only in the past decades that it has been successfully contained in

most production areas around the world (Ellis *et al.*, 2014; Figueroa *et al.*, 2017). In 1998, the emergence in Uganda of the isolate Ug99 of the wheat stem rust fungus, virulent against major wheat resistance genes, has put wheat cultivation at risk (Singh *et al.*, 2011, 2015).

The recent literature provides examples of loss estimates directly attributed to rust fungi. Overall, annual global losses in wheat production associated with rust fungi are estimated at US \$4–5 billion (Figueroa *et al.*, 2017). The wheat yellow rust disease caused by *Puccinia striiformis* f. sp. *tritici* has led to a decrease in world wheat production with an estimate market value of US\$979 million annually (Chen *et al.*, 2014; Beddow *et al.*, 2015). The wheat leaf rust *Puccinia triticina* is the most common and widely distributed wheat rust fungus and can lead to significant damage (Bolton *et al.*, 2008). Since the 1990s, the Asian soybean rust fungus *Phakopsora pachyrhizi* has spread through outbreaks in South America, and combined costs of high yield loss and intensified use of fungicides were estimated at US \$2 billion annually (Goellner *et al.*, 2010), with losses in Brazil reaching just over US\$2 billion annually between 2003 and 2014 (Godoy *et al.*, 2016). Another remarkable rust pathogen with a high impact on agriculture is the coffee leaf rust causal agent *Hemileia vastatrix*. This fungus has caused important coffee yield losses across Africa, Asia and South America with an estimated market value of US\$2–3 billion annually (Talhinhas *et al.*, 2017). Rust fungi are subjected to rigorous surveillance and monitoring against their potentially devastating effects on global crop production (Singh *et al.*, 2011; Bueno-Sancho *et al.*, 2017).

III. The different facets of rust life cycles and unresolved questions about their evolution

Rust fungi are host-specific pathogens with extremely complex life cycles. A single species can require two unrelated host plants to complete its life cycle; other rusts have less complex life cycles, missing one or more of the spore-producing stages (Aime *et al.*, 2017). Many rust fungi display life cycles that involve five different spore types or stages: pycniospores, aeciospores, urediniospores, teliospores and basidiospores (Fig. 2). Some rust fungi can be both macrocyclic and autoecious in that they produce the five spore stages on a unique host (e.g. the flax rust fungus *Melampsora lini*; Lawrence *et al.*, 2007). Macrocyclic rusts are those with five spore stages that are produced sequentially. Microcyclic refers to rusts with the shortest life cycles composed of only two spore types: basidiospores and teliospores, or pycniospores and teliospores (Fig. 2). Demicyclic rust fungi lack the urediniospore stage and can be either autoecious or heteroecious (Fig. 2). Hemicyclic rust fungi exhibit only teliospores and urediniospores and are autoecious (Aime *et al.*, 2017; Fig. 2). For some rust fungi, the complete life cycles have never been observed, as is the case for the alternate hosts for two economically important species such as the soybean rust fungus *P. pachyrhizi* and the coffee rust *H. vastatrix* (Slaminko *et al.*, 2008; Talhinhas *et al.*, 2017).

In a typical macrocyclic-heteroecious life cycle, meiosis takes place in short-lived basidia produced by germinating teliospores. Haploid basidiospores infect the aecial host within which they



Fig. 1 Rust disease symptoms in wheat and poplar. Illustration of disease symptoms for two rust fungi on their respective aecial host plants: (a–c) *Puccinia triticina* on wheat; (d–f) *Melampsora larici-populina* on poplar. (a) Orange uredinia pustules formed by *P. triticina* on the wheat leaf surface; (b) close-up of wheat leaf rust uredinia showing the release of newly formed urediniospores on the leaf surface; (c) typical symptoms observed on infected wheat leaves in the field in wheat leaf rust epidemics; (d) uredinia pustules formed by *M. larici-populina* on the lower epidermis of a poplar leaf; (e) transverse section of a poplar leaf showing an uredinia full of urediniospores (note the orange colour of the leaf mesophyll invaded by infection hyphae right below the uredinia); (f) typical defoliated poplar trees observed in plantation after a strong epidemic at the end of summer. Photos courtesy of: (a) H. Goyeau and B. Parriaud, INRA France; (b) F. Suffert, INRA, France; (c) B. McCallum, Agriculture and AgriFood Canada; (d, e) B. Petre, INRA France; (f) C. Lorrain and S. Duplessis.

differentiate various fungal structures, such as protoaecia and pycnia. Haploid pycniospores and receptive hyphae are produced within pycnial nectar droplets. Fertilization can occur between spores and receptive hyphae of compatible mating types with pycniospores formed inside and exuded in droplets. Following plasmogamy, dikaryotic aecia differentiate inside the host and aeciospores are released and dispersed by the wind. Aeciospores infect the telial host in which uredinia and urediniospores are produced, which is followed by repeated cycles of vegetative growth on the telial host for an extended period of several weeks or months, typically during the summer (e.g. Fig. 1). In early autumn, uredinia differentiate into telia, which go through an overwintering period during which karyogamy will occur, giving rise to diploid dormant teliospores (examples of typical heteroecious cycles can be found in Hacquard *et al.* (2011) and Aime *et al.* (2014)).

The evolutionary processes behind the highly complex life cycles and diversification of rust fungi have been the subject of intense interest, particularly the notion of coevolution with their host plants. As obligate parasites, rust fungi were thought to have coevolved extensively with their hosts, suggesting that ancestral rust fungi may have infected ferns, then gymnosperms and then angiosperms, mirroring the diversification of available plant hosts. However, phylogenetic studies have shown that fern rust fungi were not ancestors and that some rust fungi infecting angiosperms derived from an ancestral family (Aime *et al.*, 2006, 2017). So far, the basal rust fungus *Caecoma torreyae* is found on gymnosperm, suggesting that rust fungi are more recent than

first thought (Aime *et al.*, 2006). Recent data based on phylogenetic comparisons between rust fungi and their host plants support a diversification through host jumps during coevolution with their hosts (McTaggart *et al.*, 2016; Aime *et al.*, 2017), and support the relative importance of the aecial host on the evolution of Pucciniales (Aime *et al.*, 2018). Heteroecious rust fungi alternate between unrelated host plants such as gymnosperms and angiosperms or monocots and eudicots; however, these hosts must share common ecosystems. This indicates that alternate hosts of rust fungi with unresolved life cycles may be found in the same ecosystem as the current known host, or that the alternate host has become extinct or replaced altogether in response to changing environmental conditions.

IV. The biology of rust infection

Advances in microscopy over the last century have led to precise descriptions of both the host infection process and the various stages of rust life cycles in a few rust fungal species (Littlefield & Heath, 1979; Harder, 1984). The concept of gene for gene hypothesis developed by Harold H. Flor pioneered identification of resistance genes in plants and enhanced our understanding of the plant immune system (Flor, 1971; Wu *et al.*, 2018). It is only in the last two decades or so that the first resistance and avirulence genes have been cloned in rust pathosystems (Ellis *et al.*, 2007). By the end of the 20th century, molecular studies of biotrophic fungi ushered in a new era of understanding of the mechanisms

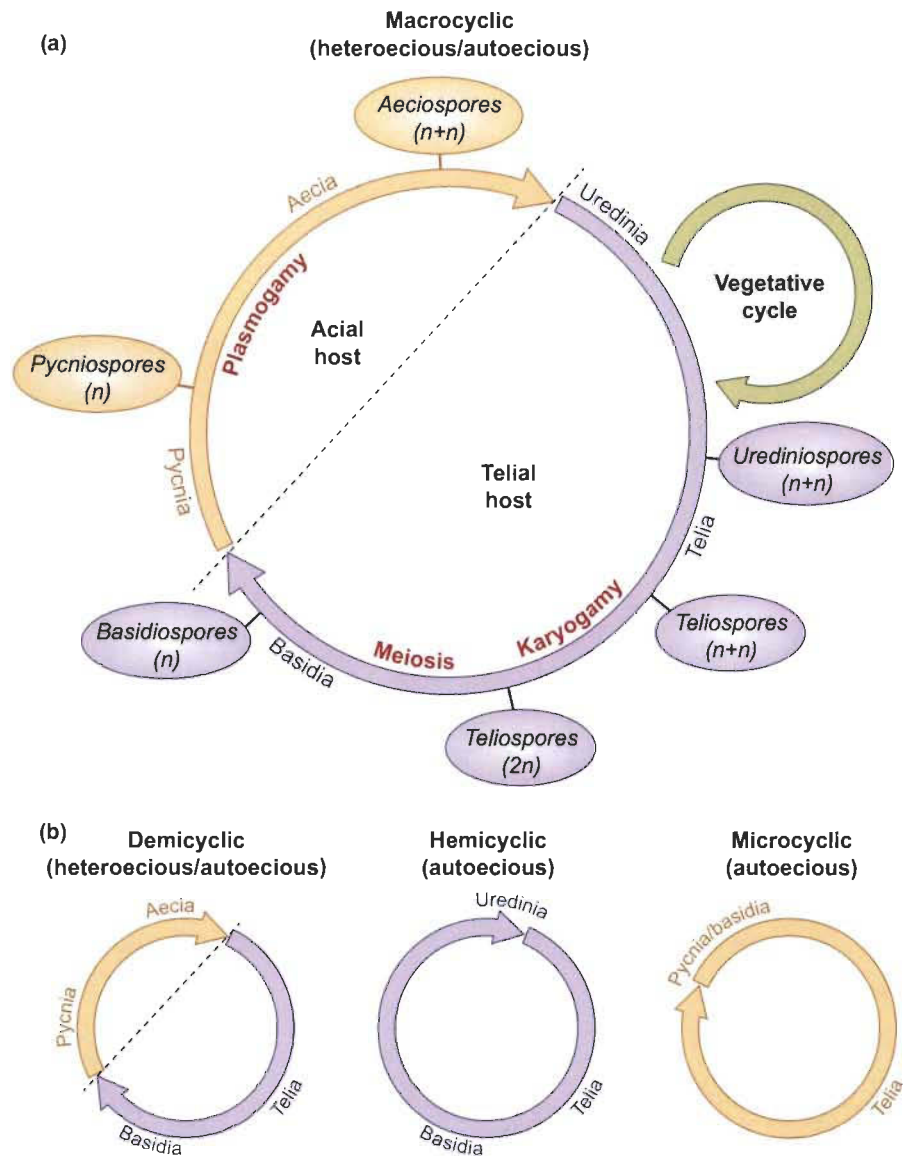


Fig. 2 Life cycles of rust fungi. (a) Macrocytic life cycle: macrocytic rust fungi produce five spore stages and they are autoecious or heteroecious (i.e. life cycle realized on one or two host plants, respectively); haploid basidiospores (n) infect the aecial host; pycnia contain pycniospores (n) of different mating types; after fertilization and plasmogamy, dikaryotic aecia release aeciospores ($n+n$) that infect the telial host; dikaryotic urediniospores ($n+n$) are then produced repeatedly through vegetative cycle during the summer season; teliospores ($n+n$) are formed in telia during autumn; karyogamy occurs in telia and meiosis initiates; basidia ($2n$) differentiate and produce basidiospores (n) after winter. (b) Derived rust life cycles: demicyclic rust fungi can be heteroecious or autoecious and do not form uredinia; hemicyclic rust fungi are autoecious with the absence of pycnia and aecial stages; the microcytic life cycle is reduced to telia and basidia or pycnia, according to species.

underlying the infection process. The following section provides an overview of these early key molecular investigations.

1. Early stages of infection: host penetration and haustorium formation

In heteroecious rust fungi, basidiospores infect the aecial host, and aeciospores and urediniospores infect the telial host by means of different penetration modalities. Monokaryotic basidiospores germinate and directly penetrate epidermal cells following appressorium differentiation (Hahn, 2000). To date, relatively little is known about the infection process in the aecial host. Dikaryotic aeciospores and urediniospores enter the host leaf through openings in the stomata, after appressorium differentiation (Voegelé *et al.*, 2009). Several transcriptomic analyses of appressoria formation of urediniospores led to identification of the genes responsible for metabolism and regulation of the cell division cycle, indicating that this structure plays an active role in the penetration

process (Hu *et al.*, 2007; Stone *et al.*, 2012; Talhinas *et al.*, 2014). Urediniospores from the soybean rust fungus directly penetrate their host through the epidermis, in a similar manner to basidiospores of heteroecious fungi, and form an appressorium with a high cell turgor pressure of 5.13 MPa (Loehrer *et al.*, 2014a), which correlates with what has been observed in other fungal plant pathogens (Ryder & Talbot, 2015). However, *P. pachyrhizi* basidiospores are hyaline and devoid of melanine, indicating a different penetration modality. To date, almost all physiological studies of host infection by rust fungi have focused on urediniospores (Voegelé *et al.*, 2009).

The haustorium represents a distinctive feature of obligate biotrophs. After getting inside the host tissue, the penetration hypha forms a substomatal vesicle from which an infection hypha differentiates; then a haustorium mother cell starts forming a penetration peg that secretes cell wall-degrading enzymes to penetrate the host cell wall in order to differentiate a haustorium within the host cell cavity (Hahn, 2000; Garnica *et al.*, 2014).

Fig. 3 Schematic representation of telial host infection by rust urediniospores. (a) The different steps of rust infection on the telial host from spore germination to sporulation. (b) Expression profiles of selected gene categories during the infection process (according to time-course transcriptomics in Duplessis *et al.*, 2011b; Dobon *et al.*, 2016). These profiles have been simplified according to the global average expression. Orange curves represent coordinated waves of expression for secreted protein genes; red curve represents the main expression profile for CAZymes, proteases and lipases (CPL) encoding genes; purple curve represents expression for transporter encoding genes. (c) Schematic representation of the interaction between rust infection hyphae and haustoria and plant host cells. Urediniospores germinate on the leaf surface and penetrate plant tissue through stomata. Infection hyphae grow between host cells and a haustorium is formed inside the host cell cavity. After several days of biotrophic growth, new urediniospores are produced and released on the leaf surface. Immunolocalization of candidate effectors is presented. *Melampsora lini* AvrM is localized at the periphery of fungal hyphae, in the haustorium, in the extrahaustorial matrix (EHM) and in flax cell cytosol. Rust-transferred proteins (RTP1p) from *Uromyces fabae* and *Uromyces striatus* are localized in the haustorium, in the EHM and in the nucleus and the cytosol in the later stages of colonization. SSP1, SSP2, SSP3 and SSP4 correspond to *Melampsora larici-populina* small secreted proteins (SSPs) Mlp123523, Mlp123227, Mlp123932 and Mlp37347, respectively. *Melampsora larici-populina* SSPs are localized at the periphery of the haustorium (SSP4, SSP3) and of infection hyphae colonizing plant mesophyll (SSP1), and at both the periphery of haustoria and sporogenous hyphae (SSP2). Close-up of haustoria–host cell interface shows enzymes and transporters of rust fungi characterized in *U. fabae*. PMA1 is a H⁺-ATPase that is active as the first haustoria formed and it localizes at the haustorial membrane. PMA1 sets up an electrochemical gradient necessary for nutrient uptake. ARD1 is a NADP⁺-dependent D-arabitol dehydrogenase localized in the lumen of haustoria. HXT1 represents the proton-motive force-driven hexose (D-glucose and D-fructose) transporter characterized in the haustorial membrane of the bean rust fungus. MAD1 is a mannitol dehydrogenase localized in haustoria, in sporogenous hyphae and in spores. AATs represent the three amino acid transporters (AAT1, AAT2, AAT3) of *U. fabae*. They all localize at the haustorial membrane.

The haustorium is an invagination within the host cell plasmalemma, which remains intact. In contrast to other obligate biotrophs, rust fungal haustoria are marked by the presence of a dense neckband that separates the extrahaustorial matrix surrounding the haustorial structure from the apoplasm (Garnica *et al.*, 2014). The formation of this specialized infection structure induces structural changes inside the infected host cell (Voegelé & Mendgen, 2003). It is also an intense site of expression of secreted proteins.

2. Nutrient uptake from the host

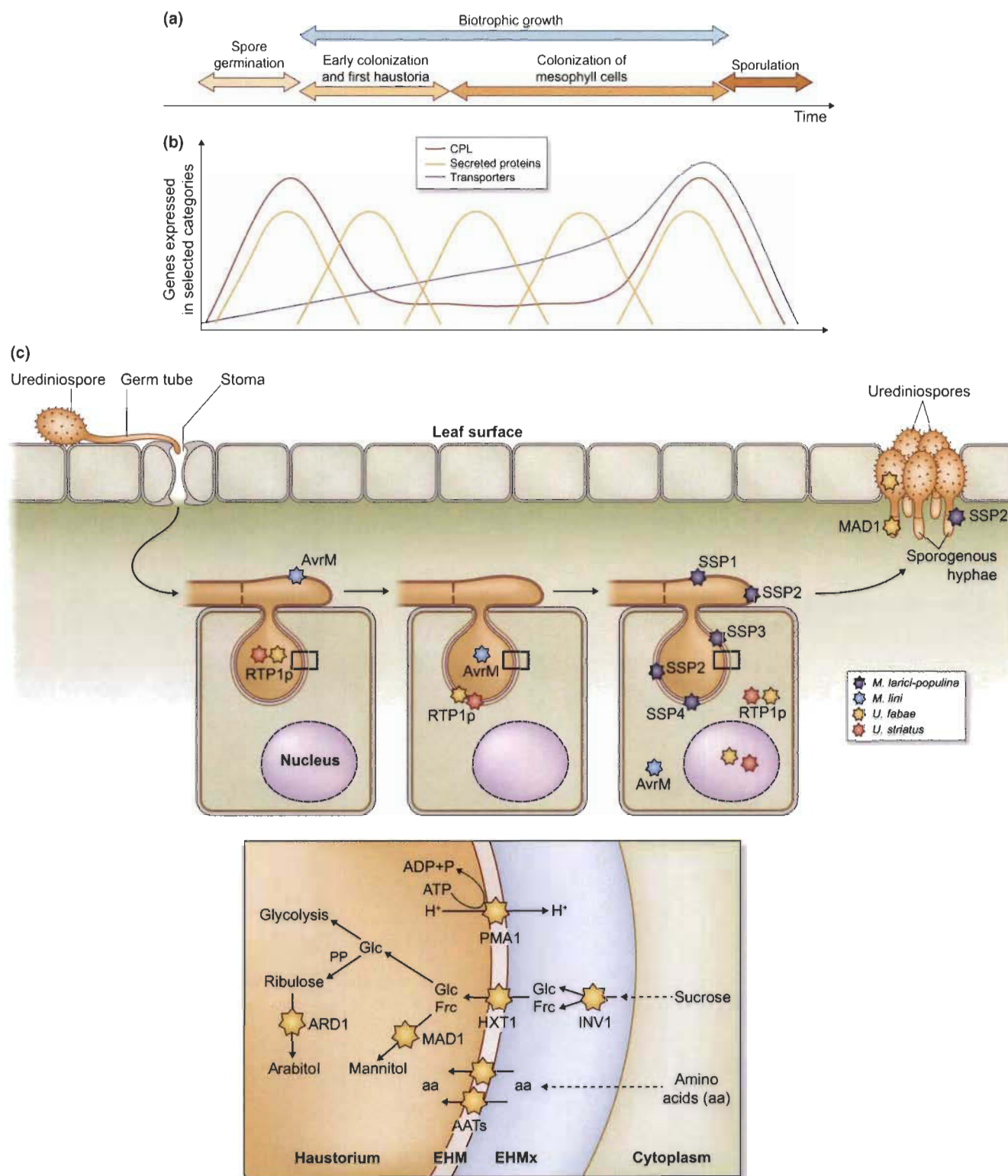
Rust fungi are completely dependent on the nutrients provided by living plant hosts to reproduce and complete their life cycles (Staples, 2000). Known nutrient uptake components of rust haustoria remain limited to a few sugar and amino acid transporters (Voegelé & Mendgen, 2011; Struck, 2015). Rust fungi are deficient in several essential pathways (e.g. nitrate and sulphur assimilation), which largely explains their dependence on nutrient uptake from their hosts (Duplessis *et al.*, 2011a; Kemen *et al.*, 2015). The first discovered amino acid permeases were AAT1p, AAT2p and AAT3p from *Uromyces fabae*, which showed a specific or a preferential expression in haustoria (Fig. 3; Hahn & Mendgen, 1997; Hahn *et al.*, 1997; Struck, 2015). Homologues of AAT genes upregulated in haustoria were later identified in other rust fungi (Hacquard *et al.*, 2010; Garnica *et al.*, 2013). The oligopeptide transporter (OPT) genes are more numerous in rust fungi than in other basidiomycetes and have been shown to be highly expressed during host infection and in haustoria of rust fungi (Duplessis *et al.*, 2011a; Garnica *et al.*, 2013). Recently, OPT genes were associated with virulence in the smut fungus *Ustilago maydis*, suggesting that extracellular peptides represent an important source of nutrients for the fungus (Lanver *et al.*, 2018). So far, no functional characterization of OPT has been conducted in rust fungi. Host oligopeptides could also represent an important source of nitrogen and sulphur for rust fungi.

Uromyces fabae H⁺-ATPase PMA1p was identified from isolated haustoria and functionally characterized in yeast. PMA1p activity establishes the electrochemical proton gradient that provides

energy for transport processes and nutrient uptake (Fig. 3; Struck *et al.*, 1996, 1998). The important role of proton gradient in the absorption of sugar was later shown for the *U. fabae* hexose transporter HXT1p localized at the haustorium membrane and its role as a proton-motive D-glucose and D-fructose transporter was demonstrated in yeast (Fig. 3; Voegelé *et al.*, 2001). Plants present low concentrations of monomeric hexoses in leaves and sucrose is the main sugar in this organ (Lohaus *et al.*, 2001; Weber & Roitsch, 2000). The *U. fabae* invertase INV1p expressed during infection and localized in the extrahaustorial matrix was reported to be necessary for the breakdown of sucrose in D-glucose and D-fructose, the substrates for HXT1p (Fig. 3; Voegelé *et al.*, 2006). Recently, the invertase PsINV was investigated in the wheat rust *P. striiformis* sp. *tritici*, and showed a high efficiency for sucrose hydrolysis as well as an increased expression during wheat infection (Chang *et al.*, 2017). Other genes related to metabolism, such as the glucokinase GLK1p, the mannitol dehydrogenase 1 (MAD1p) and the NADH⁺ dependent D-arabitol dehydrogenase, ARD1p, from *U. fabae* have provided valuable information about carbon storage and utilisation in this rust fungus (Fig. 3; Link *et al.*, 2005; Voegelé *et al.*, 2005; see Voegelé & Mendgen, 2011 for details). Homologues of these different genes were also found in other rust fungi, and they showed high levels of expression during infection of their respective hosts, supporting the mechanisms of glucose uptake in rust fungi (Hacquard *et al.*, 2010; Duplessis *et al.*, 2011a,b; Garnica *et al.*, 2013; Zheng *et al.*, 2013). No sucrose transporter has been identified in rust fungi (Duplessis *et al.*, 2011a), indicating that hexose transport may be essential in haustoria. No recent functional analyses of uptake mechanisms have been conducted in rust fungi, except for *P. striiformis* f. sp. *tritici*.

3. Infection control centre: the haustorium as a multifaceted structure

The haustorium is not only a specialized machine for nutrient diversion, it is also a site of intense expression of secreted proteins, including effectors (Duplessis *et al.*, 2012; Garnica *et al.*, 2014). In rust fungi, secreted effector proteins have been shown to reside inside host cell cytoplasm; however, the way in which effectors



translocate from the haustorium to the host cell remains unknown (Petre & Kamoun, 2014). Only a handful of effectors have been described in rust fungi (Petre *et al.*, 2014). The products of several avirulence genes of the flax rust fungus *M. lini* are recognized by resistance proteins inside plant cells where they are expected to play roles as effectors (Duplessis *et al.*, 2012). Rust-transferred proteins

RTP1p from *Uromyces* spp. were the first fungal proteins identified as being capable of translocating into their host cells (Kemen *et al.*, 2005). The proteins were localized inside host cells and then in the host nucleus, suggesting a specific traffic towards its final localization (Fig. 3). The *M. lini* effector AvrM has also been detected inside flax cells by immunogold transmission electron

microscopy (Fig. 3; Rafiqi *et al.*, 2010). In both examples, trafficking from the fungal haustoria to the host cell was temporally regulated. RTP1 is part of a multigene family conserved across Pucciniales, including basal species like *H. vastatrix* (Fernandez *et al.*, 2012; Pretsch *et al.*, 2013; Link T. *et al.*, 2014; Link T.I. *et al.*, 2014). Detailed analyses of RTP1 indicated a dual function as a protease inhibitor *in vitro* and a filament-forming protein in the extrahaustorial matrix that could act as a stabilizing factor for the fungal structure (Kemen *et al.*, 2013; Pretsch *et al.*, 2013). Immunolocalization of two RTP homologues in the poplar rust *Melampsora larici-populina* revealed a dual localization at the periphery of haustoria and infection hyphae (Fig. 3; Hacquard *et al.*, 2012). Two other poplar rust-secreted proteins were detected *in planta* by immunolocalization: a homologue of the flax rust avirulence effector AvrL567, which was found at the periphery of poplar rust haustoria but not beyond in the host cells; and SSP15, which is localized at the periphery of both haustoria and spore-forming cells in uredinia (Fig. 3; Hacquard *et al.*, 2012). Although all reported observations pertaining to rust proteins show evidence of complex and diverse patterns of localization within infected host tissues, their function during the infection process remains unclear (Petre *et al.*, 2014).

V. Rusts in the genomics era: the ever-expanding list of candidate effector genes

In the last 10 years, as many as 20 genomes of 11 different rust fungi have been reported (Supporting Information Table S1; see Aime *et al.*, 2017 for details). Beyond the basic groundwork these studies provided about predicted genes and the genome architecture and organisation; advances in genome-wide comparative analyses, transcriptome and population studies and prediction of secreted proteins to identify rust effector proteins have considerably extended our understanding of rust fungal biology (Table S1).

1. Sequencing rust fungal genomes

Rust fungi have large genomes in the range 60 to > 300 Mb (Aime *et al.*, 2017) but flow cytometry estimates the largest rust genome at > 2 Gb (Tavares *et al.*, 2014; Ramos *et al.*, 2015). Rust genomes are sequenced from dikaryotic spores and they present a remarkable degree of heterozygosity (Persoons *et al.*, 2014; Upadhyaya *et al.*, 2014; Cuomo *et al.*, 2017). High proportions of repeat elements in the rust genome, ranging from 18% to 75%, have made rust genome assembly extremely difficult. This is particularly true for rust genomes sequenced only by using short-read technologies (see Aime *et al.*, 2017 for details). Rust fungi possess large numbers of genes (15 000–20 000 per genome) and about half of them lack functional annotation. Many genes have been reported as specific to Pucciniales, which may also be reflective of innovations in their biology (Aime *et al.*, 2017). The lack of complete and high-quality rust genome assemblies and standardized practices in the community have somehow restricted the scope of comparative studies. Next-generation long-read sequencing technologies combined with improved assembly tools both support improved assembly of large rust genomes. Long-read sequencing approaches

recently used for genomes of *P. striiformis* f. sp. *tritici* 104E and two isolates of the oat crown rust *Puccinia coronata* f. sp. *avenae* provided higher-quality *de novo* assemblies (Miller *et al.*, 2017; Schwessinger *et al.*, 2017) and supported phasing of haplotypes for a significant portion of these dikaryotic genomes (50–92%). The genome size of these two rust species were smaller on average than the range of estimated rust genome sizes (83–105 Mb), and it has not yet been determined if genomes in the range of 1 Gb or more with > 70% repetitive content will affect assembly performance. Generating complete and phased genome assembly has proved to be important not only for highlighting intrinsic differences between two rust nuclei, but also for elucidating the architecture of rust genomes and, ultimately, insights into the impact of transposition on rust virulence and evolution.

2. Prediction of candidate effectors in rust secretomes

Soon after the conceptualization of plant pathogen effectors and their importance in the infection process, it was established that secreted proteins of unknown function can represent candidate effectors (Win *et al.*, 2012). Therefore, secretome prediction has been systematically applied to rust genomes with the promising possibility of identifying effectors. Some features, such as a smaller size, a higher cysteine content or taxonomical specificity, have been used to identify candidate effectors (Lorrain *et al.*, 2015; Sperschneider *et al.*, 2017a). Based on these criteria, long lists of rust candidate effectors have been identified (Table S1; Pendleton *et al.*, 2014; Duplessis *et al.*, 2011a,b; Nemri *et al.*, 2014; Saunders *et al.*, 2012; Cantu *et al.*, 2013). The number of secreted protein-encoding genes can vary depending on the precise criteria used to flag them as candidates, but they generally range between 1000 and 2000 per genome (*c.* 10% of the coding space). Both -orphan and gene families are found among Pucciniales-specific candidate effectors. Expanded gene families of unknown function are a common feature of rust fungal genomes, among which are several secreted protein families (Aime *et al.*, 2017). A recent approach has been to use machine-learning prediction tools to identify candidate effectors within fungal secretomes, improve prediction of their localization *in planta* and differentiate between apoplastic and cytoplasmic effectors (see details in Table S2; Sperschneider *et al.*, 2015, 2016, 2017b,c, 2018). These tools are particularly valuable as they are not based on *a priori* criteria and can be used to revisit sets of candidate effectors predicted in rust genomes (Table S2). The report of more *bona fide* effectors should improve future machine learning-based prediction, provided that rust effectors do have conserved features to truly support their prediction.

3. Transcriptomics of rust infection: pointing the needle in the haystack

Specific gene expression *in planta* is a critical feature in defining pathogenic effectors. Thus, transcriptomics has been the approach of choice for reducing large sets of rust candidate effectors to smaller numbers of priority targets for functional study (Table S1; Duplessis *et al.*, 2012, 2014). The first pioneer transcriptomic studies used cDNA libraries to identify haustorially expressed

secreted proteins from isolated haustoria of *U. fabae* and *M. lini* (Hahn & Mendgen, 1997; Catanzariti *et al.*, 2006). RNA-sequencing (RNAseq) has been instrumental in providing valuable insights about rust fungi without reference genomes, for example *H. vastatrix*, *Gymnosporangium* spp., *P. pachyrhizi*, *Uromyces appendiculatus* (Table S1; Loehrer *et al.*, 2014b; Link T.I. *et al.*, 2014; Cristancho *et al.*, 2014; Tao *et al.*, 2017). Transcriptomic analyses of host infection and isolated haustoria have helped to prioritize candidate effectors in different rust species (Duplessis *et al.*, 2014; Aime *et al.*, 2017). Many genes expressed in the late stages of colonization encode carbohydrate active enzymes, proteases, lipases and transporters (Fig. 3; Duplessis *et al.*, 2011b; Hacquard *et al.*, 2011, 2012). These findings are consistent with nutrient uptake by sugar and amino acid transporters once haustoria are established in the plant tissue (Voegelé & Mendgen, 2011; Struck, 2015). Different sets of candidate effectors are expressed sequentially in coordinated waves of expression from early to late stages of infection (i.e. from spore germination to sporulation; Fig. 3; Dobon *et al.*, 2016; Duplessis *et al.*, 2011b; Hacquard *et al.*, 2011, 2012, 2013; Bruce *et al.*, 2014; Cantu *et al.*, 2013; Fernandez *et al.*, 2012; Huang *et al.*, 2011; Tremblay *et al.*, 2013; Rutter *et al.*, 2017). Small secreted proteins are among the most highly expressed genes in fungal infection structures at the sporulation stage (Hacquard *et al.*, 2010). Such coordinated expression patterns indicate that rust fungi may release and deliver early and late effectors with different roles in establishing and maintaining biotrophy. Although studying the infection process in the primary host of rust fungi has proved highly informative, gene expression at other stages of the rust life cycle has largely been overlooked. Very little information is available about infection mechanisms of alternate hosts and sexual stages. To date, only a handful of studies have considered gene expression on the alternate host and they have revealed concomitant expression of sets of candidate effectors in both hosts as well as in each host individually (Xu *et al.*, 2011; Liu *et al.*, 2015; Cuomo *et al.*, 2017; Lorrain *et al.*, 2018a). This type of approach is particularly helpful for narrowing down the list of effectors related to the main host where disease causes the most significant damage.

4. Population genomics for unveiling effector and avirulence genes

Analysing population genomics and studying evolutionary processes have proved valuable in identifying avirulence genes among candidate effectors (Møller & Stukenbrock, 2017). The pressure exerted by components of the host plant immune system can be detected by traces of positive selection in effector genes under surveillance. Signatures of rapid evolution and polymorphism have been extensively studied in filamentous pathogens, including rust fungi (Raffaele & Kamoun, 2012; Stukenbrock, 2013; Persoons *et al.*, 2014; Upadhyaya *et al.*, 2014). Three *M. lini* avirulence genes (*AvrL567*, *AvrP4* and *Avr123*) were found under positive selection at the intraspecific and/or interspecific levels (Dodds *et al.*, 2004; Barrett *et al.*, 2009). A genome-wide search for single nucleotide polymorphisms or insertion/deletion in rust fungal isolates has enabled the detection of avirulence candidates among

secreted protein genes (Cantu *et al.*, 2013; Bruce *et al.*, 2014; Persoons *et al.*, 2014; Upadhyaya *et al.*, 2014; Bueno-Sancho *et al.*, 2017; Cuomo *et al.*, 2017; Wu *et al.*, 2017). Sequencing large volumes of isolates is now possible for relatively low cost, and high-density genetic maps can be constructed through the resequencing of progenies or by restriction site detection sequencing (RAD-seq) in rust fungi (Pernaci *et al.*, 2014; Anderson *et al.*, 2016; Yuan *et al.*, 2018). Two recent studies demonstrated the degree to which genomics has impacted the field by combining genomics and resequencing approaches to identify *AvrSr35* and *AvrSr50*, two *P. graminis* f. sp. *tritici* avirulence effector genes (see details in Section VI; Chen *et al.*, 2017; Salcedo *et al.*, 2017). It is very likely that population genomics and genome-wide association studies will unveil even more key pathogenicity determinants over the next few years.

VI. Functional characterization of rust effectors

Lists of candidate effectors have now been established for most rust fungi that are major threats to agriculture: some may play a critical role in establishing a successful infection, some can be targets of host resistance proteins (avirulence proteins) and some may reveal host susceptibility by identifying their targets inside the host cell. Investigating candidate effectors can foster selection of more durable resistance to rust diseases. That said, the number of effectors to be scrutinized is strikingly large and rust fungi as obligate biotrophs are particularly difficult to manipulate. To respond to these two challenging aspects, high- to medium-throughput pipelines have been developed for more effective functional characterization of candidate rust effectors and to help formulate hypotheses about their functions in plant cells. Effectoromic screens essentially allow us to identify cellular and molecular targets of given candidates and to draw hypotheses regarding their role in pathogenesis. Heterologous plant systems are favoured for investigating nonmodel candidate effector functions (Petre *et al.*, 2014; Lorrain *et al.*, 2018b). Fig. 4 shows current approaches used to identify and unravel functions of candidate effectors in the plant cell. In the following section, we describe the different strategies and approaches applied over the past 5 years to investigating effectors in rust fungi. We choose to report these studies by rust species according to the importance of the community-driven efforts and to distinguish screening approaches and work on specific genes. Table 1 summarizes the number of candidates assayed in each rust species.

1. Effectoromics of the yellow rust fungus *P. striiformis*

The rust species *P. striiformis* f. sp. *tritici* has attracted the greatest attention from the scientific community. Both medium-throughput screening and dedicated investigations have been used to functionally characterize candidate effectors and pathogenesis-related functions. Multiple efforts from different research groups have used different pipelines and approaches. The host-induced gene silencing (HIGS) approach developed to silence *Puccinia* spp. genes using the *Barley stripe mosaic virus* (BSMV) on wheat (Yin *et al.*, 2011) has been instrumental in assessing rust genes in wheat–rust interactions (Tang *et al.*, 2017).

(a)

Effector mining		
Genome mining	Transcriptomics/proteomics	Comparative genomics
<ul style="list-style-type: none"> • Secretion prediction • Functional annotation • Localization prediction (LOCALIZER/ApoplastP) • Effector prediction (EffectorP) • Cysteine content 	<ul style="list-style-type: none"> • Expression during infection (<i>in planta</i>) • Specific host expression (heteroecious species) • Expression in haustoria 	<ul style="list-style-type: none"> • Population genomics (selection indices) • Variant identification (SNPs, InDels, ...) • Taxonomic specificity (order, family)

(b)

Candidate secreted effector proteins (CSEPs)

Functional characterization		
Secretion validation	Virulence properties	Gene silencing
<ul style="list-style-type: none"> • Yeast signal sequence trap system 	<ul style="list-style-type: none"> • Plant cell death assay (HR induction/suppression) • Callose deposition • ROS accumulation • Infection assays 	<ul style="list-style-type: none"> • Host/virus-induced gene silencing • Pathogenicity assays (pustule number/area, number/size of infection hyphae and haustoria)
Expression <i>in planta</i>	Plant interactors	Biochemistry
<ul style="list-style-type: none"> • Subcellular localization of fluorescent-tagged CSEPs (host, heterologous plants) • Immunolocalization • Phenotyping plants expressing CSEPs • Infection assays on plants expressing CSEPs 	<ul style="list-style-type: none"> • Coimmunoprecipitation • Yeast two-hybrid assays • Bimolecular fluorescence complementation • Pull-down assay 	<ul style="list-style-type: none"> • Recombinant protein • Expression (Western blot) • NMR/crystallography • Structural homologies

Fig. 4 Common approaches applied to study effector biology of rust fungi. (a) Effector mining shows genomic and transcriptomic approaches applied to rust fungi to identify candidate secreted effector proteins (CSEPs). (b) Functional characterization details the different approaches and methods deployed in the scientific community to scrutinize role and functions of rust fungal CSEPs. SNP, single nucleotide polymorphism; HR, hypersensitive response; ROS, reactive oxygen species; NMR, nuclear magnetic resonance.

Two medium-throughput screens have focused on *P. striiformis* f. sp. *tritici* candidate effectors. Ramachandran *et al.* (2017) scrutinized 20 Pucciniaceae candidates, including nine from *P. striiformis* f. sp. *tritici*. Seven were able to suppress a hypersensitive response (HR) triggered in *Nicotiana benthamiana*, and plant cell death suppression was confirmed in wheat for one candidate through delivery by *Pseudomonas fluorescens* (Ramachandran *et al.*, 2017). In another study, 16 mature candidate effectors were transiently expressed in fusion with green fluorescent protein (GFP) in *N. benthamiana* to determine their subcellular localization and to identify their putative interactors by coimmunoprecipitation (coIP) and mass spectrometry (MS) (Petre *et al.*, 2016a). Nine were informative in terms of localization or specific interactions with plant proteins. One candidate showed accumulation in processing bodies and interaction with an enhancer of mRNA decapping protein. The interaction between the wheat enhancer of mRNA decapping protein 4 (TaEDC4) and the yellow rust effector was confirmed by coIP in *N. benthamiana* (Petre *et al.*, 2016a). Taken together, these two screens led to the identification of 16 candidates of interest with virulence-related activities, a specific subcellular localization and/or a putative plant protein interactor (Table 1). A small-throughput screen for immunity suppression was set for six other candidate effectors, identifying PEC6 as a candidate with virulence properties (Liu C. *et al.*, 2016). PEC6 suppressed plant immune responses in *N. benthamiana*, enhanced bacterial growth in *Arabidopsis* and showed interaction with an adenosine kinase by yeast two-hybrid and bimolecular

fluorescence complementation (Liu C. *et al.*, 2016; Liu J. *et al.*, 2016). Fungal infection was also reduced in wheat when PEC6 was silenced by HIGS (Liu C. *et al.*, 2016; Liu J. *et al.*, 2016).

Several studies investigated yellow rust candidate effectors and signalling proteins through separate efforts. PstSCR1, a small cysteine-rich candidate effector, induces plant cell death when expressed with its signal peptide or when injected as a purified protein in *N. benthamiana*, but its potential role in *P. striiformis* f. sp. *tritici* remains unknown (Dagvadorj *et al.*, 2017). A pipeline of diversified molecular approaches has been established to unravel the role of *P. striiformis* f. sp. *tritici* proteins during interaction with wheat and facilitated the study of more than 10 yellow rust genes over the past 5 years (Tang *et al.*, 2015a,b; Tang *et al.*, 2017). Selected candidate genes are most often assayed by HIGS to validate their involvement in pathogenicity through quantitative estimation of infection-related traits. Depending on the given candidate, different complementary approaches were applied, such as subcellular localization in heterologous systems or in wheat protoplasts (Cheng *et al.*, 2015, 2016a; Wang *et al.*, 2016; Zhu *et al.*, 2017a,b), complementation assays on yeast or fungal pathogens (Guo *et al.*, 2011; Liu *et al.*, 2014; Liu J. *et al.*, 2016; Zhu *et al.*, 2017a), yeast invertase secretion assays (Cheng *et al.*, 2016a; Liu J. *et al.*, 2016; Wang *et al.*, 2016), plant cell death BAX-triggered suppression assays (Cheng *et al.*, 2016a,b; Liu J. *et al.*, 2016) and overexpression in yeast coupled with the monitoring of stress responses (Cheng *et al.*, 2015, 2016b; Tang *et al.*, 2015a,b; Jiao *et al.*, 2017; Zhu

Table 1 Effectoromic screens conducted in rust fungi to identify candidate secreted effector proteins (CSEPs).

References	Approaches	Screened CSEPs	Highlighted CSEPs
<i>Hemileia vastatrix</i> Maia <i>et al.</i> (2017)	Plant cell death suppression in coffee leaves	30	1
<i>Melampsora larici-populina</i> Petre <i>et al.</i> (2015, 2016a,b)	Localization and ColP/MS ^a of CSEPs in <i>Nicotiana benthamiana</i>	20	8
Germain <i>et al.</i> (2017)	Infection assays and localization of CSEPs in <i>Arabidopsis</i>	16	3
<i>Melampsora lini</i> Zhang <i>et al.</i> (2016)	Recombinant protein production and structural analysis	4	4
<i>Phakopsora pachyrhizi</i> Qi <i>et al.</i> (2018)	Localization and plant cell death suppression in <i>N. benthamiana</i> , <i>Arabidopsis</i> and yeast, immune response in tomato and pepper	82	17
de Carvalho <i>et al.</i> (2017)	Localization and plant cell death suppression in <i>N. benthamiana</i>	6	6
Kunjeti <i>et al.</i> (2016)	Plant cell death suppression in <i>N. benthamiana</i>	6	2
<i>Puccinia graminis</i> f. sp. <i>tritici</i> Yin <i>et al.</i> (2015)	HIGS ^a to select proteins related to pathogenesis	76	10
Upadhyaya <i>et al.</i> (2014)	HIGS to select proteins related to pathogenesis	25	1
Ramachandran <i>et al.</i> (2017)	Plant cell death suppression in <i>N. benthamiana</i>	11	2
Chen <i>et al.</i> (2017)	Coexpression with <i>R</i> gene in <i>N. benthamiana</i>	41	1
<i>Puccinia striiformis</i> f. sp. <i>tritici</i> Petre <i>et al.</i> (2016b)	Localization and ColP/MS of CSEPs–fluorescent tag in <i>N. benthamiana</i>	16	9
Ramachandran <i>et al.</i> (2017)	Plant cell death suppression in <i>N. benthamiana</i>	9	7
Liu C. <i>et al.</i> (2016)	Plant cell death suppression in <i>N. benthamiana</i> and infection assays in <i>Arabidopsis</i>	6	1
<i>Uromyces appendiculatus</i> Cooper & Campbell (2017)	HIGS to select proteins related to pathogenesis	5	3

^aColP/MS, coimmunoprecipitation coupled with mass spectrometry; HIGS, host-induced gene silencing.

et al., 2017a,b). Among the proteins characterized through the aforementioned pipeline, six were identified as protein kinases or regulators of kinases in fungal transduction pathways (Guo *et al.*, 2011; Jiao *et al.*, 2018; Cheng *et al.*, 2015, 2016b; Qi, Zhu *et al.*, 2018) whereas others corresponded to a mitochondrial adenine nucleotide translocase antiporter (PsANT; Tang *et al.*, 2015a,b), a superoxide dismutase (Liu J. *et al.*, 2016), an isocitrate lyase required for spore germination and one transcription factor. Two of the studied proteins were identified as candidate effectors (Cheng *et al.*, 2016a; Wang *et al.*, 2016). PSTha5a23 is a specific *P. striiformis* f. sp. *tritici* candidate highly expressed during interaction with wheat that localized in the cytoplasm of wheat protoplasts, suppressed cell death in *N. benthamiana*, and supported increased pustule number when overexpressed in wheat (Cheng *et al.*, 2016a). The candidate effector PNPI was found to interact in wheat with the homologue of the key defence regulator NPR1 where it competes with the transcription factor TGA2.2 which is known to activate *Pathogenesis-related* gene expression (Wang *et al.*, 2016). The functional pipeline applied to characterize molecular determinants of *P. striiformis* f. sp. *tritici* proved to be both useful and efficient in delivering new insights into pathogenesis of the wheat yellow rust fungus.

2. Effectoromics of the wheat stem rust fungus *P. graminis* f. sp. *tritici*

A high-throughput effectoromic screen based on BSMV-mediated HIGS was applied to a large set of Pucciniaceae candidate effectors conserved in the three wheat rust fungi *P. graminis* f. sp. *tritici*,

P. striiformis f. sp. *tritici* and *P. tritici* (Yin *et al.*, 2015). Among 86 silenced genes, 10 were required for full infection by *P. graminis* f. sp. *tritici*, including secreted proteins, metabolic and transport-related genes (Yin *et al.*, 2015). Reduced development of *P. striiformis* f. sp. *tritici* and *P. tritici* was also observed through transient silencing of four and three of those genes, respectively, indicating that screens performed at the level of taxonomical families can reveal crucial functions for infection shared between rust species (Yin *et al.*, 2015). A tryptophan 2-monooxygenase (Pgt-IaaM) gene involved in auxin synthesis is required for full virulence of *P. graminis* f. sp. *tritici* and displayed pleiotropic auxin-related phenotypes when expressed in *Arabidopsis* (Yin *et al.*, 2014). This study demonstrates the modulation of the plant hormone balance by a rust fungus. In their assay, Ramachandran *et al.* (2017) also screened plant cell death suppression properties of 11 wheat stem rust candidate effectors, and two were able to suppress HR in *N. benthamiana*. Another effectoromic screen was established to deliver rust candidate effectors in wheat through a specific type III secretion system (*Pseudomonas fluorescens* effector to host analyser strain EtHAN) (Upadhyaya *et al.*, 2014). A total of 25 *P. graminis* f. sp. *tritici* candidates were monitored for plant cell death induction and one was able to induce a genotype-specific HR, suggesting it might be an avirulence protein (Upadhyaya *et al.*, 2014). Four candidate effectors predicted to accumulate in nuclei or in chloroplasts of plant cells by LOCALIZER (cf. Section V.2) were validated in *N. benthamiana* (Sperschneider *et al.*, 2017b).

The wheat stem rust avirulence genes *AvrSr35* and *AvrSr50* were recently identified through genomic and functional approaches. In

one case, *AvrSr35* was identified in rust fungal mutants obtained by induced mutagenesis of an avirulent fungal isolate. *AvrSr35* encodes a small secreted protein of 578 amino acids, and the interaction with the wheat resistance gene *Sr35* was verified by transient coexpression in *N. benthamiana* as well as by infiltration of the purified protein in wheat leaves (Salcedo *et al.*, 2017). *AvrSr35* and *Sr35* colocalized with the endoplasmic reticulum and interacted in plant cells; however, whether the interaction is direct or indirect remains to be determined. In the second example, sequencing the genomes of an avirulent *AvrSr50* isolate and a spontaneous virulent mutant derived from this isolate identified a 2.5 Mb region showing loss-of-heterozygosity (Chen *et al.*, 2017). A total of 41 genes encoding haustorially expressed secreted proteins present in this region were transiently coexpressed without their signal peptide with the wheat resistance gene *Sr50*, and a single candidate triggered a HR (Chen *et al.*, 2017). *AvrSr50* showed a nucleocytoplasmic localization in *N. benthamiana* and exhibited a cell death suppression activity, indicating that it could have a role in suppressing defence response (Chen *et al.*, 2017). These back-to-back papers demonstrate the great promises of combining functional assays with (population) genomics and transcriptomics to identify new rust avirulence genes.

3. Effectoromics of the wheat leaf rust *P. tritici*

Progress has been made in the study of wheat leaf rust pathogenicity through methodological developments. Complementation of a *U. maydis* mitogen-activated protein kinase (MAPK) knockout mutant with a *P. tritici* MAPK restored virulence, demonstrating that mutant complementation could be an effective way of studying rust functions (Hu *et al.*, 2007). The knockdown of *P. tritici* signalling genes encoding an adenylate cyclase, a MAPK and a calcineurin through BSMV-mediated HIGS resulted in suppressed disease phenotypes (Panwar *et al.*, 2013). Such a host-delivered RNA interference system could be effective through generations in transgenic wheat lines, conferring protection against rust disease (Panwar & Bakkeren, 2018). A co-bombardment assay of selected secreted proteins with a β -glucuronidase (GUS) expressing vector was also developed to detect potential avirulence candidates. In this system, a reduction in GUS expression may result from a HR caused by the recognition of the secreted protein by the immune system. Two secreted proteins were flagged as potential avirulence candidates using this biolistic assay (Segovia *et al.*, 2016).

4. Effectoromics of the Asian soybean rust fungus *P. pachyrhizi*

One of the most advanced effectoromic screens performed in rust fungi was recently reported for *P. pachyrhizi* (Qi *et al.*, 2018). The selection of 82 candidate effectors (PpECs) was based on haustoria RNAseq (Link T. *et al.*, 2014; Link T.I. *et al.*, 2014). Qi *et al.* (2018) cloned the 82 PpECs in different vectors in order to: scrutinize their subcellular localization in *N. benthamiana*; assay induction of an immune response in pepper and tomato; and assay cell death suppression in *N. benthamiana*, *Arabidopsis* and yeast. Among the 82 PpECs, 17 suppress plant immunity and

30 PpECs target different cell compartments, such as nucleus, nucleolus and cytoplasm (Qi *et al.*, 2018). The secretion of one candidate, PpEC23, was confirmed in yeast and it was further shown to suppress the immune response in soybean. Interaction of PpEC23 with the soybean transcription repressor GmSPL121 was shown by yeast two-hybrid and pull-down. PpEC23 interacts with itself in the cytoplasm and it localizes in the nucleus when it interacts with GmSPL121 (Qi *et al.*, 2016). With this large effectoromic screen, the authors were able to propose a shortlist of 17 *P. pachyrhizi* priority candidates that can now be specifically addressed in dedicated efforts (Qi *et al.*, 2018). This clearly illustrates the power of systematic and methodological approaches to large-scale screening applied to rust fungi. Parallel screening efforts identified other soybean rust candidate effectors. For instance, six candidates showed suppression of plant cell death triggered by *Pseudomonas syringae* pv. *tomato* DC3000 and an accumulation in nucleus and cytosol after transient expression in *N. benthamiana* (de Carvalho *et al.*, 2017). In another study, two out of six selected candidates were found to increase *P. infestans* infection in *N. benthamiana*, suggesting a role in the virulence of the soybean rust fungus (Kunjeteri *et al.*, 2016).

5. Effectoromics of the poplar rust fungus *M. larici-populina*

Twenty candidate effectors of the poplar rust fungus have been investigated in *N. benthamiana* to identify their cellular localization and their putative plant partners (Pette *et al.*, 2015). Eight candidates showed either a specific localization when transiently expressed without their signal peptide in fusion with GFP or a specific interaction with plant proteins by coIP/MS. The candidate Mlp124017 was shown to interact with the poplar TOPLESS-related protein 4 by coIP when coexpressed in *N. benthamiana* (Pette *et al.*, 2015). This study also revealed a candidate effector targeting chloroplasts (chloroplast-targeted protein 1; MlpCTP1). A transit peptide in MlpCTP1 is necessary and sufficient to accumulate in chloroplasts (Pette *et al.*, 2016b). This gene is part of a Melampsoraceae family and other members were shown to possess a transit peptide to target chloroplasts in plant cells. Sixteen other *M. larici-populina* candidates were assayed in *A. thaliana* through stable expression of GFP fusion to determine their subcellular localization (Germain *et al.*, 2017). Three candidates showed an accumulation in plasmodesmata, chloroplasts and cytoplasmic bodies, respectively; 11 promoted *Hyaloperonospora arabidopsidis* growth *in planta*; and five promoted *P. syringae* growth *in planta*. Three candidates promoted both oomycete and bacterial growth and may be considered as virulence factors (Germain *et al.*, 2017). Investigations conducted in two different heterologous systems highlighted 11 poplar-rust candidate effectors with specific localization and/or plant protein interactors (Table 1).

6. Effectoromics of the flax rust fungus *M. lini*

Melampsora lini is an advanced rust model for the genetic dissection of avirulence genes, and most known *bona fide* rust fungal effectors have been identified in this fungus (Pette *et al.*, 2014). Detailed reviews have addressed the identification and study of avirulence

genes in the flax rust fungus and have been a prime source of information (e.g. Ellis *et al.*, 2007). A recent small screen investigated biochemical properties and structures of *M. lini* effectors (Zhang *et al.*, 2017a). This is the first screen to be aimed specifically at the production of recombinant rust effector proteins. Avirulence proteins AvrP, AvrP123, AvrP4 and AvrM were produced in several *Escherichia coli* strains. To date, AvrL567, AvrM and AvrP are the only rust fungal effectors with characterized three-dimensional structures. They exhibit completely different structures (i.e. β -sandwich fold, L-shape fold and elongated zinc-finger-like fold, respectively), but they do, however, all display surface polymorphic residues involved in the recognition by the flax resistance proteins L5/L6/L7, M and P, respectively (Wang *et al.*, 2007; Ve *et al.*, 2013; Zhang *et al.*, 2017b). No structural homology was found for the three flax rust fungal effectors in structure databases (Wang *et al.*, 2007; Ve *et al.*, 2013; Zhang *et al.*, 2017b). Resolving more rust effector structures may reveal structural homologies and inform their putative functions, which largely remain undetermined.

7. Effectoromics of the coffee rust fungus *H. vastatrix*

Hundreds of candidate effectors have been identified in *H. vastatrix* by genomic and transcriptomic approaches (Fernandez *et al.*, 2012; Cristancho *et al.*, 2014; Talhinhos *et al.*, 2014). Similar to effectoromic screens conducted in *P. graminis* f. sp. *tritici*, a medium-throughput screen has been established to deliver candidates in coffee leaves through *Pseudomonas syringae* pv. *garcae* (Maia *et al.*, 2017). The effect of 30 *H. vastatrix* candidate effectors delivered into coffee leaves on the symptoms caused by *P. syringae* pv. *garcae* infection was assessed to identify avirulence candidates. HvEC-016 reduced bacterial growth and suppressed plant cell death in coffee genotypes carrying a specific resistance gene (Maia *et al.*, 2017). This elegant approach, successfully applied to a complex rust pathosystem, opens promising perspectives for the identification of avirulence genes.

8. *Uromyces* spp. effectoromics

Uromyces fabae has been a pioneer rust species for the analysis of the infection process, notably through the description of haustorially expressed genes (Hahn & Mendgen, 1997). The secretion of c. 100 *U. fabae* candidate secreted proteins has been validated using the yeast signal sequence trap system (Link & Voegelé, 2008). Most functions studied in this fungus are discussed in Section IV. A small-throughput screen has been established to assay five *Uromyces appendiculatus* proteins through *Bean pod mottle virus*-mediated HIGS. Silencing of a trehalose phosphatase, a chitinase-like protein and a glycoside hydrolase prevented *U. appendiculatus* infection, suggesting they might play an important role in fungal development and plant infection (Cooper *et al.*, 2016; Cooper & Campbell, 2017).

VII. Putting rusts to sleep: Pucciniales research outlooks

Genomics has been a fantastic lever in fostering progress in the field of rust research (Duplessis *et al.*, 2014). Despite the complexity of

rust genomes, the latest genomic reports are promising and the systematic use of long fragment sequencing techniques will be key to resolving their assembly. It remains to be determined how these methods will perform in rust fungi with expected genome sizes > 1 Gb. The complexity of very large rust genomes may very well remain an impediment to completing genomic projects. On the other hand, many reports have proved that transcriptomics is a promising alternative in the absence of a reference genome (for *P. pachyrhizi*, see Link T. *et al.*, (2014), Link T.I. *et al.*, (2014) and Qi *et al.* (2018); for *H. vastatrix*, see Fernandez *et al.* (2012) and Maia *et al.* (2017)). The large repertoires of small secreted proteins presenting typical features of effectors in rust genomes is impressive, with hundreds to thousands of candidates per species. Expression of these genes in successive waves indicates that cohorts of effectors may act in a concerted way to allow the fungus to progress, unbeknownst to the host plant, and to maintain the biotrophic state. Whether effectors act alone or in complex with other effectors within the host cell will need to be determined. The many facets of the complex life cycles of rust fungi should be more systematically explored to improve our understanding of heteroecism. The recent transcriptomic studies conducted during alternate host infection for several rust species provide a deeper understanding of rust pathogenesis. Understanding reproductive mechanisms on alternate hosts in heteroecious rust fungi of major agricultural crops is essential for developing effective and sustainable strategies for managing rust diseases.

By providing an exhaustive overview of efforts deployed towards searching for effector proteins among hundreds of candidates, our review illustrates the speed with which the field of rust research has progressed. The multiplicity of the rust pathosystems and the systematic application of cellular and molecular screens based on heterologous plant systems are examples of many high-value innovations that are paving the way towards a better understanding of the pathogenesis process. Similar efforts can be applied to any major rust fungus impacting agriculture or the environment. The same tools could be employed to identify and validate avirulence genes ultimately to inform host resistance management (Moscou & van Esse, 2017).

Efficient tools with which to unravel the precise molecular mechanisms of rust–plant interactions are still urgently needed. For years, attempts to genetically transform rust fungi have had limited success (Lawrence *et al.*, 2010; Djulic *et al.*, 2011). Gene silencing is the best proxy at the moment and it is being widely and successfully applied to *P. striiformis* f. sp. *tritici* (Panwar & Bakkeren, 2017; Tang *et al.*, 2017). Efforts are currently ongoing to achieve genetic transformation in several Pucciniales, which, at the moment, appears to be the most direct way to address critical functions during interaction with the host plant (Fig. 5). Development of standardized routines in genomics and the design of functional pipelines are also needed in the community to allow for comparative studies. The more recent focus on effector genes has, however, somewhat overshadowed other categories related to infection. If understanding how the pathogen is able to rewire the plant immune system to avoid recognition and establish infection in the host is important, it will be just as crucial to determine how it is able to reroute nutrients and sustain its growth and spore multiplication.

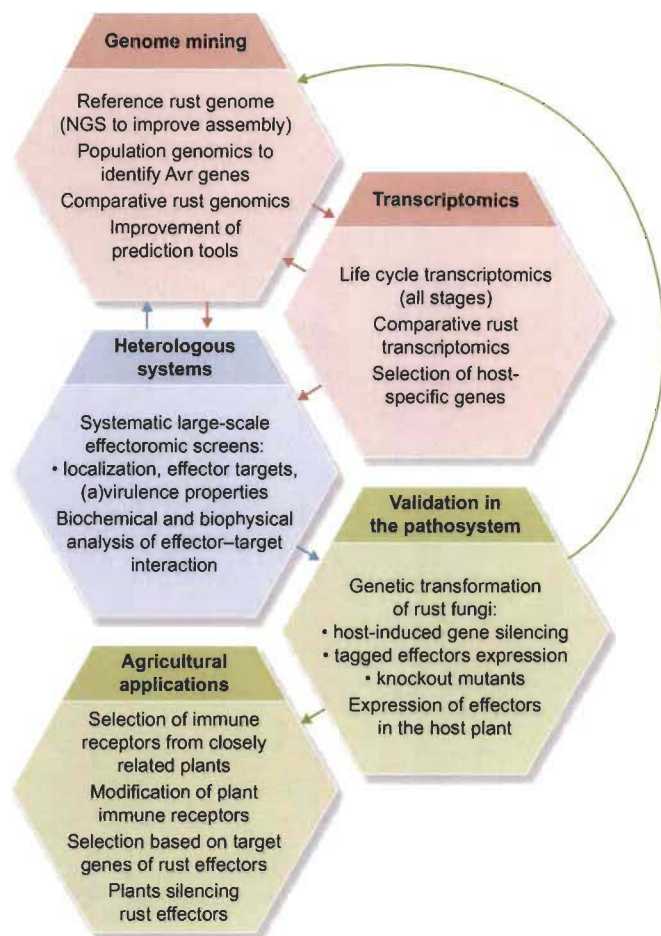


Fig. 5 Perspectives for rust fungal effector biology. Perspectives related to genomic and transcriptomic approaches are indicated in pink; and those related to applications in heterologous systems or in the pathosystem are in blue and green, respectively. Improved assembly of reference genomes, access to populational information and life cycle-informed transcriptomics will support mining for candidate effectors and avirulence gene identification in rust fungi; comparative rust genomics will enable better identification of core and lineage-specific effectors, and improvement of prediction tools will help their prioritization. Effectoromic screens pinpoint priority candidates with particular localization, plant targets and avirulence properties. The main challenge of rust effector biology is to validate candidate effectors in the pathosystem through transformation of rust fungi and their hosts. Some approaches, such as host-induced gene silencing and expression of effectors in the host plant, are readily available but so far the inability to obtain knockout mutants remains a drawback. The main output for rust effector research is translation to field disease management with selection of new resistances or less susceptible plants for future applications in agriculture. NGS, next-generation sequencing.

Progress in the field of rust effector biology and pathogenesis in rust fungi can inform the selection of new resistances in host plants and the development of novel means of controlling or containing crop diseases (Fig. 5). A number of recent reviews have specifically discussed how findings from molecular analysis of the effector mechanisms can translate into developing integrated resistance management approaches (Boch *et al.*, 2014; Moscou & van Esse, 2017; Nelson *et al.*, 2018). Although the gap between fundamental and applied research remains strong, recent rapid advances on all fronts in understanding obligate biotrophy significantly support

the fight against the fascinating plant pathogens that are the rust fungi.

Acknowledgements

The authors would like to thank Aimee Orsini for her patience during language editing. We would also like to warmly thank all colleagues who shared pictures of their favourite rust fungus in order to set Fig. 1: Brent McCallum and Guus Bakkeren at Agriculture and Agri-Food Canada; and Henriette Goyeau, Bénédicte Parriaud, Benjamin Petre and Frédéric Suffert at INRA France. CL, AH and SD are supported by the French 'Investissements d'Avenir' programme (ANR-11-LABX-0002-01, Lab of Excellence ARBRE). CL is supported by a young scientist grant (CJS) from INRA. KCGS is supported by graduate scholarships from MITACS Canada, FRQNT and UQTR. SD, AH and HG benefited from a CFQCU French-Quebec Cooperative project established between the Université de Lorraine (France) and the Université du Québec à Trois-Rivières (Canada).

ORCID

Sébastien Duplessis <https://orcid.org/0000-0002-2072-2989>
Hugo Germain <https://orcid.org/0000-0002-7046-6194>
Karen Cristine Gonçalves dos Santos <https://orcid.org/0000-0003-3270-3133>
Arnaud Hecker <https://orcid.org/0000-0003-4511-5410>
Cécile Lorrain <https://orcid.org/0000-0001-9727-2616>

References

- Aime MC, Bell CD, Wilson AW. 2018. Deconstructing the evolutionary complexity of rust fungi (Pucciniales) and their plant hosts. *Studies in Mycology* 89: 143–152.
- Aime MC, MacTaggart AR, Mondo SJ, Duplessis S. 2017. Phylogenetics and phylogenomics of rust fungi. *Advances in Fungal Genetics* 100: 267–307.
- Aime MC, Matheny PB, Henk DA, Frieders EM, Nilsson RH, Piepenbring M, McLaughlin DJ, Szabo LJ, Begerow D, Sampaio JP *et al.* 2006. An overview of the higher-level classification of Pucciniomycotina based on combined analyses of nuclear large and small subunit rDNA sequences. *Mycologia* 98: 896–905.
- Aime M, Toome M, McLaughlin D. 2014. The Pucciniomycotina. In: McLaughlin D, Spatafora JW, eds. *The Mycota VII Part A*. Berlin/Heidelberg, Germany: Springer, 271–294.
- Anderson C, Khan MA, Catanzariti A-M, Jack CA, Nemri A, Lawrence GJ, Jones DA. 2016. Genome analysis and avirulence gene cloning using a high-density RADseq linkage map of the flax rust fungus, *Melampsora lini*. *BMC Genomics* 17: 667.
- Barrett LG, Thrall PH, Dodds PN, van der Merwe M, Linde CC, Lawrence GJ, Burdon JJ. 2009. Diversity and evolution of effector loci in natural populations of the plant Pathogen *Melampsora lini*. *Molecular Biology and Evolution* 26: 2499–2513.
- Beddow JM, Pardey PG, Chai Y, Hurley TM, Kriticos DJ, Braun H-J, Yonow T. 2015. Research investment implications of shifts in the global geography of wheat stripe rust. *Nature Plants* 1: 15132.
- Boch J, Bonas U, Lahaye T. 2014. TAL effectors-pathogen strategies and plant resistance engineering. *New Phytologist* 204: 823–832.
- Bolton MD, Kolmer JA, Garvin DF. 2008. Wheat leaf rust caused by *Puccinia triticina*. *Molecular Plant Pathology* 9: 563–575.

- Bruce M, Neugebauer K, Joly D, Migeon P, Cuomo C, Wang S, Fellers J. 2014. Using transcription of six *Puccinia triticina* races to identify the effective secretome during infection of wheat. *Frontiers in Plant Science* 4: 520.
- Bueno-Sancho V, Persoons A, Hubbard A, Cabrera-Quio LE, Lewis CM, Corredor-Moreno P, Bunting DCE, Ali S, Cheng S, Hodson DP *et al.* 2017. Pathogenomic analysis of wheat yellow rust lineages detects seasonal variation and host specificity. *Genome Biology and Evolution* 9: 3282–3296.
- Cantu D, Segovia V, MacLean D, Bayles R, Chen X, Kamoun S, Dubcovsky J, Saunders DGO, Uauy C. 2013. Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics* 14: 270.
- de Carvalho MC, da CG, Nascimento LC, Darben LM, Polizel-Podanosqui AM, Lopes-Caitar VS, Qi M, Rocha CS, Carazzolle MF, Kuwahara MK *et al.* 2017. Prediction of the in planta *Phakopsora pachyrhizi* secretome and potential effector families. *Molecular Plant Pathology* 18: 363–377.
- Catanzariti AM, Dodds PN, Lawrence GJ, Ayliffe MA, Ellis JG. 2006. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* 18: 243–256.
- Chang Q, Liu J, Lin X, Hu S, Yang Y, Li D, Chen L, Huai B, Huang L, Voegelé RT *et al.* 2017. A unique invertase is important for sugar absorption of an obligate biotrophic pathogen during infection. *New Phytologist* 215: 1548–1561.
- Chen J, Upadhyaya NM, Ortiz D, Sperschneider J, Li F, Bouton C, Breen S, Dong C, Xu B, Zhang X *et al.* 2017. Loss of *AvrSr50* by somatic exchange in stem rust leads to virulence for *Sr50* resistance in wheat. *Science* 358: 1607–1610.
- Chen W, Wellings C, Chen X, Kang Z, Liu T. 2014. Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici*. *Molecular Plant Pathology* 15: 433–446.
- Cheng Y, Wang W, Yao J, Huang L, Voegelé RT, Wang X, Kang Z. 2016b. Two distinct *Ras* genes from *Puccinia striiformis* exhibit differential roles in rust pathogenicity and cell death. *Environmental Microbiology* 18: 3910–3922.
- Cheng Y, Wang X, Yao J, Voegelé RT, Zhang Y, Wang W, Huang L, Kang Z. 2015. Characterization of protein kinase PsSRPKL, a novel pathogenicity factor in the wheat stripe rust fungus. *Environmental Microbiology* 17: 2601–2617.
- Cheng Y, Wu K, Yao J, Li S, Wang X, Huang L, Kang Z. 2016a. PSTha5a23, a candidate effector from the obligate biotrophic pathogen *Puccinia striiformis* f. sp. *tritici*, is involved in plant defense suppression and rust pathogenicity. *Environmental Microbiology* 19: 1717–1729.
- Cooper B, Campbell KB. 2017. Protection against common bean rust conferred by a gene-silencing method. *Phytopathology* 107: 920–927.
- Cooper B, Campbell KB, Beard HS, Garrett WM, Islam N. 2016. Putative rust fungal effector proteins in infected bean and soybean leaves. *Phytopathology* 106: 491–499.
- Cristancho MA, Botero-Rozo DO, Giraldo W, Tabima J, Riaño-Pachón DM, Escobar C, Rozo Y, Rivera LF, Duran A, Restrepo S *et al.* 2014. Annotation of a hybrid partial genome of the coffee rust (*Hemileia vastatrix*) contributes to the gene repertoire catalog of the Pucciniales. *Frontiers in Plant Science* 5: 594.
- Cuomo CA, Bakkeren G, Khalil HB, Panwar V, Joly D, Linning R, Sakthikumar S, Song X, Adiconis X, Fan L *et al.* 2017. Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. *G3: Genes – Genomes – Genetics* 7: 361–376.
- Dagvadorj B, Ozketen AC, Andac A, Duggan C, Bozkurt TO, Akkaya MS. 2017. A *Puccinia striiformis* f. sp. *tritici* secreted protein activates plant immunity at the cell surface. *Scientific Reports* 7: 1141.
- Dean R, van Kan JAL, Pretorius ZA, Hammond-Kosack KE, di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J *et al.* 2012. The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology* 13: 414–430.
- Djulic A, Schmid A, Lenz H, Sharma P, Koch C, Wirsig SGR, Voegelé RT. 2011. Transient transformation of the obligate biotrophic rust fungus *Uromyces fabae* using biolistics. *Fungal Biology* 115: 633–642.
- Dobon A, Bunting DCE, Cabrera-Quio LE, Uauy C, Saunders DGO. 2016. The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. *BMC Genomics* 17: 380.
- Dodds PN, Lawrence GJ, Catanzariti AM, Ayliffe MA, Ellis JG. 2004. The *Melampsora lini* *Avr567* avirulence genes are expressed in haustoria and their products are recognized inside plant cell. *Plant Cell* 16: 755–768.
- Dracatos PM, Haghdoust R, Singh D, Park RF. 2018. Exploring and exploiting the boundaries of host specificity using the cereal rust and mildew models. *New Phytologist* 218: 453–462.
- Duplessis S, Bakkeren G, Hamelin R. 2014. Advancing knowledge on biology of rust fungi through genomics. *Advances in Botanic Research* 70: 173–209.
- Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel B *et al.* 2011a. Obligate biotrophy features unravelled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences, USA* 108: 9166–9171.
- Duplessis S, Hacquard S, Delaruelle C, Tisserant E, Frey P, Martin F, Kohler A. 2011b. *Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Molecular Plant–Microbe Interactions* 24: 808–818.
- Duplessis S, Joly DJ, Dodds PN. 2012. Rust effectors. In: Martin F, Kamoun S, eds. *Effectors in plant-microbes interactions*. Oxford, UK: Wiley-Blackwell, 155–193.
- Ellis JG, Dodds PN, Lawrence GJ. 2007. Flax rust resistance gene specificity is based on direct resistance-avirulence protein interactions. *Annual Review of Phytopathology* 45: 289–306.
- Ellis JG, Lagudah ES, Spielmeier W, Dodds PN. 2014. The past, present and future of breeding rust resistant wheat. *Frontiers in Plant Science* 24: 641.
- Fernandez D, Tisserant E, Talhinas P, Azinheira H, Vieira A, Petitot A-S, Loureiro A, Poulain J, Da Silva C, Silva MDC *et al.* 2012. 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. *Molecular Plant Pathology* 13: 17–37.
- Figuerola M, Hammond-Kosack KE, Solomon PS. 2017. A review of wheat diseases – a field perspective. *Molecular Plant Pathology* 19: 1523–1536.
- Figuerola M, Upadhyaya NM, Sperschneider J, Park RF, Szabo LJ, Steffenson B, Ellis GJ, Dodds PN. 2016. Changing the game: using integrative genomics to probe virulence mechanisms of the stem rust pathogen *Puccinia graminis* f. sp. *tritici*. *Frontiers in Plant Science* 7: 205.
- Fischer MC, Henck DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484: 186–194.
- Flor H. 1971. Current status of the gene-for-gene concept. *Annual Review of Phytopathology* 9: 275–296.
- Garnica DP, Nemri A, Upadhyaya NM, Rathjen JP, Dodds PN. 2014. The ins and outs of rust haustoria. *PLoS Pathogens* 10: e1004329.
- Garnica DP, Upadhyaya NM, Dodds PN, Rathjen JP. 2013. Strategies for wheat stripe rust pathogenicity identified by transcriptome sequencing. *PLoS ONE* 8: e67150.
- Germain H, Joly DL, Mireault C, Plourde MB, Leranneur C, Stewart D, Morency MJ, Petre B, Duplessis S, Seguin A. 2017. Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Molecular Plant Pathology* 19: 191–200.
- Godoy CV, Seixas CDS, Soares RM, Marcelino-Guimaraes FC, Meyer MC, Costamilan LM. 2016. Asian soybean rust in Brazil: past, present and future. *Pesquisa Agropecuária Brasileira* 5: 407–421.
- Goellner K, Loehrer M, Langenbach C, Conrath U, Koch E, Schaffrath U. 2010. *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. *Molecular Plant Pathology* 11: 169–177.
- Guo J, Dai X, Xu J-R, Wang Y, Bai P, Liu F, Duan Y, Zhang H, Huang L, Kang Z. 2011. Molecular characterization of a Fus3/Kss1 Type MAPK from *Puccinia striiformis* f. sp. *tritici*, PsMAPK1. *PLoS ONE* 6: e21895.
- Hacquard S, Delaruelle C, Frey P, Tisserant E, Kohler A, Duplessis S. 2013. Transcriptome analysis of poplar rust telia reveals overwintering adaptation and tightly coordinated karyogamy and meiosis processes. *Frontiers in Plant Science* 4: 456.
- Hacquard S, Delaruelle C, Legue V, Tisserant E, Kohler A, Frey P, Martin F, Duplessis S. 2010. Laser capture microdissection of uredinia formed by *Melampsora larici-populina* revealed a transcriptional switch between biotrophy and sporulation. *Molecular Plant–Microbe Interactions* 23: 1275–1286.
- Hacquard S, Joly DL, Lin Y-C, Tisserant E, Feau N, Delaruelle C, Legue V, Kohler A, Tanguay P, Petre B *et al.* 2012. A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Molecular Plant–Microbe Interactions* 25: 279–293.

- Hacquard S, Petre B, Frey P, Hecker A, Rouhier N, Duplessis S. 2011. The poplar-poplar rust interaction: insights from genomics and transcriptomics. *Journal of Pathogens* 2011: 716041.
- Hahn M. 2000. The rust fungi: cytology, physiology and molecular biology of infection. In: Kronstad J, ed. *Fungal pathology*. Dordrecht, the Netherlands: Kluwer Academic Publishers, 267–306.
- Hahn M, Mendgen K. 1997. Characterization of in plant-induced rust genes isolated from a haustorium-specific cDNA library. *Molecular Plant-Microbe Interactions* 10: 427–437.
- Hahn M, Neef U, Struck C, Göttfert M, Mendgen K. 1997. A putative amino acid transporter is specifically expressed in haustoria of the rust fungus *Uromyces fabae*. *Molecular Plant-Microbe Interactions* 10: 438–445.
- Harder DE. 1984. Developmental ultrastructure of hyphae and spores. In: Bushnell W, Roelfs AP, eds. *The cereal rusts*. Cambridge, MA, USA: Academic Press, 333–373.
- Hu GG, Linning R, McCallum B, Banks T, Cloutier S, Butterfield Y, Liu J, Kirkpatrick R, Stott J, Yang G *et al.* 2007. Generation of a wheat leaf rust, *Puccinia triticina*, EST database from stage-specific cDNA libraries. *Molecular Plant Pathology* 8: 451–467.
- Huang X, Chen X, Coram T, Wang M, Kang Z. 2011. Gene expression profiling of *Puccinia striiformis* f. sp. *tritici* during development reveals a highly dynamic transcriptome. *Journal of Genetics and Genomics* 38: 357–371.
- Jiao M, Yu D, Tan C, Guo J, Lan D, Han E, Guo J, Lan D, Han E, Qi T *et al.* 2017. Basidiomycete-specific *PtCaMKL1* encoding a CaMK-like protein kinase is required for full virulence of *Puccinia striiformis* f. sp. *tritici*. *Environmental Microbiology* 19: 4177–4189.
- Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444: 323–329.
- Kemen AC, Alger MT, Kemen E. 2015. Host-microbe and microbe-microbe interactions in the evolution of obligate plant parasitism. *New Phytologist* 206: 1207–1228.
- Kemen E, Jones JDG. 2012. Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends in Plant Science* 17: 448–457.
- Kemen E, Kemen A, Ehlers A, Voegelé R, Mendgen K. 2013. A novel structural effector from rust fungi is capable of fibril formation. *The Plant Journal* 75: 767–780.
- Kemen E, Kemen AC, Rafiqi M, Hempel U, Mendgen K, Hahn M, Voegelé RT. 2005. Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Molecular Plant-Microbe Interactions* 18: 1130–1139.
- Kunjeti SG, Iyer G, Johnson E, Li E, Broglie KE, Rauscher G, Raidan GJ. 2016. Identification of *Phakopsora pachyrhizi* candidate effectors with virulence activity in a distantly related pathosystem. *Frontiers in Plant Science* 7: 269.
- Lanver D, Müller AN, Happel P, Schweizer G, Haas FB, Franitza M, Pellegrin C, Reissmann S, Altmüller J, Rensing SA *et al.* 2018. The biotrophic development of *Ustilago maydis* studied by RNAseq analysis. *Plant Cell* 30: 300–323.
- Lawrence GJ, Dodds PN, Ellis JG. 2007. Rust of flax and linseed caused by *Melampsora lini*. *Molecular Plant Pathology* 8: 349–364.
- Lawrence GJ, Dodds PN, Ellis JG. 2010. Transformation of the flax rust fungus, *Melampsora lini*: selection via silencing of an avirulence gene. *The Plant Journal* 61: 364–369.
- Link TI, Lang P, Scheffler BE, Duke MV, Graham MA, Cooper B, Tucker ML, van de Mortel M, Voegelé RT, Mendgen K *et al.* 2014. The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakopsora pachyrhizi* and their candidate effector families. *Molecular Plant Pathology* 15: 379–393.
- Link T, Lohaus G, Heiser I, Mendgen K, Hahn M, Voegelé RT. 2005. Characterization of a novel NADP⁺-dependent D-arabitol dehydrogenase from the plant pathogen *Uromyces fabae*. *Biochemical Journal* 389: 289–295.
- Link T, Seibel C, Voegelé RT. 2014. Early insights into the genome sequence of *Uromyces fabae*. *Frontiers in Plant Science* 5: 587.
- Link TI, Voegelé RT. 2008. Secreted proteins of *Uromyces fabae*: similarities and stage specificity. *Molecular Plant Pathology* 9: 59–66.
- Littlefield LJ, Heath MC. 1979. *Ultrastructure of rust fungi*. New York, NY, USA: Academic Press.
- Liu C, Pedersen C, Schultz-Larsen T, Aguilar GB, Madriz-Ordeñana K, Hovmøller MS, Thordal-Christensen H. 2016. The stripe rust fungal effector PEC6 suppresses pattern-triggered immunity in a host species-independent manner and interacts with adenosine kinases. *New Phytologist*. doi:10.1111/nph.14034.
- Liu J, Guan T, Zheng P, Chen L, Yang Y, Huai B, Li D, Chang Q, Huang L, Kang Z. 2016. An extracellular Zn-only superoxide dismutase from *Puccinia striiformis* confers enhanced resistance to host-derived oxidative stress. *Environmental Microbiology* 18: 4118–4135.
- Liu J, Wang Q-L, Chang Q, Han L-N, Pei G-L, Xue Y-Q, Jia L-M, Zhang K, Duan Y-Y, Kang Z. 2014. Isocitrate lyase is required for urediniospore germination of *Puccinia striiformis* f. sp. *tritici*. *Molecular Biology Reports* 41: 7797–7806.
- Liu JJ, Sturrock RN, Snieszko RA, Williams H, Benton R, Zamany A. 2015. Transcriptome analysis of the white pine blister rust pathogen *Cronartium ribicola*: de novo assembly, expression profiling, and identification of candidate effectors. *BMC Genomics* 16: 678.
- Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. 2015. Fungal effectors and plant susceptibility. *Annual Review of Plant Biology* 66: 513–545.
- Loehrer M, Botterweck J, Jahnke J, Mahlmann DM, Gaertgens J, Oldiges M, Horbach R, Deising H, Schaffrath U. 2014a. In vivo assessment by Mach-Zehnder double-beam interferometry of the invasive force exerted by the Asian soybean rust fungus (*Phakopsora pachyrhizi*). *New Phytologist* 203: 620–631.
- Loehrer M, Vogel A, Huettel B, Reinhardt R, Benes V, Duplessis S, Usadel B, Schaffrath U. 2014b. On the current status of *Phakopsora pachyrhizi* genome sequencing. *Frontiers in Plant Science* 5: 377.
- Lohaus G, Pennewiss K, Sattelmacher B, Hussmann M, Hermann Muehling K. 2001. Is the infiltration-centrifugation technique appropriate for the isolation of apoplastic fluid? A critical evaluation with different plant species. *Physiologia Plantarum* 111: 457–465.
- Lorrain C, Hecker A, Duplessis S. 2015. Effector-mining in the poplar rust fungus *Melampsora larici-populina* secretome. *Frontiers in Plant Science* 6: 1051.
- Lorrain C, Marchal C, Hacquard S, Delaruelle C, Petrowski J, Petre B, Hecker A, Frey P, Duplessis S. 2018a. The rust fungus *Melampsora larici-populina* expresses a conserved genetic program and distinct sets of secreted protein genes during infection of its two host plants, larch and poplar. *Molecular Plant-Microbe Interactions* 31: 695–706.
- Lorrain C, Petre B, Duplessis S. 2018b. Show me the way: rust effector targets in heterologous plant systems. *Current Opinion in Microbiology* 14: 19–25.
- Maia T, Badel JL, Marin-Ramirez G, Rocha CM, Fernandes MB, da Silva JCF, de Azevedo-Junior G-H, Brommonschenkel SH. 2017. The *Hemileia vastatrix* effector HvEC-016 suppresses bacterial blight symptoms in coffee genotypes with the SH 1 rust resistance gene. *New Phytologist* 213: 1315–1329.
- McTaggart AR, Shivas RG, van der Nest MA, Roux J, Wingfield BD, Wingfield MJ. 2016. Host jumps shaped the diversity of extant rust fungi (Pucciniales). *New Phytologist* 209: 1149–1158.
- Miller ME, Zhang Y, Omidvar V, Sperschneider J, Schwessinger B, Raley C, Raley C, Palmer JM, Garnica D, Upadhyaya N *et al.* 2017. De novo assembly and phasing of dikaryotic genomes from two isolates of *Puccinia coronata* f. sp. *avenae*, the causal agent of oat crown rust. *mBio* 9: e01650–17.
- Moller M, Strukenbrock EH. 2017. Evolution and genome architecture in fungal plant pathogens. *Nature Reviews. Microbiology* 15: 756–771.
- Moscou M, van Esse P. 2017. The quest for durable resistance. *Science* 358: 1541–1542.
- Nelson R, Wiesner-Hanks T, Wissner R, Balint-Kurti P. 2018. Navigating complexity to breed disease-resistant crops. *Nature Genetics Reviews* 19: 21–33.
- Nemri A, Saunders DGO, Anderson C, Upadhyaya NM, Win J, Lawrence GJ, Jones DA, Kamoun S, Ellis JG, Dodds PN. 2014. The genome sequence and effector complement of the flax rust pathogen *Melampsora lini*. *Frontiers in Plant Science* 5: 98.
- Panwar V, Bakkeren G. 2018. Investigating gene function in cereal rust fungi by plant-mediated virus-induced gene silencing. *Methods in Molecular Biology* 1659: 115–124.
- Panwar V, McCallum B, Bakkeren G. 2013. Endogenous silencing of *Puccinia triticina* pathogenicity genes through in planta-expressed sequences leads to the suppression of rust diseases on wheat. *The Plant Journal* 73: 521–532.
- Pendleton AL, Smith KE, Feau N, Martin F, Grigoriev IV, Hamelin R, Nelson CD, Burleigh JG, Davis JM. 2014. Duplications and losses in gene families of rust pathogens highlight putative effectors. *Frontiers in Plant Science* 5: 299.
- Pernaci M, De Mita S, Andrieux A, Petrowski J, Halkett F, Duplessis S, Frey P. 2014. Genome-wide patterns of segregation and linkage disequilibrium: the

- construction of a linkage genetic map of the poplar rust fungus *Melampsora larici-populina*. *Frontiers in Plant Science* 5: 454.
- Persoons A, Morin E, Delaruelle C, Payen T, Halkett F, Frey P, De Mita S, Duplessis S. 2014. Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors. *Frontiers in Plant Science* 5: 450.
- Petre B, Joly DL, Duplessis S. 2014. Effector proteins of rust fungi. *Front in Plant Science* 5: 146.
- Petre B, Kamoun S. 2014. How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biology* 12: e1001801.
- Petre B, Lorrain C, Saunders DGO, Win J, Sklenar J, Duplessis S, Kamoun S. 2016b. Rust fungal effectors mimic host transit peptides to translocate into chloroplasts. *Cellular Microbiology* 18: 453–465.
- Petre B, Saunders DGO, Sklenar J, Lorrain C, Krasileva KV, Win J, Duplessis S, Kamoun S. 2016a. Heterologous expression screens in *Nicotiana benthamiana* identify a candidate effector of the wheat yellow rust pathogen that associates with processing bodies. *PLoS ONE* 11: e0149035.
- Petre B, Saunders DGO, Sklenar J, Lorrain C, Win J, Duplessis S, Kamoun S. 2015. Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. *Molecular Plant–Microbe Interactions* 28: 689–700.
- Pretsch K, Kemen A, Kemen E, Geiger M, Mendgen K, Voegele R. 2013. The rust transferred proteins – a new family of effector proteins exhibiting protease inhibitor function. *Molecular Plant Pathology* 14: 96–107.
- Qi M, Grayczyk JP, Seitz JM, Lee Y, Link TI, Choi D, Pedley K, Voegele R, Baum T, Whitham SA. 2018. Suppression or activation of immune responses by predicted secreted proteins of the soybean rust pathogen *Phakopsora pachyrhizi*. *Molecular Plant–Microbe Interactions* 31: 163–174.
- Qi M, Link TI, Muller M, Hirschburger D, Pudake RN, Pedley KF, Braun E, Voegele RT, Baum TJ, Whitham SA. 2016. A small cysteine-rich protein from the asian soybean rust fungus, *Phakopsora pachyrhizi*, suppresses plant immunity. *PLoS Pathogens* 12: e1005827.
- Qi T, Zhu X, Tan C, Liu P, Guo J, Kang Z, Guo J. 2018. Host-induced gene silencing of an important pathogenicity factor PsCPK1 in *Puccinia striiformis* f. sp. *tritici* enhances resistance of wheat to stripe rust. *Plant Biotechnology Journal* 16: 797–807.
- Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* 10: 417.
- Rafiqi M, Gan PHP, Ravensdale M, Lawrence GJ, Ellis JG, Jones DA, Hardham AR, Dodds PN. 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.
- Ramachandran SR, Yin C, Kud J, Tanaka K, Mahoney AK, Xiao F, Hulbert SH. 2017. Effectors from wheat rust fungi suppress multiple plant defense responses. *Phytopathology* 107: 75–83.
- Ramos AP, Tavares S, Tavares D, Silva MDC, Loureiro J, Talhinas P. 2015. Flow cytometry reveals that the rust fungus, *Uromyces bidentis* (Pucciniales), possesses the largest fungal genome reported – 2489Mbp. *Molecular Plant Pathology* 16: 1006–1010.
- Rutter WB, Salcedo A, Akhunova A, He F, Wang S, Liang H, Bowden RL, Akhunov E. 2017. Divergent and convergent modes of interaction between wheat and *Puccinia graminis* f. sp. *tritici* isolates revealed by the comparative gene co-expression network and genome analyses. *BMC Genomics* 18: 291.
- Ryder LS, Talbot NJ. 2015. Regulation of appressorium development in pathogenic fungi. *Current Opinion in Plant Biology* 26: 8–13.
- Salcedo A, Rutter W, Wang S, Akhunova A, Bolus S, Chao S, Anderson N, De Soto MF, Rouse M, Szabo L *et al.* 2017. Variation in the *AvrSr35* gene determines *Sr35* resistance against wheat stem rust race Ug99. *Science* 358: 1604–1606.
- Saunders DGO, Win J, Cano LM, Szabo LJ, Kamoun S, Raffaele S. 2012. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS ONE* 7: e29847.
- Schwessinger B, Sperschneider J, Cuddy W, Miller M, Garnica D, Taylor J, Dodds PN, Figueroa M, Park R, Rathjen J. 2017. A near complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* reveals high inter-haplome diversity. *mBio* 9: e02275–17.
- Segovia V, Bruce M, Shoup Rupp JL, Huang L, Bakkeren G, Trick HN, Fellers JP. 2016. Two small secreted proteins from *Puccinia trititica* induce reduction of β -glucuronidase transient expression in wheat isolines containing *Lr9*, *Lr24* and *Lr26*. *Canadian Journal of Plant Pathology* 38: 91–106.
- Singh RP, Hodson DP, Huerta-Espino J, Jin Y, Bhavani S, Njau P, Herrera-Foessel S, Singh PK, Singh S, Govindan V. 2011. Races of the wheat stem rust fungus is a threat to world wheat production. *Annual Review Phytopathology* 49: 465–481.
- Singh RP, Hodson DP, Jin Y, Lagudah ES, Ayliffe MA, Bhavani S, Rouse BM, Pretorius ZA, Szabo LJ, Huerta-Espino J *et al.* 2015. Emergence and spread of new races of wheat stem rust fungus: continued threat to food security and prospects of genetic control. *Phytopathology* 105: 872–884.
- Slaminko TL, Miles MR, Frederick RD, Bonde MR, Hartman GL. 2008. New legume hosts of *Phakopsora pachyrhizi* based on greenhouse evaluations. *Plant Disease* 92: 767–771.
- Snieszko RA, Yanchuk AD, Kliejunas JT, Palmieri KM, Alexander JM, Frankel SJ. 2012. *Proceedings of the fourth international workshop on the genetics of host-parasite interactions in forestry: disease and insect resistance in forest trees*. General Technical Report PSW-GTR-240. Albany, CA, USA: US Department of Agriculture, Forest Service, Pacific Southwest Research Station. 372p.
- Spanu PD. 2012. The genomics of obligate (and non-obligate) biotrophs. *Annual Review Phytopathology* 50: 91–109.
- Sperschneider J, Catanzariti AM, DeBoer K, Petre B, Gardiner DM, Singh KB, Dodds PN, Taylor JM. 2017b. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Report* 7: 44598.
- Sperschneider J, Dodds PN, Gardiner DM, Mannes JM, Singh KB, Taylor JM. 2015. Advances and challenges in computational prediction of effectors from plant pathogenic fungi. *PLoS Pathogens* 11: e1004806.
- Sperschneider J, Dodds P, Gardiner D, Singh K, Taylor J. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology* 19: 2094–2110.
- Sperschneider J, Dodds PN, Singh KB, Taylor JM. 2017c. ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *New Phytologist* 217: 1764–1778.
- Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Mannes JM, Taylor JM. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytologist* 210: 743–761.
- Sperschneider J, Taylor J, Dodds PN, Duplessis S. 2017a. Computational methods for predicting effectors in rust pathogens. *Wheat Rust Disease – Methods and Protocols* 1659: 73–83.
- Staples RC. 2000. Research on the rust fungi during the twentieth century. *Annual Review of Phytopathology* 38: 49–69.
- Stone CL, McMahon MB, Fortis Nuñez A, Smythers GW, Luster DG, Frederick RD. 2012. Gene expression and proteomic analysis of the formation of *Phakopsora pachyrhizi* appressoria. *BMC Genomics* 13: 269.
- Struck C. 2015. Amino acid uptake in rust fungi. *Frontiers in Plant Science* 6: 40.
- Struck C, Hahn M, Mendgen K. 1996. Plasma membrane H^+ -ATPase activity in spores, germ tubes, and haustoria of the rust fungus *Uromyces viciae-fabae*. *Fungal Genetics and Biology* 20: 30–35.
- Struck C, Siebels C, Rommel O, Wernitz M, Hahn M. 1998. The plasma membrane H^+ -ATPase from the biotrophic rust fungus *Uromyces fabae*: molecular characterization of the gene (*PMAl*) and functional expression of the enzyme in yeast. *Molecular Plant–Microbe Interactions* 11: 458–465.
- Stukenbrock EH. 2013. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytologist* 199: 895–907.
- Talhinas P, Azinheira HG, Vieira B, Loureiro A, Tavares S, Batista D, Morin E, Petitot AS, Paulo OS, Poulain J *et al.* 2014. Overview of the functional virulent genome of the coffee leaf rust pathogen *Hemileia vastatrix* with an emphasis on early stages of infection. *Frontiers in Plant Science* 5: 88.
- Talhinas P, Batista D, Diniz I, Vieira A, Silva DN, Loureiro A, Tavares S, Pereira A-P, Azinheira H-G, Guerra-Guimaraes L *et al.* 2017. The coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. *Molecular Plant Pathology* 18: 1039–1051.
- Tang CL, Wang XJ, Cheng YL, Liu MJ, Zhao MX, Wei JP, Kang ZS. 2015a. New insights in the battle between wheat and *Puccinia striiformis*. *Frontiers in Agricultural Science and Engineering* 2: 101–114.

- Tang C, Wei J, Han Q, Liu R, Duan X, Fu Y, Huang X, Wang X, Kang Z. 2015b. PsANT, the adenine nucleotide translocase of *Puccinia striiformis*, promotes cell death and fungal growth. *Scientific Reports* 5: 11241.
- Tang C, Xu Q, Zhao M, Wang X, Kang Z. 2017. Understanding the lifestyles and pathogenicity mechanisms of obligate biotrophic fungi in wheat: the emerging genomics era. *Crop Journal* 6: 60–67.
- Tao S-Q, Cao B, Tian C-M, Liang Y-M. 2017. Comparative transcriptome analysis and identification of candidate effectors in two related rust species (*Gymnosporangium yamadai* and *Gymnosporangium asiaticum*). *BMC Genomics* 18: 651.
- Tavares S, Ramos AP, Pires AS, Azinheira HG, Caldeirinha P, Link T, Abranches R, Silva MDC, Voegelé RT, Loureiro J *et al.* 2014. Genome size analyses of Pucciniales reveal the largest fungal genomes. *Frontiers in Plant Science* 5: 422.
- Tremblay A, Hosseini P, Li S, Alkharouf NW, Matthews BF. 2013. Analysis of *Phakopsora pachyrhizi* transcript abundance in critical pathways at four time-points during infection of a susceptible soybean cultivar using deep sequencing. *BMC Genomics* 14: 614.
- Upadhyaya NM, Mago R, Staskawicz BJ, Ayliffe MA, Ellis JG, Dodds PN. 2014. A bacterial type III secretion assay for delivery of fungal effector proteins into wheat. *Molecular Plant-Microbe Interactions* 27: 255–264.
- Ve T, Williams SJ, Catanzariti A-M, Rafiqi M, Rahman M, Ellis JG, Hardham AR, Jones DA, Anderson PA, Dodds PN *et al.* 2013. Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. *Proceedings of the National Academy of Sciences, USA* 110: 17594–17599.
- Voegelé RT, Hahn M, Lohaus G, Link T, Heiser I, Mendgen K. 2005. possible roles for mannitol and mannitol dehydrogenase in the biotrophic plant pathogen *Uromyces fabae*. *Plant Physiology* 137: 190–198.
- Voegelé RT, Hahn M, Mendgen K. 2009. The Uredinales: cytology, biochemistry, and molecular biology. In: Deising HB, eds. *Plant relationships. The Mycota (A comprehensive treatise on fungi as experimental systems for basic and applied research)*. Berlin, Germany: Springer, 69–98.
- Voegelé RT, Mendgen K. 2003. Rust haustorium: nutrient uptake and beyond. *New Phytologist* 159: 93–100.
- Voegelé RT, Mendgen K. 2011. Nutrient uptake in rust fungi: how sweet is parasitic life? *Euphytica* 179: 41–55.
- Voegelé RT, Struck C, Hahn M, Mendgen K. 2001. The role of haustoria in sugar supply during infection of broad bean by the rust fungus *Uromyces fabae*. *Proceedings of the National Academy of Sciences, USA* 98: 8133–8138.
- Voegelé RT, Wirsal S, Moll U, Lechner M, Mendgen K. 2006. Cloning and characterization of a novel invertase from the obligate biotroph *Uromyces fabae* and analysis of expression patterns of host and pathogen invertases in the course of infection. *Molecular Plant-Microbe Interactions* 19: 625–634.
- Wang CIA, Guncar G, Forwood JK, Teh T, Catanzariti AM, Lawrence GJ, Loughlin FE, Mackay JP, Schirra HJ, Anderson PA *et al.* 2007. 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.
- Wang X, Yang B, Li K, Kang Z, Cantu D, Dubcovsky J. 2016. A conserved *Puccinia striiformis* protein interacts with wheat NPR1 and reduces induction of pathogenesis-related genes in response to pathogens. *Molecular Plant-Microbe Interactions* 29: 977–989.
- Weber H, Roitsch T. 2000. Invertases and life beyond sucrose cleavage. *Trends in Plant Science* 5: 47–48.
- Win J, Chaparro-Garcia A, Belhaj K, Saunders DG, Yoshida K, Dong S, Schornack S, Zipfel C, Robatzek S, Hogenhout SA *et al.* 2012. Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harbor Symposia on Quantitative Biology* 77: 235–247.
- Wu C-H, Derevnina L, Kamoun S. 2018. Receptor networks underpin plant immunity. *Science* 360: 1300–1301.
- Wu J, Wang Q, Liu S, Huang S, Mu J, Zeng Q, Huang L, Han D, Kang Z. 2017. Saturation mapping of a major effect QTL for stripe rust resistance on wheat chromosome 2B in cultivar Napo 63 using SNP genotyping arrays. *Frontiers in Plant Science* 8: 653.
- Xu J, Linning R, Fellers J, Dickinson M, Zhu W, Antonov I, Joly DL, Donaldson ME, Eilam T, Anikster Y *et al.* 2011. Gene discovery in EST sequences from the wheat leaf rust fungus *Puccinia tritica* sexual spores, asexual spores and haustoria, compared to other rust and corn smut fungi. *BMC Genomics* 12: 161.
- Yin C, Downey SI, Klages-Mundt NL, Ramachandran S, Chen X, Szabo LJ, Kigawa T, Kamoun S, Hulbert SH. 2015. Identification of promising host-induced silencing targets among genes preferentially transcribed in haustoria of *Puccinia*. *BMC Genomics* 16: 579.
- Yin C, Jurgenson JE, Hulbert SH. 2011. Development of a host-induced RNAi system in the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici*. *Molecular Plant-Microbe Interactions* 24: 554–561.
- Yin C, Park J-J, Gang DR, Hulbert SH. 2014. Characterization of a *tryptophan 2-monoxygenase* gene from *Puccinia graminis* f. sp. *tritici* involved in auxin biosynthesis and rust pathogenicity. *Molecular Plant-Microbe Interactions* 27: 227–235.
- Yuan C, Wang M, Skinner DZ, See DR, Xia C, Guo X, Chen X. 2018. Inheritance of virulence, construction of a linkage map, and mapping dominant virulence genes in *Puccinia striiformis* f. sp. *tritici* through characterization of a sexual population with genotyping-by-sequencing. *Phytopathology* 108: 133–141.
- Zhang X, Farah N, Rolston L, Ericsson DJ, Catanzariti AM, Bernoux M, Ve T, Bendak K, Chen C, Mackay JP *et al.* 2017b. Crystal structure of the *Melampsora lini* effector AvrP reveals insights into a possible nuclear function and recognition by the flax disease resistance protein. *Molecular Plant Pathology* 19: 1196–1209.
- Zhang X, Nguyen N, Breen S, Outram MA, Dodds PN, Kobe B, Solomon PS, Williams SJ. 2017a. Production of small cysteine-rich effector proteins in *Escherichia coli* for structural and functional studies. *Molecular Plant Pathology* 18: 141–151.
- Zheng W, Huang L, Huang J, Wang X, Chen X, Zhao J, Guo J, Zhuang H, Qiu C, Liu J *et al.* 2013. High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nature Communications* 4: 2673.
- Zhu X, Liu W, Chu X, Sun Q, Tan C, Yang Q, Jiao M, Guo J, Kang Z. 2017a. The transcription factor PstSTE12 is required for virulence of *Puccinia striiformis* f. sp. *tritici*. *Molecular Plant Pathology* 19: 961–974.
- Zhu X, Qi T, Yang Q, He F, Tan C, Ma W, Voegelé RT, Kang Z, Guo J. 2017b. Host-induced gene silencing of the mapkk gene *PsFUZ7* confers stable resistance to wheat stripe rust. *Plant Physiology* 175: 1853–1863.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article:

Table S1 Effector-mining in rust fungal genomic resources.

Table S2 Prediction and analysis tools dedicated to effector-mining in fungal genomes.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

ANNEX B

A RUST FUNGAL EFFECTOR BINDS PLANT DNA AND MODULATES TRANSCRIPTION

Md Bulbul Ahmed, Karen Cristine Gonçalves dos Santos, Ingrid Berenice Sanchez,
Benjamin Petre, Cécile Lorrain, Mélodie B. Plourde, Sébastien Duplessis,
Isabel Desgagné-Penix, Hugo Germain

Scientific Reports (2018), 8: 14718. doi:10.1038/s41598-018-32825-0

SCIENTIFIC REPORTS

OPEN

A rust fungal effector binds plant DNA and modulates transcription

Md Bulbul Ahmed^{1,2}, Karen Cristine Gonçalves dos Santos^{1,2}, Ingrid Benerice Sanchez^{1,2,3}, Benjamin Petre^{4,5,6}, Cécile Lorrain⁵, Mélodie B. Plourde^{1,2}, Sébastien Duplessis⁵, Isabel Desagné-Penix^{1,2} & Hugo Germain^{1,2}

Received: 2 March 2018

Accepted: 21 June 2018

Published online: 03 October 2018

The basidiomycete *Melampsora larici-populina* causes poplar rust disease by invading leaf tissues and secreting effector proteins through specialized infection structures known as haustoria. The mechanisms by which rust effectors promote pathogen virulence are poorly understood. The present study characterized Mlp124478, a candidate effector of *M. larici-populina*. We used the models *Arabidopsis thaliana* and *Nicotiana benthamiana* to investigate the function of Mlp124478 in plant cells. We established that Mlp124478 accumulates in the nucleus and nucleolus, however its nucleolar accumulation is not required to promote growth of the oomycete pathogen *Hyaloperonospora arabidopsidis*. Stable constitutive expression of Mlp124478 in *A. thaliana* repressed the expression of genes involved in immune responses, and also altered leaf morphology by increasing the waviness of rosette leaves. Chip-PCR experiments showed that Mlp124478 associates with the TGA1a-binding DNA sequence. Our results suggest that Mlp124478 exerts a virulence activity and binds the TGA1a promoter to suppress genes induced in response to pathogen infection.

Plant pathogens secrete molecules, known as effectors, into host tissues to promote parasitic growth. Effectors target various host cell compartments and interact with molecules, such as proteins and DNA, to modulate their location, stability and function^{1–4}. Nowadays, molecular plant pathologists employ effectors as probes to identify and understand the plant processes targeted by pathogens and exploit this insight to develop resistant crops. Genomic approaches coupled with heterologous expression studies in *Arabidopsis thaliana* and *Nicotiana benthamiana* are commonly undertaken to decipher the mechanisms by which effectors promote pathogen virulence^{5–9}.

Many effectors interfere with transcription to alter plant immune responses^{10–12}. For instance, bacterial transcription activator-like effectors (TAL) function as transcription factors and alter host gene expression levels, which may result in substantial influence on host phenotypes^{13,14}. The oomycete *Hyaloperonospora arabidopsidis*, a filamentous obligate biotrophic pathogen, has effectors that target the nucleus. One of them, HaRxL44, accumulates in the nucleus and interacts with the Mediator complex MED19a, inducing its proteasome-mediated degradation. This, in turn, leads to transcriptional changes resembling jasmonic acid and ethylene induction with repressed salicylic acid signaling enhancing susceptibility to biotrophs¹⁵. Similarly, global expression profiling of the fungal biotroph *Ustilago maydis*-maize interaction demonstrated early induction of the defense response genes which are later quenched¹⁶, indicating that host transcriptional reprogramming is a conserved mechanism amongst biotrophs.

Rust fungi (order *Pucciniales*) are notorious plant pathogens and are among the most studied obligate biotrophic fungal pathogens^{17,18}. *Melampsora larici-populina* causes poplar leaf rust disease, which threatens poplar plantations worldwide¹⁹. Genome analysis of *M. larici-populina* has predicted 1,184 small secreted proteins (SSPs)²⁰. Several features, such as expression in poplar leaves during infection, homology to other known

¹Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières (UQTR), Trois-Rivières, QC, G9A 5H7, Canada. ²Groupe de recherche en biologie végétale, UQTR, Trois-Rivières, QC, G9A 5H7, Canada.

³Department of Biotechnology and Engineering in Chemistry, Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Estado de México (ITESM CEM), Margarita Maza de Juárez, 52926, Cd, López Mateos, Mexico.

⁴The Sainsbury Laboratory, Norwich Research Park, Norwich, NR4 7UH, UK. ⁵INRA, UMR 1136 Interactions Arbres/Microorganismes, INRA/Université de Lorraine, Centre INRA Grand Est - Nancy, 54280, Champenoux, France.

⁶Université de Lorraine, UMR 1136 Interactions Arbres/Microorganismes, INRA/Université de Lorraine, Faculté des Sciences et Technologies - Campus Aiguillettes, BP, 70239–54506, Vandœuvre-lès-Nancy, France. Correspondence and requests for materials should be addressed to M.B.P. (email: Melodie.Bplourde@uqtr.ca) or H.G. (email: hugo.germain@uqtr.ca)

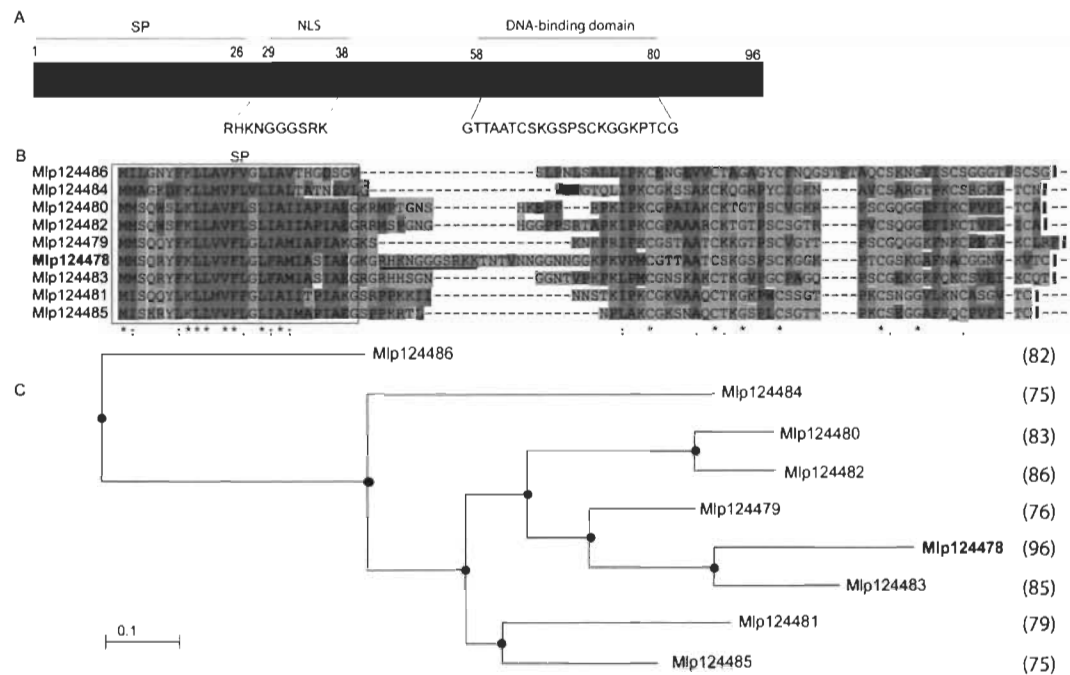


Figure 1. Sequence alignment and phylogenetic tree of the *M. larici-populina* CPG2811 SSP family. (A) Schematic representation of Mlp124478 protein topology, signal peptide (SP), nuclear localization sequence (NLS) and DNA-binding domain are shown. (B) Multiple sequence alignment of the nine members of the *M. larici-populina* CPG2811 SSP family. Predicted Signal peptides (SP) are boxed. Identical/highly conserved residues (*); semi conserved residues (:); and conserved residues (.) are marked. Predicted nuclear localization signal (NLS) is indicated by solid black underline. (C) Phylogenetic tree of the nine members of the CPG2811 gene family obtained with COBALT using Kimura distance value and neighbor joining tree method. Amino acid length is indicated in parenthesis.

rust effectors, signature of positive selection, specificity to Pucciniales order, and lack of a predicted function, were considered to select candidate secreted effector proteins (CSEPs)^{21,22}. Recently, twenty *M. larici-populina* candidate effectors were shown to accumulate in multiple leaf cell compartments and target several protein complexes when expressed heterologously in *N. benthamiana*²³. Of the CSEPs analyzed by Petre *et al.* (2015) and Germain *et al.* 2018²³, Mlp124478 is the only one to localize to the nucleus and nucleolus both in *N. benthamiana* and *Arabidopsis*. Mlp124478 is part of a gene family of nine members (CPG2811), which are specific to the order Pucciniales (Hacquard *et al.*²¹). Mlp124478 expression is strongly induced during infection and reaches 50-fold induction at 96 h after infection. Given the kinetics of *M. larici-populina* infection, this corresponds to the biotrophic growth stage in mesophyll cells²⁴. In addition, the CPG2811 group presents a signature of rapid evolution, a feature of pathogen effector families²¹. These different features observed for Mlp124478 prompted us to investigate its functional role more precisely.

Here, we confirmed the localization of Mlp124478 in epithelial cells of *A. thaliana*, we identify the sequence responsible for the nucleolar accumulation and investigate the effector cellular function *in planta*. Since the constitutive expression of Mlp124478 in *A. thaliana* affects morphology and susceptibility to *H. arabidopsidis*, we used a transcriptomic approach to test whether the effector induces transcriptional reprogramming. Our results indicate that Mlp124478 nucleolar accumulation is dispensable for the effector to exert its virulence activity.

Results

Mlp124478 carries a putative nuclear localization signal and a putative DNA-binding domain.

Mlp124478 is part of the CPG2811 multigenic family, which is specific to rust fungi with nine members; each is composed of a predicted signal peptide followed by two exons encoding short peptides (75–96 amino acids) (Fig. 1A,B). Except for the six conserved cysteine residues, amino acid conservation is low in the family. Amino acid identity ranges from 28% to 60% between Mlp124478 and the other family members (Fig. 1C). Mlp124478 is the only member of the CPG2811 family that exhibits a putative nuclear localization signal (NLS) and a putative DNA-binding domain (amino acids 29–38 and 58 to 80, respectively, Fig. 1A). The infection specific expression of Mlp124478 and its uniqueness among his family prompted us to investigate if it played a role *in planta* during pathogen growth.

Mlp124478 affects *Arabidopsis* leaf shape and accumulates in the nucleus and the nucleolus of *A. thaliana* cells. To evaluate the biological consequences of Mlp124478's presence in plant cells, we used functional genomic assays as summarized in Fig. 2. We generated a stable transgenic *A. thaliana* line expressing the mature form of Mlp124478 (i.e., without the signal peptide) fused to GFP under the control of a 35S promoter (pro35S::Mlp124478-GFP) in the Col-0 background (Fig. 2). Interestingly, the transgenic lines exhibited altered

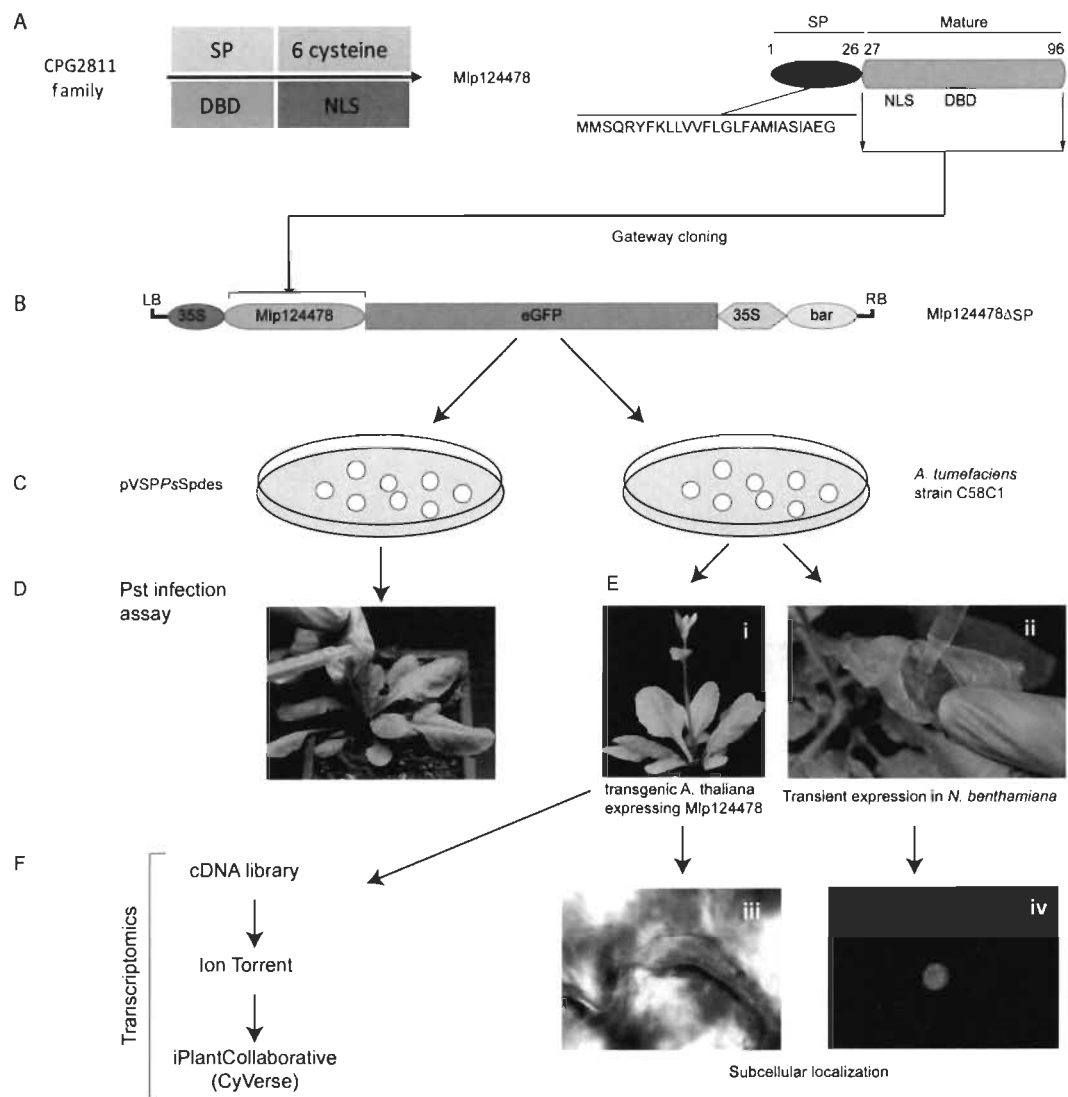


Figure 2. Overview of functional approaches applied to Mlp124478. (A) Mlp124478 was mined from CPG2811 family and has a signal peptide (SP), a putative nuclear localization signal (NLS) and a putative DNA-binding domain (DBD). (B) The mature coding sequences of Mlp124478 was cloned in frame with the green fluorescent protein (GFP). (C) Mlp124478 was recombined into pVSPSPsSpdes vector for Pst infection assay (effector delivery) and pB7FWG2.0 was then inserted into *A. tumefaciens* strain C58C1. (D) Pst expressing Mlp124478 was syringe infiltrated into the abaxial side of the leaves of *Arabidopsis thaliana*. (E) *A. tumefaciens* strain C58C1 expressing Mlp124478 was used to develop stable transgenic *A. thaliana* plants expressing Mlp124478 and perform transient expression, both were viewed by confocal microscopy (F) Transcriptomic study was performed with cDNA library preparation from the RNA extracted from the transgenic *A. thaliana* expressing Mlp124478 and control.

leaf morphology, characterized by waviness of leaf margins, while no curvature in the margins was evident in Col-0 plants (Fig. 3A). Anti-GFP immunoblotting for proteins extracted from Mlp124478-GFP and Col-0 lines revealed a band signal at the expected size of 37 kDa only in the transgenic line (Fig. 3B), indicating that the full length fusion accumulates in plant cells. Our results suggest that the constitutive *in planta* expression of the Mlp124478-GFP fusion alters leaf morphology.

To ascertain the subcellular localization of Mlp124478, we undertook confocal laser scanning microscopy of leaves from 4-day-old *A. thaliana* seedlings expressing Mlp124478-GFP fusion. We detected the GFP signal in the nucleolus, the nucleoplasm, and the cytosol of epithelial cells (Fig. 3C) similar to the localization observed in *N. benthamiana* by Petre *et al.* (2015b). In contrast, in control plants expressing GFP, the fluorescent signal accumulated only in the nucleoplasm and cytosol, and was excluded from the nucleolus (Fig. 3C). We conclude that Mlp124478-GFP mainly accumulates in the nucleolus and nucleoplasm of leaf cells, with a weak accumulation in the cytosol.

Mlp124478 Nuclear Localization Signal (NLS) is required for nucleolar accumulation. Mlp124478 carries a predicted NLS consisting of 10 amino acids within the N-terminal part of the mature form (Mlp124478₂₉₋₃₈;

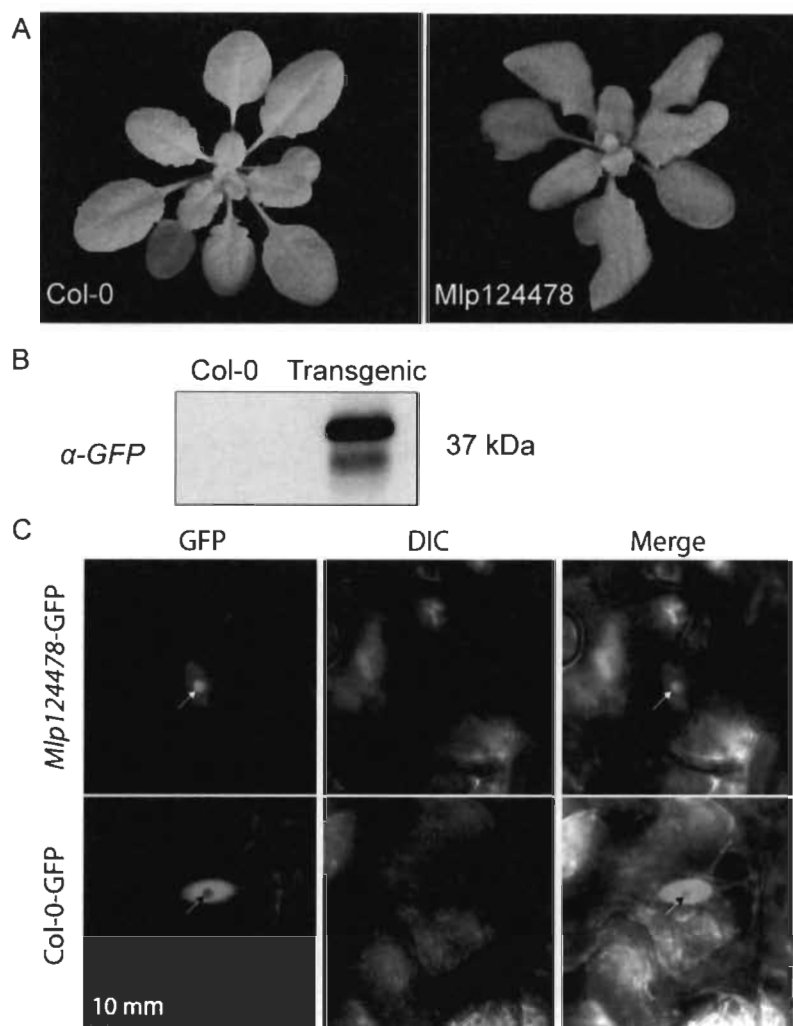


Figure 3. Mlp124478 carries a putative nuclear localization signal and a putative DNA-binding domain. (A) Morphology of 4-week-old soil grown *A. thaliana* Col-0 and stable transgenic plant expressing *Mlp124478* grown at 22 °C under 14 h/10 h photoperiod in growth chamber. (B) Immunodetection of GFP protein in Col-0 and stable transgenic seedlings from 12 days old plantlets. (C) Live cell imaging using confocal microscope of epidermal cells of 4-days-old *A. thaliana* stable transgenic *Mlp124478*-GFP plantlets. GFP in the Col-0 background was used as control. Left panel shows GFP, middle panel shows DIC and right panel shows merge. Nucleoli are pointed with black or white arrowheads.

RHKNNGGSRK) (Fig. 1). To assess whether the predicted NLS was required for nuclear localization, we designed a GFP tagged construct lacking the predicted NLS, hereafter named *Mlp124478* $_{\Delta 29-38}$ -GFP, and expressed it transiently in *N. benthamiana* leaf cells by agro-infiltration (Fig. 4A). Consistent with our observation in *A. thaliana* and from those of Petre *et al.* (2015), *Mlp124478*-GFP fusion accumulated in both the nucleus and nucleolus of *N. benthamiana* epithelial cells (Fig. 4B). However, *Mlp124478* $_{\Delta 29-38}$ -GFP accumulated solely in the nucleus and cytosol, and its signal was excluded from the nucleolus (Fig. 4B). To quantify the changes in subcellular distribution, we generated intensity plots of the fluorescent signals, which showed decreased fluorescence in the nucleolus between the two *Mlp124478* constructs (Fig. 4C and Supplementary Fig. 1). *Mlp124478*-GFP had a significantly higher nucleolar/nuclear ratio of 5.55 (SD = 1.55) compared to *Mlp124478* $_{\Delta 29-38}$ with a ratio of 0.8 (SD = 0.77) (Fig. 4C). Taken together, these results suggest that the predicted NLS of *Mlp124478* also acts as a nucleolar localization signal.

***Mlp124478*-GFP and *Mlp124478* $_{\Delta 29-38}$ -GFP increase *H. arabidopsidis* growth on *A. thaliana*.** In order to assess whether *Mlp124478* accumulation and localization in plant cell affects susceptibility to pathogen growth, we generated an additional transgenic line expressing *Mlp124478* $_{\Delta 29-38}$ -GFP and conducted two different pathogen assays. Firstly, we infected the stable transgenic *A. thaliana* that constitutively express effectors with *H. arabidopsidis*. Secondly, we used a *P. syringae* effector-delivery system.

Firstly, we observed that the wavy leaves phenotype observed in *Mlp124478*-GFP was strongly enhanced in *Mlp124478* $_{\Delta 29-38}$ -GFP. It resulted in twisted and larger leaves (Fig. 5A-ii) and *Mlp124478* $_{\Delta 29-38}$ -GFP plants also displayed early bolting (Fig. 5A-iii). We then confirmed that the effector localization in *A. thaliana*

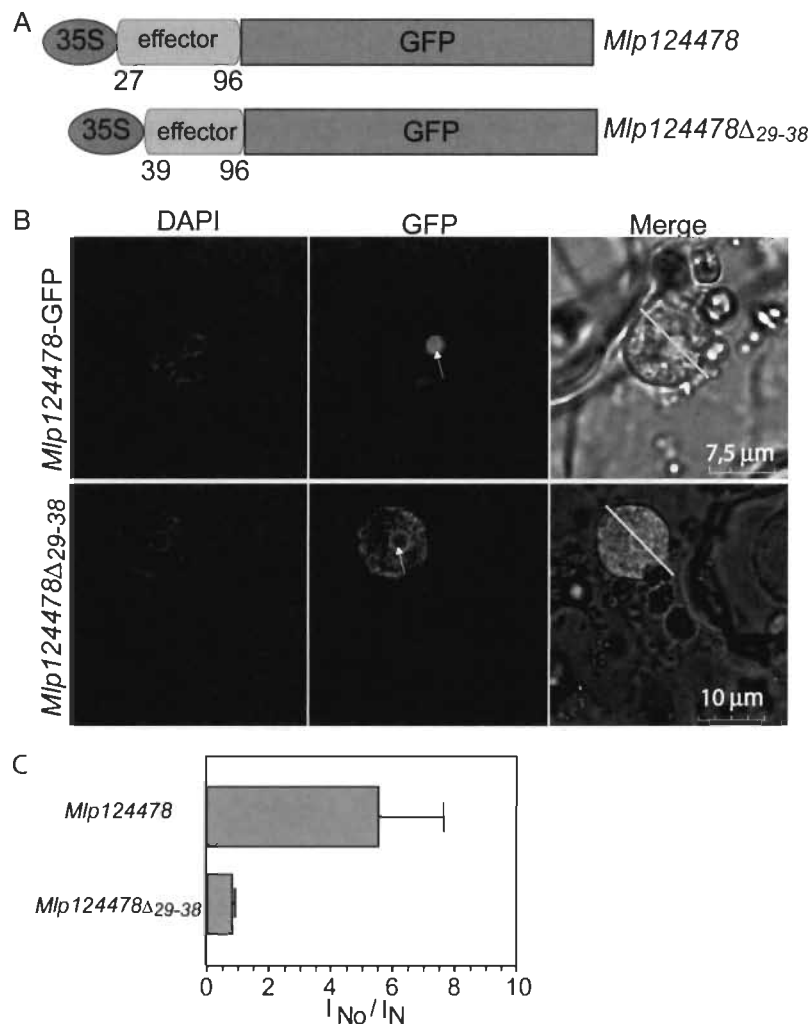


Figure 4. Mlp124478 Nuclear Localization Signal (NLS) is required for nucleolar accumulation. (A) Schematic representation of the constructs (Mlp124478 and Mlp124478 Δ_{29-38}) used for transient expression. (B) Subcellular accumulation of Mlp124478-GFP and Mlp124478 Δ_{29-38} -GFP in *N. benthamiana* epidermal cells at 4-days post-infiltration, the nucleus was stained by DAPI and epidermal cells were observed under the blue channel (left panel), green channel (middle panel) and merge of all channels (right panel). Arrowheads point the nucleolus. (C) Nuclear-nucleolar distribution of the fluorescent fusion proteins according to the fluorescence intensity ratios: nucleolar intensity (I_{No}) divided by nuclear intensity (I_N).

corroborated with the one observed in *N. benthamiana* and again observed a similar localization as before; that is Mlp124478-GFP effector accumulates in the nucleolus, nucleus and cytosol while the Mlp124478 Δ_{29-38} -GFP is excluded from the nucleolus but still accumulate in the nucleoplasm and to a lesser extent in the cytoplasm (Fig. 5B). Secondly, we performed infection assays to evaluate the susceptibility of the transgenic plants. Seven days following *H. arabidopsidis* spores inoculation, we quantified the number of spores and counted 3.8 times more spores on Mlp124478-GFP than on Col-0, 4.3 times more spores on Mlp124478 Δ_{29-38} -GFP than on Col-0 and 14.5 times more spores on *eds1-1* than on Col-0. We noted a significant increased susceptibility in Mlp124478 and Mlp124478 Δ_{29-38} -GFP transgenic plants compared to Col-0 ($P < 0.0001$), although not as strong as that encountered in *eds1-1* plants, used as positive control (Fig. 5C). These findings demonstrate that the nucleolar localization of Mlp124478 is not necessary for the augmented plant susceptibility to *H. arabidopsidis*.

Experiments using the plant bacterial pathogen *Pst*DC3000 Δ CEL carrying Mlp124478 or an empty vector (Supplementary Fig. 2A) and additional experimental assay using *Pst* in which the effector was expressed in *planta* did not demonstrate alteration of pathogen growth (Supplementary Fig. 2B). These results indicate that neither the full-length mature effector nor the truncated effector excluded from the nucleolus increased plant susceptibility to this bacterial pathogen. From this experiment set, we conclude that Mlp124478 enhances the growth of the filamentous pathogen *H. arabidopsidis* but not of the bacterial pathogen *P. syringae* in *A. thaliana*.

The expression of Mlp124478 in plant cells alters *A. thaliana* transcriptome. To better understand how Mlp124478 functions in plant cells, and since it alters plant morphology and susceptibility to pathogen, we investigated whether Mlp124478 alters gene expression in *A. thaliana*. We performed transcriptome profiling

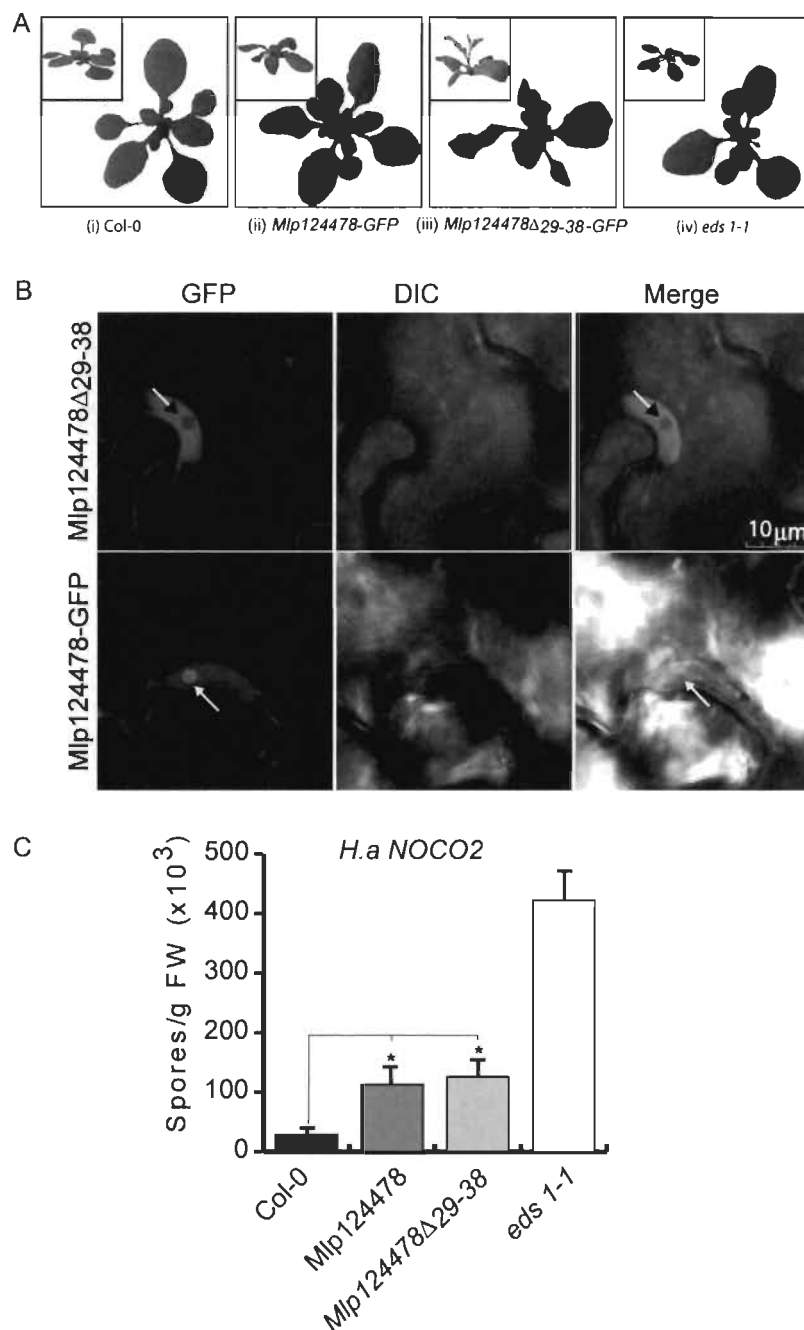


Figure 5. *Mlp124478-GFP* and *Mlp124478 Δ 29-38-GFP* increase *H. arabidopsidis* growth on *A. thaliana*. (A) Col-0 plant showing normal leaves (i); wavy leaves phenotype observed in *Mlp124478-GFP* (ii); strongly enhanced leave waviness and early bolting in *Mlp124478 Δ 29-38-GFP* (iii) morphology of *eds1-1* (iv). (B) Live cell imaging using confocal microscopy of epidermal cells of 4-days-old stable transgenic *Mlp124478 Δ 29-38-GFP* and *Mlp124478-GFP* plantlets. Left panel displays GFP, middle panel shows DIC, right panel shows the merge. Nucleoli are pointed with arrows. Scale bar: 10 μ m. (C) Four weeks old soil grown Col-0, stable transgenic *Mlp124478*, *Mlp124478 Δ 29-38-GFP* and *eds1-1* plants were spray inoculated with *Hyaloperonospora arabidopsidis* Noco2 (50,000 conidiospores/mL) and the number of conidiospores were quantified at 7 days after inoculation. Statistical significance was evaluated using student's *t* test. Asterisk denotes significant difference to Col-0, $p < 0.0001$ for *Mlp124478* and $p < 0.002$ for *Mlp124478 Δ 29-38*. Experiments were repeated three times with similar results.

of 4-days-old *A. thaliana* *Mlp124478* stable transgenic line and control plants expressing GFP. We identified 98 up- and 294 down-regulated genes, respectively (Fig. 6A, Supplementary Dataset). To test the robustness of the transcriptome data, we used qRT-PCR to assess the expression of 3 randomly selected up-regulated genes and 7 down-regulated genes. Transcriptome and qRT-PCR data correlated well, although quantitative differences were detected (Supplementary Fig. 3).

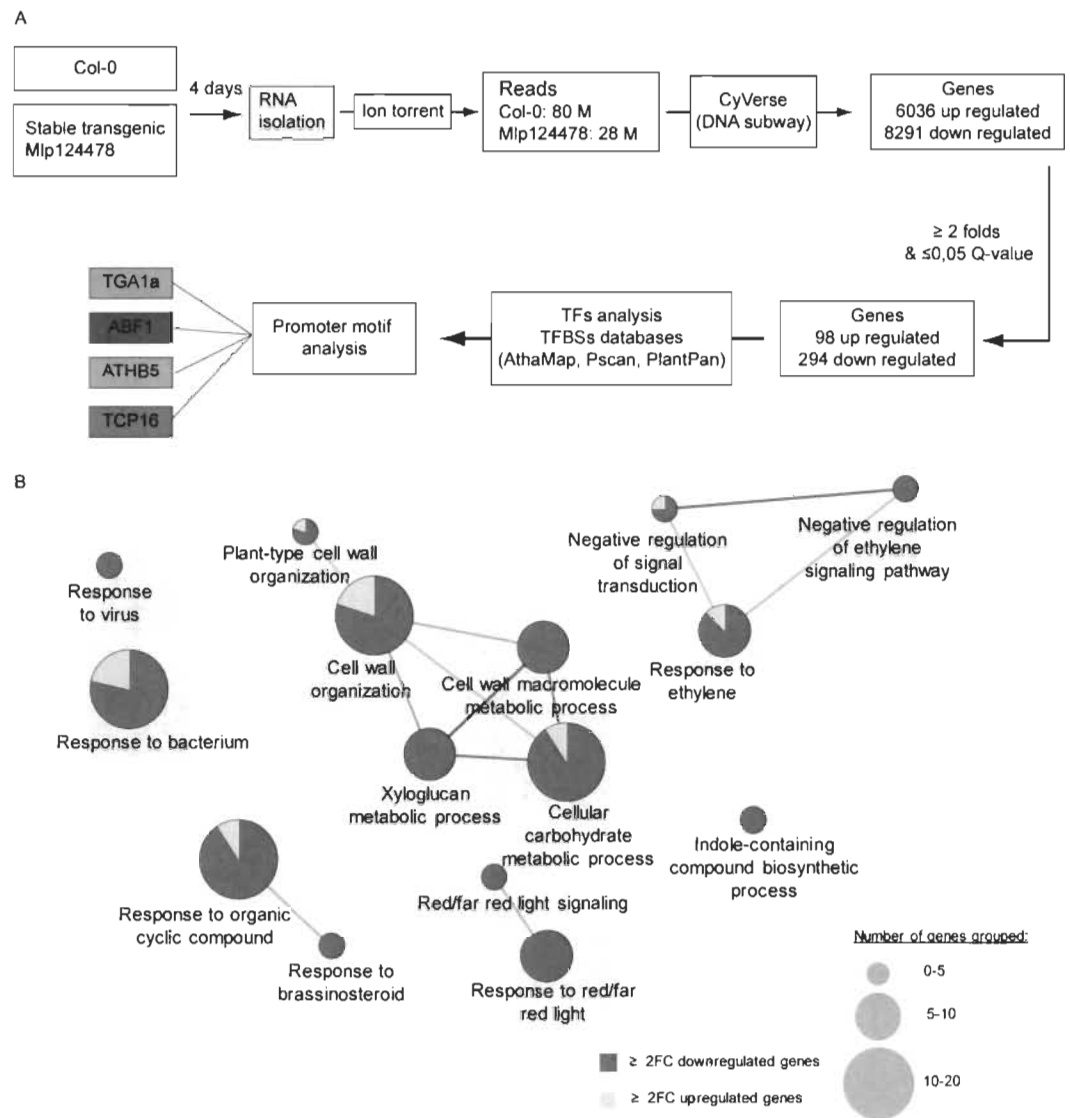


Figure 6. The expression of Mlp124478 in plant cells alters *A. thaliana* transcriptome. (A) Schematic illustration of transcriptomic work flow. RNA was isolated from 4-days old *A. thaliana* Mlp124478 stable transgenic and Col-0 plants and sequenced using Ion torrent. Transcripts were analyzed using iPlantCollaborative DNA subway and deregulated genes were considered for further analysis. (B) Go term enrichment analysis was performed with deregulated genes filtered with Q-value ≤ 0.05 and fold-change ≥ 2 using the Cytoscape software (version 3.1.1). Cytoscape was performed with the plug-in ClueGO and CluePedia to visualize functions enriched in the deregulated genes. The GO terms presented are significantly enriched in up-regulated and down-regulated genes with FDR ≤ 0.05 (Benjamini-Hochberg p-value correction) and revealed 15 GO terms belonging to 7 functional groups.

GO term enrichment analysis was applied to the deregulated genes to determine relevant biological processes. Seven functional groups (groups 0–6) of GO terms were significantly enriched among deregulated genes (Fig. 6B). The up-regulated genes with related GO terms are presented in (Fig. 6B). Among the 294 down-regulated genes and out of the 42 genes of the “cell wall organization”, 37 belong to the xyloglucan transglycosylase XTH, XRT and EXT families. The defense-related transcription factors WRKY18, WRKY27, WRKY33, MYB51, the defense-related proteins NHL3, RPP8, YLS9, AZI1, CRK11 and the jasmonate pathway and regulation genes JAZ1, ASA1, ASB1 were down-regulated in the Mlp124478-GFP transgenic lines compared to the GFP transgenic plants. Other genes involved in diverse mechanisms were down-regulated such as the chitinase CHI, the brassinosteroid-related genes BAS1, BES1, PAR1, BEE1, the salicylic acid-related genes NPR3, the ethylene-related response genes ARGOS and ARGOS-like (ARL), EBF2, ERF6, ETR2, RTE1, the carbon metabolism-related genes EXO, the red/far red light signalization-related genes FAR1, GA2OX2, PAR1, PIF3, PKS4. The changes in Mlp124478-GFP *A. thaliana* transgenic line transcriptomes occurred mostly by a down-regulation of the expression of genes involved in diverse functions, frequently related to defense response regulation.

Next, we analyzed the gene expression profiles of up- and down-regulated genes during different biotic perturbations. We accessed Genevestigator towards this end. Expression levels in five different infection conditions

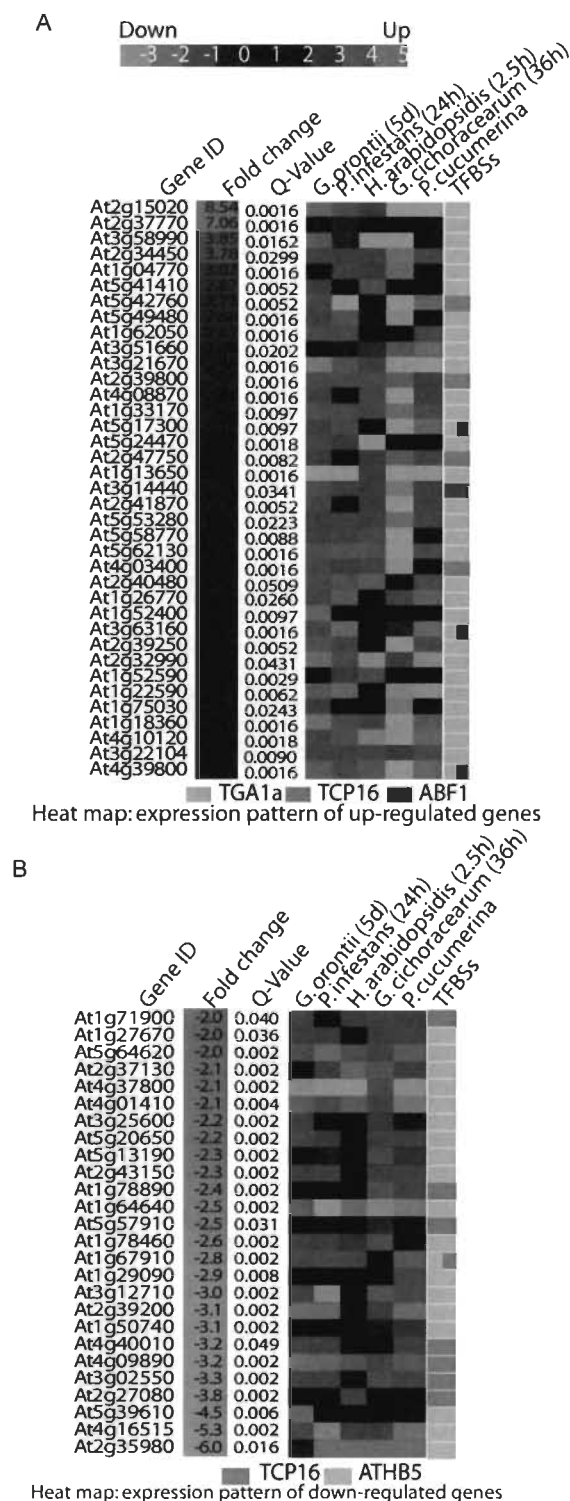


Figure 7. Regulation of gene expression level. Heat map of biotrophic pathogens response of genes in two groups: (A) upregulated genes and (B) down regulated genes. Genevestigator was used for differential expression analysis.

(*Golovinomyces orontii*, *Phytophthora infestans*, *H. arabidopsidis*, *Golovinomyces cichoracearum*, *Plectosphaerella cucumerina*) were retrieved for all up- and down-regulated genes in the *A. thaliana* transgenic line overexpressing *Mlp124478* (Fig. 7A,B). Almost all up-regulated genes (92%) in our transcriptome analysis were down-regulated in response to these pathogens. Only one gene (At3g51660) also appeared up-regulated (maximum fold change of 2.5) in most conditions (Fig. 7A). Of the 30 down-regulated genes, 8 were up-regulated in almost all conditions surveyed (At2g37130, At3g25600, At5g13190, At5g57910, At2g39200, At1g50740, At5g39610, At2g35980; Fig. 7B). We further analyzed the identity of these genes. At2g37130 encodes a peroxidase, which was strongly

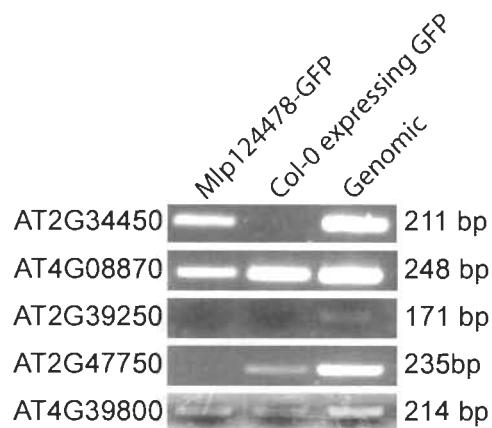


Figure 8. Mlp124478 binds DNA. Two-weeks-old plants tissues of Col-0 expressing GFP or stable transgenic Mlp124478 were used for chromatin preparation using ChIP assay with antibody against GFP as described in the material and methods section and *A. thaliana* genomic DNA was used as a positive control. TGA1a associated site was PCR amplified with TGA1a specific primer pair. Expected bands (211 bp) was obtained from transgenic and *Arabidopsis* genomic DNA for TGA1a at the promoter region of AT2G34450 gene. Mlp124478-GFP indicates chromatin IP from that line. Col-0 indicates that chromatin was immunoprecipitated from Col-0 expressing GFP (negative control). Genomic Col-0 DNA, not immunoprecipitated served as a positive control. The other genes shown (At4g08870, At2g39250, At2g47750, At4g39800) are examples of non specific reaction. Col-0 expressing GFP DNA: negative control; *A. thaliana* genomic DNA: positive control.

up-regulated in response to fungal infection. At5g13190 encodes a plasma membrane protein regulating cell death. At2g39200 encodes *MILDEW RESISTANCE LOCUS O 12* (*AtMLO12*) whereas the product of the At2g35980 gene was very similar to *Arabidopsis* *NON-RACE-SPECIFIC DISEASE RESISTANCE 1* (*NDR1*). These results mirror those obtained with Cytoscape and further confirm that Mlp124478 rewires host transcription specifically to induce genes not normally expressed during infection against those five pathogens, whereas gene normally up-regulated in response to these pathogens are down-regulated upon expression of Mlp124478.

Mlp124478 binds DNA. The localization of Mlp124478 in *planta*, the presence of a DNA-binding motif and alterations at the transcriptional, morphological, and immune levels prompted us to investigate whether Mlp124478 associates with DNA molecules. In a first time, we screened for Transcriptional Factor Binding Sites (TFBS) in the promoter sequences of all up- and down-regulated genes identified in our transcriptome analysis. We identified four different TFBS, which were very abundant among the up- (43 genes out of 98) and down-regulated (30 genes out of 294) genes (Supplementary Dataset). TFBS in the up-regulated gene set included ABF1 and TGA1a which belongs to the basic region/leucine zipper motif (bZIP) transcription factor (TF) family; and TCP16 belongs to the TCP (*TEOSINTE BRANCHED 1*, *CYCLOIDEA* and *PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR 1*) TF family. The TFBS ATHB5 and TCP16 were also among the down-regulated genes. Thus, these TFBS were selected as candidate targets for Mlp124478-DNA interaction studies.

In a second time, we performed a ChIP-PCR experiment. We cross-linked proteins and DNA using formaldehyde, and then immunoprecipitated (IP) Mlp124478-GFP fusion with anti-GFP beads from transgenic plants. We designed primer pairs that could amplify the promoter regions most abundant among de-regulated genes (TCP16, ATHB5, TGA1a and ABF1). One of the primer sets amplified DNA, revealing an interaction of Mlp124478 with the TGA1a-binding site of AT2G34450, one of the gene among the most strongly up-regulated genes in Mlp124478-expressing plants. We did not observe any band in the IP with Col-0 plant expressing GFP, which served as negative control, but a band was produced with *A. thaliana* Col-0 genomic DNA as positive control (Fig. 8) (four examples (At4g08870, At2g39250, At2g47750 and Atg39800) of non specific interaction are shown below At2g34450 (all interaction are shown in Supplementary Fig. 4)). AT2G34450 was up-regulated in the presence of Mlp124478 and showed down-regulation against biotrophic pathogens (Fig. 7A). We attempted EMSA with a synthetic peptide encompassing the DNA-binding domain of Mlp124478 and a double-stranded oligonucleotide displaying the consensus TGA1a sequence, but did not discern any interaction (Supplementary Fig. 5). This results confirms that Mlp124478 interacts with DNA but the exact binding domain, sufficient for interaction could not be identified.

Since poplars are natural the hosts of *M. larici-populina*, we searched for the sequence immunoprecipitated in the ChIP experiment in the *Populus trichocarpa* genome. The promoter of the gene model POPTR_0004s13630.1 exhibits 57% of identity with the AT2G34450 promoter. Both genes encode protein with a similar predicted function, with a conserved exon-intron structure (6 exons and 5 introns), and a TGA1a regulatory sequence in their promoter sequence (Supplementary Fig. 6), suggesting that DNA interaction in *Arabidopsis* could also occur in poplars.

Discussion

Recently, several groups reported on the use of heterologous systems to investigate the function, localization, and interaction of effectors from biotrophic pathogens^{6,7,22,25–29}. It has also been shown that many pathogens effectors target the nucleus and, in some cases, alter transcription^{30–32}. Here, we undertook a functional genomics approach to study Mlp124478, a CSEP from the poplar leaf rust pathogen *M. larici-populina*. We conducted *in planta* pathogen assays, live-cell imaging, comparative transcriptomics, and protein-nucleic acid interaction to assess Mlp124478 function.

One of the major findings of our study is that Mlp124478 represses the expression of genes involved in defense response. The GO terms that were most significantly enriched were response to virus, response to bacterium, response to brassinosteroid, indole-containing compound biosynthetic process, cell wall organization, response to red or far red light signaling and negative regulation of ethylene-activated signaling pathway (Fig. 6B). Among the most down-regulated genes, several belonged to the defense-related transcription factors WRKY18, WRKY27, WRKY33, MYB51^{33,34}, the defense-related proteins NHL3³⁵, RPP8³⁶, YLS9³⁷, AZI1³⁸, CRK11³⁹ and the jasmonate pathway such as JAZ1⁴⁰, ASA1, ASB1⁴¹, a chitinase which is involved in defense against fungi and salicylic acid-related genes NPR3⁴². Thus, the changes in the transcriptome of the transgenic line expressing *Mlp124478-GFP* occur mostly by the down-regulation of expression of genes involved in functions frequently related to defense response regulation.

The presence of a DNA-binding domain in Mlp124478 and the fact that we could confirm Mlp124478 interaction with DNA in a sequence-specific manner, besides alteration of the transcriptome through downregulation of defense related genes strongly suggest that it may alter gene expression to deceive plant immune systems. Recently, effectors from filamentous pathogens that bind DNA have been identified. CgEP1, a *Colletotrichum graminicola* effector with DNA-binding properties has been shown to enhance anthracnose development in maize³. Like Mlp124478, the oomycete effector PsCRN108 exhibits a putative DNA-binding domain, localizes to the nucleus and downregulate the expression of defense-related genes⁴³. Several down-regulated genes found in our study corresponded to genes also reported recently highlighted in a transcriptomic analysis of *A. thaliana* responses during colonization by the two fungi *Colletotrichum tofieldiae* (symbiont) and *Colletotrichum incanum* (parasite). This study and our results share eight GO terms, with the exception that genes induced during the colonization by *C. incanum* are down regulated in Mlp124478 transgenic lines (Supplementary table 2). Hence, the expression of this single effector (Mlp124478) appears to bear broad transcriptional impact as it appears to counter the normal gene regulation described by Hacquard (2016) using a very similar analysis. However, since the EMSA assay could not confirm a direct protein-DNA interaction, the broad transcriptional changes caused by the presence of Mlp124478 could be caused by an indirect effect or interaction with a transcriptional regulator.

When *Arabidopsis* was exposed to a filamentous pathogen that possesses a mode of infection similar to rust fungi, we observed more susceptibility to pathogen growth, but susceptibility was not enhanced to infection by bacterial pathogen. This result indicates that this effector may target an immunity component specifically affected by pathogens with filamentous lifestyle. The morphology of the plants expressing *Mlp124478* or *Mlp124478_{Δ29–38}* is altered, the plants show wavy leaves and *Mlp124478_{Δ29–38}* displays early bolting. Altered phenotype has previously been associated to altered susceptibility to pathogen. For instance, the *scn1* mutant plants have been thoroughly described as having increased resistance to pathogen and accumulate elevated salicylic acid level⁴⁴, however *scn1* plants have very short stature unlike the plants expressing *Mlp124478*. Although wavy leaf phenotype has been reported before it does not appear to be linked to plant susceptibility⁴⁵.

Since the default GFP distribution in plant cells is nucleo-cytoplasmic, the localization of a GFP-tagged effector displaying nucleo-cytoplasmic distribution is considered non-informative. However, in the case of Mlp124478, the localization in nucleoli indicates that GFP is not masking the Mlp124478 localization sequence, thus localization is driven by the effector sequence. Nucleolus targeting has long been recognized as a hallmark of virus infection^{46–48}, essentially to recruit nucleolar proteins and facilitate virus replication⁴⁸, but has also been observed for other pathogens, including oomycete⁴⁹ and bacteria⁵⁰. While viral lifestyle easily explains the need to target the nucleolus, the reasons why a rust effector would do so are not as clear. Since the virulence activity of Mlp124478 does not require nucleolar accumulation, the accumulation of Mlp124478 in the nucleolus remains unexplained; however Mlp124478 could have additional function in the nucleolus which is undetected in our virulence assays.

Taken together, our results suggest that Mlp124478 likely manipulates plants by targeting DNA, remodeling transcription via DNA-binding, to suppress normal transcriptional responses to pathogens, and mislead the host into up-regulating the expression of genes unrelated to defense.

Materials and Methods

Plant material and growth conditions. *A. thaliana* and *N. benthamiana* plants were soil-grown in a growth chamber under a 14 h/10 h light/dark cycle with temperature set at 22 °C and relative humidity of 60%. The plants were grown in Petri dishes for the selection of single-insertion homozygous transgenic *Mlp124478* with ½ Murashige and Skoog medium containing 0.6% agar and 15 mg/ml Basta.

Growth of *Pseudomonas syringae* pv. *tomato*, *H. arabidopsidis* Noco2 and infection assay. *Pseudomonas syringae* strain ΔCEL⁵¹ containing *Mlp124478* was grown overnight and infiltrated in the leaves of 4-weeks-old Col-0 and of transgenic *Mlp124478* plants at an optical density at 600 nm (OD₆₀₀) of 0.001. *Pst* infections were produced by syringe infiltration of 4-weeks-old *Arabidopsis* plant leaves, and *H. arabidopsidis* Noco2 spray infections were performed as previously described⁵².

Plasmid construction. Constructs were developed with Gateway cloning systems (Invitrogen, Life Technologies). The *Mlp124478* coding sequence without the signal peptide (lacking amino acids 1–27,

hereafter referred to as *Mlp124478*) was ordered from GenScript in pUC57 in lyophilized form, and primer pairs (Supplementary Table 1, Primers Nos 1–3) were used to amplify the open reading frame (ORF) of *Mlp124478* from pUC57 by the polymerase chain reaction (PCR). Amplicons were then cloned into pDONRTM221 entry vector by Gateway BP recombination, followed by recombination with Gateway LR reaction either into pVSPSPsSpdes vector for *Pst* infection assay (effector delivery) or pB7FWG2.0 vector³³ to express C-terminal green fluorescent protein (GFP)-tagged *Mlp124478* fusion in *planta*. pVSPSPsSpdes harbors the AvrRpm1 secretion signal³¹.

Transient expression in *N. benthamiana* leaf cells. Solutions of *A. tumefaciens*-carrying recombinant plasmids were infiltrated into leaf pavement cells of 4-weeks-old *N. benthamiana* plants⁵⁵. Briefly, *A. tumefaciens* AGL1-competent cells were transformed with pB7FWG2-containing *Mlp124478* and grown overnight in yeast extract peptone medium supplemented with spectinomycin (50 mg/L). The cells were precipitated by centrifugation at 300g and adjusted to OD₆₀₀ of 0.5 in infiltration buffer (10 mM MgCl₂ and 150 μM acetosyringone). After 1 h, the agro-suspension was infiltrated into the abaxial side of leaves, and the plants were returned to the growth chamber. At 2 days post-infiltration (dpi), water-mounted slides of leaf tissue from agro-infected leaves were visualized by confocal microscopy.

Confocal laser scanning microscopy. Leaves were observed under a Leica TCS SP8 confocal laser scanning microscopy (Leica Microsystems). Images were obtained with HC PL APO CS2 40X/1.40 oil immersion objective, and acquired sequentially to exclude excitation and emission crosstalk (when required). Leaves were immersed in water containing 0.2 μg/ml DAPI for 15 min for nuclei staining at room temperature. The samples were then observed at excitation/emission wavelengths of 405/444–477 nm and 488/503–521 nm for DAPI and eGFP, respectively. Images were annotated with LAS AF Lite software.

Chromatin immunoprecipitation (ChIP)-polymerase chain reaction assay. ChIP assays were conducted, as described previously, with minor modifications⁵⁶. Briefly, 300 mg of 2-weeks-old *A. thaliana* *Mlp124478* stable transgenics and Col-0 plants expressing GFP were collected in tubes containing 10 mL of phosphate-buffered saline (PBS), which were replaced by 10 mL of 1% formaldehyde to cross-link tissue under vacuum infiltration. To quench the cross-linker, 0.125 M glycine was added after removal of formaldehyde, followed by vacuuming, incubation for 5 min, and tissue-rinsing with 10 mL cold PBS. Cross-linked tissues were dried on paper towel for nuclei isolation. Sonicated chromatin was immunoprecipitated with 50 μL/mL anti-GFP microbeads (MACS, Miltenyi Biotec Inc.) and incubated for 2 h at 4 °C. The beads were placed in the μ-column, in the magnetic field of a μMACS separator, and washed twice. After reverse cross-linking of DNA-protein, ChIP samples underwent DNA purification according to a previously-described method⁵⁶, followed by PCR amplification with specific primer pairs listed in Supplementary Table 1 (Primer Nos 4–38).

Electrophoretic mobility shift assay (EMSA). EMSA was undertaken, as described earlier⁵⁷, with minor modifications. Unlabeled and digoxigenin (DIG)-labeled forward TGA1a oligonucleotides were ordered from Integrated DNA Technologies. Double-stranded (DS) oligonucleotides were annealed by heating 1 nmol of each oligonucleotide at 95 °C for 10 min, then slowly cooled down to 20 °C. DS oligonucleotides were diluted in TEN buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA, pH 8, 100 mM NaCl) to a final concentration of 50 pmol/μL. Dot blotting was carried out by serial dilutions and spotted on positively-charged nylon membranes to test efficiency of the DIG-labeled probe. 3 pmol of probe was found to be efficient for detection with anti-DIG primary antibody. Gel shift reaction was performed with 3 pmol of DS oligonucleotides and 100 ng of synthetic peptide in binding buffer (100 mM HEPES, pH 7.6, 5 mM EDTA, 50 mM (NH₄)₂SO₄, 5 mM DTT, 1% Tween 20 and 150 mM KCl). After binding reaction at 25 °C for 15 min, the samples were placed on ice for 15 min, and the mixtures were electrophoresed immediately through 0.25X TBE 20% polyacrylamide gel at 12.5 volts/cm. Bio-Rad semi-dry transfer cells were electroblotted on positively-charged nylon membranes at 25 volts for 10 min. DNA was then cross-linked to the membrane by baking at 80 °C for 40 min. For DIG detection, the membranes were blocked in TBS (50 mM Tris, 150 mM NaCl, 1% BSA), followed by 2 washes with TBS for 10 min and 1 wash with TBST (TBS and 1% Tween 20), then incubated overnight at 4 °C with anti-DIG monoclonal antibody diluted 1:1,000 in TBS with 1% BSA. The membranes were washed 4 times in TBS for 5 min and once in TBST. Finally, they were incubated with HRP-conjugated secondary antibody diluted 1:3,000 in TBST with 5% milk at room temperature for 45 min. The membranes were washed 4 times in TBS and once in TBST for 5 min. Bio-Rad's Clarity Western ECL blotting substrate was then applied for detection. EMSA was performed at least 3 times with independent dilution of synthetic peptides and freshly-hybridized DIG probe.

RNA extraction and transcriptome analysis. Total RNA was extracted from 4-days-old *A. thaliana* *Mlp124478* stable transgenics and from control plants expressing GFP with the RNeasy Plant Mini Kit (Qiagen, Inc.), according to the manufacturer's specifications. The growth stage (Petri grown 4-days-old seedlings) was chosen to avoid variation due to growth chamber variation or micro-environmental variation, which would result in noise in the transcriptome analysis. Control and transgenic plants were extracted in triplicate. Eluted total RNA was quantified and sent to the Plateforme d'Analyses Génomiques of the Institut de Biologie Intégrative et des Systèmes (Université Laval, Quebec City, Canada) for library construction and sequencing with the Ion Torrent Technology. Differential expression was analyzed with green line workflow of the DNA subway in the iPlant collaborative pipeline (now CYVERSE) (Cold Spring Harbor Laboratory), including *A. thaliana*-Ensembl TAIR 10 as reference genome. Genes with a Q-value ≤ 0.05 and a fold-change ≥ 2 were considered as significantly differentially expressed and were further investigated for Gene Ontology (GO) enrichment. The Cytoscape software (version 3.1.1)⁵⁸ with the plug-in ClueGO and CluePedia⁵⁹ was used to visualize functions enriched among deregulated genes. The threshold for GO terms deregulation was set as FDR ≤ 0.05 (Benjamini-Hochberg p-value correction).

qRT-PCR validation of the transcriptomic analysis. For qRT-PCR total RNA was extracted with the RNeasy Plant Mini Kit (Qiagen, Inc., Valencia, CA, USA) according to the manufacturer's instructions. RNA quality was assessed by agarose gel electrophoresis and quantified by spectrophotometry. One μg of each sample was reverse transcribed into cDNA with the High Capacity cDNA Archive Kit (Life Technologies, Burlington, ON, Can). Quantitative RT-PCR (RT-qPCR) amplification was undertaken with a BioRad Detection system using SYBR Green PCR Master Mix (Bioline, London, U.K.). 100 ng cDNA template and 0.4 μM of each primer (listed in Supplementary Table 1) (were used in a final volume of 20 μl). The qRT-PCR thermal profile was: 95 °C for 2 min, 40 cycles of 95 °C for 5 s, 58 °C for 10 s, and 72 °C for 5 s. To analyze the quality of dissociation curves, the following program was added after 40 PCR cycles: 95 °C for 1 min, followed by constant temperature increases from 55 °C to 95 °C. Actin 1 served to normalize all RT-qPCR results. The expression levels of each gene were calculated according to the $\Delta\Delta\text{Ct}$ method⁶⁰. Three technical replicates for each treatment were analyzed. Standard deviation was computed by the error propagation rule.

Bioinformatics analyses. Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) was used to align sequences of the nine gene members of the CPG2811 SSP family, which were later manually annotated. Phylogenetic trees were generated by COBALT (<http://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi>). SignalP 4.0 (<http://www.cbs.dtu.dk/services/SignalP/>) predicted signal peptides. NLStradamus (<http://www.mose-slab.csb.utoronto.ca/NLStradamus/>) forecast nuclear-localizing signals. Transcription factor-binding sites (TFBS) were identified and analyzed with the AthaMap (http://www.athamap.de/search_gene.php)⁶¹, Pscan (<http://159.149.160.88/pscan/>)⁶² and PlantPan (<http://plantpan2.itsp.ncku.edu.tw/index.html>)⁶³ databases. Consensus TFBS sequences were retrieved from the Pscan database. Promoter sequences were obtained individually with TAIR's SeqViewer (<http://tairvm09.tacc.utexas.edu/servlets/sv>), and TFBS-specific primers (Supplementary Table 1, Primer Nos 4–38) were designed with Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>). Gene expression data under different biological conditions were retrieved from Genevestigator (https://genevestigator.com/gv/doc/intro_plant.jsp). Protein DNA-binding sites were predicted by MetaDBSite (<http://projects.biotec.tu-dresden.de/metadbsite/>)^{64,65}.

References

- Chaudhari, P., Ahmed, B., Joly, D. L. & Germain, H. Effector biology during biotrophic invasion of plant cells. *Virulence* 5, 703–709, <https://doi.org/10.4161/viru.29652> (2014).
- Lewis, J. D., Guttman, D. S. & Desveaux, D. The targeting of plant cellular systems by injected type III effector proteins. *Semin Cell Dev Biol* 20, 1055–1063, <https://doi.org/10.1016/j.semcdb.2009.06.003> (2009).
- Vargas, W. A. *et al.* A Fungal Effector With Host Nuclear Localization and DNA-Binding Properties Is Required for Maize Anthracnose Development. *Mol Plant Microbe Interact* 29, 83–95, <https://doi.org/10.1094/MPMI-09-15-0209-R> (2016).
- Win, J. *et al.* Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb Symp Quant Biol* 77, 235–247, <https://doi.org/10.1101/sqb.2012.77.015933> (2012).
- Fabro, G. *et al.* Genome-wide expression profiling Arabidopsis at the stage of Golovinomyces cichoracearum haustorium formation. *Plant Physiol* 146, 1421–1439, <https://doi.org/10.1104/pp.107.111286> (2008).
- Gauar, O., Morency, M.-J., Letanneur, C., Séguin, A. & Germain, H. The 124202 candidate effector of *Melampsora larici-populina* interacts with membranes in *Nicotiana* and *Arabidopsis*. *Canadian Journal of Plant Pathology*, <https://doi.org/10.1080/07060661.2016.1153523> (2016).
- Kunjeti, S. G. *et al.* Identification of *Phakopsora pachyrhizi* Candidate Effectors with Virulence Activity in a Distantly Related Pathosystem. *Front Plant Sci* 7, 269, <https://doi.org/10.3389/fpls.2016.00269> (2016).
- Rafiqi, M., Ellis, J. G., Ludowici, V. A., Hardham, A. R. & Dodds, P. N. Challenges and progress towards understanding the role of effectors in plant-fungal interactions. *Curr Opin Plant Biol* 15, 477–482, <https://doi.org/10.1016/j.pbi.2012.05.003> (2012).
- Sohn, K. H., Lei, R., Nemri, A. & Jones, J. D. The downy mildew effector proteins ATR1 and ATR13 promote disease susceptibility in *Arabidopsis thaliana*. *Plant Cell* 19, 4077–4090, <https://doi.org/10.1105/tpc.107.054262> (2007).
- Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* 326, 1509–1512, <https://doi.org/10.1126/science.1178811> (2009).
- Motion, G. B., Amaro, T. M., Kulagina, N. & Huitema, E. Nuclear processes associated with plant immunity and pathogen susceptibility. *Brief Funct Genomics*, <https://doi.org/10.1093/bfpg/ely013> (2015).
- Rivas, S. & Genin, S. A plethora of virulence strategies hidden behind nuclear targeting of microbial effectors. *Front Plant Sci* 2, 104, <https://doi.org/10.3389/fpls.2011.00104> (2011).
- Gu, K. *et al.* R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 435, 1122–1125, http://www.nature.com/nature/journal/v435/n7045/supinfo/nature03630_S1.html (2005).
- Yang, B., Sugio, A. & White, F. F. Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proc Natl Acad Sci USA* 103, 10503–10508, <https://doi.org/10.1073/pnas.0604088103> (2006).
- Caillaud, M. C. *et al.* A downy mildew effector attenuates salicylic acid-triggered immunity in *Arabidopsis* by interacting with the host mediator complex. *PLoS Biol* 11, e1001732, <https://doi.org/10.1371/journal.pbio.1001732> (2013).
- Doehlemann, G. *et al.* Reprogramming a maize plant: transcriptional and metabolic changes induced by the fungal biotroph *Ustilago maydis*. *Plant J* 56, 181–195, <https://doi.org/10.1111/j.1365-3113.2008.03590.x> (2008).
- Aime, M. C., McTaggart, A. R., Mondo, S. J. & Duplessis, S. In *Advances in Genetics* Vol. 100 (eds Townsend, J. P. & Wang, Z.) 267–307 (Academic Press, 2017).
- Dean, R. *et al.* The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol* 13, 414–430, <https://doi.org/10.1111/j.1364-3703.2011.00783.x> (2012).
- Pinon, J. & Frey, P. In *Rust diseases of willow and poplar* 139–154 (CABI Publishing, 2005).
- Duplessis, S. *et al.* Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences* 108, 9166–9171, <https://doi.org/10.1073/pnas.1019315108> (2011).
- Hacquard, S. *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* 25, 279–293, <https://doi.org/10.1094/MPMI-09-11-0238> (2012).
- Petre, B. *et al.* Candidate Effector Proteins of the Rust Pathogen *Melampsora larici-populina* Target Diverse Plant Cell Compartments. *Molecular Plant-Microbe Interactions* 28, 689–700, <https://doi.org/10.1094/MPMI-01-15-0003-R> (2015).
- Germain, H. *et al.* Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Molecular Plant Pathology* 19, 191–200, <https://doi.org/10.1111/mpp.12514> (2018).

24. Duplessis, S. *et al.* Melampsora larici-populina Transcript Profiling During Germination and Timecourse Infection of Poplar Leaves Reveals Dynamic Expression Patterns Associated with Virulence and Biotrophy. *Molecular Plant-Microbe Interactions* 24, 808–818, <https://doi.org/10.1094/MPMI-01-11-0006> (2011).
25. Caillaud, M. C. *et al.* Subcellular localization of the Hpa RxLR effector repertoire identifies a tonoplast-associated protein HaRxLR17 that confers enhanced plant susceptibility. *Plant J* 69, 252–265, <https://doi.org/10.1111/j.1365-3113.2011.04787.x> (2012).
26. Caillaud, M. C. *et al.* Mechanisms of nuclear suppression of host immunity by effectors from the Arabidopsis downy mildew pathogen Hyaloperonospora arabidopsidis (Hpa). *Cold Spring Harb Symp Quant Biol* 77, 285–293, <https://doi.org/10.1101/sqb.2012.77.015115> (2012).
27. Du, Y., Berg, J., Govers, F. & Bouwmeester, K. Immune activation mediated by the late blight resistance protein R1 requires nuclear localization of R1 and the effector AVR1. *New Phytol* 207, 735–747, <https://doi.org/10.1111/nph.13355> (2015).
28. Petre, B. *et al.* Rust fungal effectors mimic host transit peptides to translocate into chloroplasts. *Cellular Microbiology* 18, 453–465, <https://doi.org/10.1111/cmi.12530> (2016).
29. Petre, B. *et al.* Heterologous Expression Screens in Nicotiana benthamiana Identify a Candidate Effector of the Wheat Yellow Rust Pathogen that Associates with Processing Bodies. *PLoS One* 11, e0149035, <https://doi.org/10.1371/journal.pone.0149035> (2016).
30. Canonne, J. & Rivas, S. Bacterial effectors target the plant cell nucleus to subvert host transcription. *Plant Signal Behav* 7, 217–221, <https://doi.org/10.4161/psb.18885> (2012).
31. McLellan, H. *et al.* An RxLR effector from Phytophthora infestans prevents re-localisation of two plant NAC transcription factors from the endoplasmic reticulum to the nucleus. *PLoS Pathog* 9, e1003670, <https://doi.org/10.1371/journal.ppat.1003670> (2013).
32. Rennoll-Bankert, K. E., Garcia-Garcia, J. C., Sinclair, S. H. & Dummer, J. S. Chromatin-bound bacterial effector ankyrin A recruits histone deacetylase 1 and modifies host gene expression. *Cell Microbiol* 17, 1640–1652, <https://doi.org/10.1111/cmi.12461> (2015).
33. Pandey, S. P. & Somssich, I. E. The Role of WRKY Transcription Factors in Plant Immunity. *Plant Physiology* 150, 1648–1655, <https://doi.org/10.1104/pp.109.138990> (2009).
34. Gigolashvili, T. *et al.* The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in Arabidopsis thaliana. *Plant J* 50, 886–901, <https://doi.org/10.1111/j.1365-3113.2007.03099.x> (2007).
35. Varet, A., Hause, B., Hause, G., Scheel, D. & Lee, J. The Arabidopsis NHL3 gene encodes a plasma membrane protein and its overexpression correlates with increased resistance to Pseudomonas syringae pv. tomato DC3000. *Plant Physiol* 132, 2023–2033 (2003).
36. Mohr, T. J. *et al.* The Arabidopsis downy mildew resistance gene RPP8 is induced by pathogens and salicylic acid and is regulated by W box cis elements. *Mol Plant Microbe Interact* 23, 1303–1315, <https://doi.org/10.1094/mpmi-01-10-0022> (2010).
37. Yoshida, S., Ito, M., Nishida, I. & Watanabe, A. Isolation and RNA gel blot analysis of genes that could serve as potential molecular markers for leaf senescence in Arabidopsis thaliana. *Plant Cell Physiol* 42, 170–178 (2001).
38. Atkinson, N. J., Lilley, C. J. & Urwin, P. E. Identification of genes involved in the response of Arabidopsis to simultaneous biotic and abiotic stresses. *Plant Physiol* 162, 2028–2041, <https://doi.org/10.1104/pp.113.222372> (2013).
39. Chen, K., Fan, B., Du, L. & Chen, Z. Activation of hypersensitive cell death by pathogen-induced receptor-like protein kinases from Arabidopsis. *Plant Mol Biol* 56, 271–283, <https://doi.org/10.1007/s11103-004-3381-2> (2004).
40. Demianski, A. J., Chung, K. M. & Kunkel, B. N. Analysis of Arabidopsis JAZ gene expression during Pseudomonas syringae pathogenesis. *Mol Plant Pathol* 13, 46–57, <https://doi.org/10.1111/j.1364-3703.2011.00727.x> (2012).
41. Sun, J. *et al.* Arabidopsis ASA1 Is Important for Jasmonate-Mediated Regulation of Auxin Biosynthesis and Transport during Lateral Root Formation. *The Plant Cell* 21, 1495–1511, <https://doi.org/10.1105/tpc.108.064303> (2009).
42. Fu, Z. Q. *et al.* NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. *Nature* 486, 228–232, <https://doi.org/10.1038/nature11162> (2012).
43. Song, T. *et al.* An Oomycete CRN Effector Reprograms Expression of Plant HSP Genes by Targeting their Promoters. *PLoS Pathog* 11, e1005348, <https://doi.org/10.1371/journal.ppat.1005348> (2015).
44. Li, X., Clarke, J. D., Zhang, Y. & Dong, X. Activation of an EDS1-mediated R-gene pathway in the snc1 mutant leads to constitutive, NPR1-independent pathogen resistance. *Mol Plant Microbe Interact* 14, 1131–1139, <https://doi.org/10.1094/MPMI.2001.14.10.1131> (2001).
45. Abe, M. *et al.* WAVY LEAF1, an ortholog of Arabidopsis HEN1, regulates shoot development by maintaining MicroRNA and trans-acting small interfering RNA accumulation in rice. *Plant Physiol* 154, 1335–1346, <https://doi.org/10.1104/pp.110.160234> (2010).
46. Salvetti, A. & Greco, A. Viruses and the nucleolus: the fatal attraction. *Biochim Biophys Acta* 1842, 840–847, <https://doi.org/10.1016/j.bbdis.2013.12.010> (2014).
47. Hiscox, J. A. The nucleolus—a gateway to viral infection? *Arch Virol* 147, 1077–1089, <https://doi.org/10.1007/s00705-001-0792-0> (2002).
48. Hiscox, J. A. RNA viruses: hijacking the dynamic nucleolus. *Nat Rev Microbiol* 5, 119–127, <https://doi.org/10.1038/nrmicro1597> (2007).
49. Stam, R. *et al.* Identification and Characterisation CRN Effectors in Phytophthora capsici Shows Modularity and Functional Diversity. *PLOS ONE* 8, e59517, <https://doi.org/10.1371/journal.pone.0059517> (2013).
50. Dean, P. *et al.* The Enteropathogenic E. coli Effector EspF Targets and Disrupts the Nucleolus by a Process Regulated by Mitochondrial Dysfunction. *PLOS Pathogens* 6, e1000961, <https://doi.org/10.1371/journal.ppat.1000961> (2010).
51. Alfano, J. R. *et al.* The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proc Natl Acad Sci USA* 97, 4856–4861 (2000).
52. Li, X., Zhang, Y., Clarke, J. D., Li, Y. & Dong, X. Identification and cloning of a negative regulator of systemic acquired resistance, SN1, through a screen for suppressors of npr1-1. *Cell* 98, 329–339 (1999).
53. Karimi, M., Inzé, D. & Depicker, A. GATEWAY™ vectors for Agrobacterium-mediated plant transformation. *Trends in Plant Science* 7, 193–195, [https://doi.org/10.1016/S1360-1385\(02\)02251-3](https://doi.org/10.1016/S1360-1385(02)02251-3) (2002).
54. Rentel, M. C., Leonelli, L., Dahlbeck, D., Zhao, B. & Staskawicz, B. J. Recognition of the Hyaloperonospora parasitica effector ATR13 triggers resistance against oomycete, bacterial, and viral pathogens. *Proc Natl Acad Sci USA* 105, 1091–1096, <https://doi.org/10.1073/pnas.0711215105> (2008).
55. Sparkes, I. A., Runions, J., Kearns, A. & Hawes, C. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protocols* 1, 2019–2025 (2006).
56. Yamaguchi, N. *et al.* PROTOCOLS: Chromatin Immunoprecipitation from Arabidopsis Tissues. *Arabidopsis Book* 12, e0170, <https://doi.org/10.1199/tab.0170> (2014).
57. Kass, J., Artero, R. & Baylies, M. K. Non-radioactive electrophoretic mobility shift assay using digoxigenin-ddUTP labeled probes. *Drosophila Information Service* 83, 185–188 (2000).
58. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498–2504, <https://doi.org/10.1101/gr.1239303> (2003).
59. Bindea, G., Galon, J. & Mlecnik, B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics* 29, 661–663, <https://doi.org/10.1093/bioinformatics/btt019> (2013).
60. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2– $\Delta\Delta CT$ Method. *Methods* 25, 402–408, <https://doi.org/10.1006/meth.2001.1262> (2001).

61. Steffens, N. O., Galuschka, C., Schindler, M., Bulow, L. & Hehl, R. AthaMap web tools for database-assisted identification of combinatorial cis-regulatory elements and the display of highly conserved transcription factor binding sites in *Arabidopsis thaliana*. *Nucleic Acids Res* **33**, W397–402, <https://doi.org/10.1093/nar/gki395> (2005).
62. Zambelli, F., Pesole, G. & Pavesi, G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res* **37**, W247–252, <https://doi.org/10.1093/nar/gkp464> (2009).
63. Chang, W. C., Lee, T. Y., Huang, H. D., Huang, H. Y. & Pan, R. L. PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. *BMC Genomics* **9**, 561, <https://doi.org/10.1186/1471-2164-9-561> (2008).
64. Si, J., Zhang, Z., Lin, B., Schroeder, M. & Huang, B. MetaDBSite: a meta approach to improve protein DNA-binding sites prediction. *BMC Systems Biology* **5**, 1–7, <https://doi.org/10.1186/1752-0509-5-s1-s7> (2011).
65. <Jingna_2011-MetaDBsite.pdf>.

Acknowledgements

Funding for the project was provided by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants to HG. The project in HG's laboratory was also partially funded by an institutional research chair held by HG. BA was funded by an international PhD scholarship from the Fonds de Recherche du Québec sur la Nature et les Technologies (FRQNT). Travel grants from the Conseil Franco-Québécois de Coopération Universitaire (CFQCU) were also awarded to BA, KCGS and CL. IBS. obtained a mobility grant from MITACS. BP was supported by an Institut National de la Recherche Agronomique (INRA) Contrat Jeune Scientifique, by the European Union, as part of the Marie-Curie FP7 COFUND People Program, through an AgreenSkills' fellowship (under grant agreement No. 267196), and by the Laboratory of Excellence ARBRE, through a mobility grant (12RW53). CL was supported by INRA in the framework of a Contrat Jeune Scientifique and by the Labex ARBRE (ANR-11-Labex-002-01, Lab of Excellence ARBRE). KCGS was supported by the MITACS-Globalink scholarship program. SD is supported by the French National Research Agency through the Laboratory of Excellence ARBRE (ANR-12-LABXARBRE-01) and the Young Scientist Grant POPRUST (ANR-2010-JCJC-1709-01).

Author Contributions

B.A., K.C.G.S., I.B.S., M.B.P., B.P. performed the experiments and generated biological material; B.A., B.P., C.L., M.B.P., S.D., I.D.P. and G.H., analyzed the results; B.A., C.L., B. P., M.B.P., S.D., I.D.P. and H.G. wrote and/or edited the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-32825-0>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2018