



HAL
open science

Le locus de résistance Ma des Prunus vis-à-vis des nématodes à galles : Originalité structurale et évolution dans la famille des NBS-LRRs chez les plantes

Cyril van Ghelder

► **To cite this version:**

Cyril van Ghelder. Le locus de résistance Ma des Prunus vis-à-vis des nématodes à galles : Originalité structurale et évolution dans la famille des NBS-LRRs chez les plantes. Biologie moléculaire. Université Côte d'Azur, 2019. Français. NNT : 2019AZUR6006 . tel-02517633

HAL Id: tel-02517633

<https://theses.hal.science/tel-02517633>

Submitted on 24 Mar 2020

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

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

THÈSE DE DOCTORAT

Le locus de résistance *Ma* des *Prunus* vis-à-vis des
nématodes à galles : Originalité structurale et évolution
dans la famille des NBS-LRRs chez les plantes

Cyril VAN GHELDER

UMR ISA INRA 1355 – UNS – CNRS 7254

**Présentée en vue de l'obtention
du grade de docteur en**
Sciences de la Vie et de la Santé
mention interactions moléculaires et
cellulaires de l'Université Côte d'Azur

Dirigée par : Daniel ESMENJAUD,
IRHC, HDR, UMR ISA

Soutenu le : 15 Octobre 2019

Devant le jury, composé de :

M. Pierre FRENDO, Professeur, UNS INRA CNRS ISA	Président
M. Thomas KROJ, DR2, UMR BGPI, Montpellier	Rapporteur
M. Olivier LEMAIRE, DR2, UMR SVQV, Colmar	Rapporteur
Mme Marie-Claire KERLAN, IR1, UMR IGEPP, Rennes	Examineur
M. Jean-Luc POESSEL, CRHC, INRA, UR GAFL, Avignon	Examineur
M. John MACKAY, Professeur, Université d'Oxford, Royaume-Uni	Membre invité

Le locus de résistance *Ma* des *Prunus* vis-à-vis des
nématodes à galles : Originalité structurale et évolution
dans la famille des NBS-LRRs chez les plantes

Jury :

Président du jury

M. Pierre FRENDO, Professeur, UNS INRA CNRS ISA

Rapporteurs

M. Thomas KROJ, DR2, UMR BGPI, Montpellier

M. Olivier LEMAIRE, DR2, UMR SVQV, Colmar

Examineurs

Mme Marie-Claire KERLAN, IR1, UMR IGEPP, Rennes

M. Jean-Luc POESSEL, CRHC, INRA, UR GAFL, Avignon

Invité

M. John MACKAY, Professeur, Université d'Oxford, UK

Titre : Le locus de résistance *Ma* des *Prunus* vis-à-vis des nématodes à galles : Originalité structurale et évolution dans la famille des NBS-LRRs chez les plantes

Résumé

Les nématodes à galles, *Meloidogyne* spp., sont des ravageurs extrêmement polyphages qui, à l'échelle mondiale, occasionnent de graves dommages aux plantes et tout spécialement aux espèces pérennes. La résistance génétique spécifique des plantes s'appuie principalement sur les gènes de la famille des récepteurs NBS-LRR (ou NLRs, regroupant les sous-familles des TNLs, CNLs et RNLs) qui sont des facteurs-clefs du contrôle des maladies et ravageurs. Chez les *Prunus*, le gène *Ma* du prunier et le gène *RMja* de l'amandier, tous deux localisés sur le chromosome 7, expriment différents spectres de résistance (R) vis-à-vis des nématodes à galles. Le gène *Ma* confère la résistance à toutes les espèces de *Meloidogyne* testées et montre une structure TNL particulière due à sa région C-terminale constituée de cinq domaines répétés, désignés domaines post-LRRs (PLs). Dans ce contexte, notre travail de thèse a caractérisé l'originalité et la distribution de ce domaine PL au sein des gènes NBS-LRRs chez les *Prunus* et les autres plantes terrestres et a identifié la relation génétique entre les gènes *Ma* et *RMja*. Nous avons tout d'abord étudié la fréquence, la distribution et les caractéristiques structurales des i) gènes TNLs et ii) des domaines PL dans le génome du pêcher, génome de référence des Rosaceae. Le domaine PL a été détecté dans deux tiers des 195 gènes TNL identifiés sur les huit chromosomes du pêcher. Nous avons pu utiliser la séquence consensus de 124 domaines PL pour définir des motifs spécifiques permettant d'augmenter significativement le nombre de domaines PL et de gènes TNLs associés détectés dans divers génomes de Dicotylédones. Par une approche d'exploration de divers génomes nous avons montré que les TNLs disposant de domaines PL multiples sont rares chez le pêcher et les autres plantes terrestres étudiées. Le modèle à cinq domaines répétés est probablement unique à *Ma* et à ses orthologues chez les *Prunus* et les genres étroitement apparentés des Rosaceae. Cette structure a vraisemblablement été héritée de leur ancêtre commun dans l'ordre des Rosales. Il ressort de nos travaux que le domaine PL, présent dans une large majorité des TNLs des Angiospermes, est spécifique de cette sous-famille de gènes.

Nous avons ensuite recherché si les TNLs et leurs PLs associés étaient présents chez les Gymnospermes, en nous appuyant sur leurs représentants les plus importants économiquement,

les conifères. Jusqu'à présent, chez les espèces de ce groupe taxonomique, la constitution de l'arsenal de NBS-LRRs et des contingents de TNLs, CNLs et RNLs n'étaient pas claires en raison de la grande taille et de la complexité de leurs génomes. Afin d'éviter cet écueil, nous avons identifié et analysé plus de 3800 séquences NBS-LRR exprimées issus de 7 transcriptomes de référence développés à partir de sept espèces de conifères. Aucun des motifs précédemment définis chez les Angiospermes n'a été retrouvé à partir de l'ensemble des peptides prédits et aucune signature solide n'a pu être définie dans les régions en aval des domaines LRR présents dans les séquences TNLs. Cependant, alors que nous avons trouvé des motifs PL complets dans l'Angiosperme ayant divergé précocement *Amborella trichopoda*, nous n'avons trouvé que des motifs partiels chez le Gymnosperme ancien *Ginkgo biloba*. Ces observations suggèrent une acquisition partielle précoce chez les plantes à graines et une expansion adaptative chez les plantes à fleurs. En complément, nous avons montré que les conifères, tout comme les Rosaceae, possèdent de nombreux RNLs et TNLs. En étendant notre étude à diverses plantes terrestres, nous avons mis en évidence un rapport moyen de 1:10 entre effectifs de RNLs et de TNLs.

Nous avons finalement conduit une cartographie haute résolution et localisé avec précision le gène *RMja* chez l'amandier. En nous appuyant sur une banque BAC, *RMja* a été localisé dans le cluster de résistance *Ma* et l'orthologue de *Ma* est de très loin le meilleur candidat. Cette co-localisation ouvre la voie à une comparaison du polymorphisme de séquence chez trois régions orthologues du locus *Ma*, dans le prunier (spectre R complet), l'amandier (spectre R incomplet) et le pêcher (spectre R nul). Nous avons mis en évidence que le cluster de TNLs contenant *Ma* contient des orthologues de ce gène qui ont une structure conservée unique. Nos résultats suggèrent que le polymorphisme des répétitions du domaine PL sous-tend des interactions différentielles de résistance vis-à-vis des nématodes à galles et un mécanisme d'immunité original chez les plantes pérennes. Dans ces processus immuns de reconnaissance ou de signalisation, d'autres composants tels que les RNLs et/ou récepteur lectine kinases pourraient être impliqués. Notre travail ouvre la voie à des approches comparative et fonctionnelle d'identification des déterminants moléculaires impliqués dans la résistance aux nématodes à galles chez les *Prunus*.

Mots-clés : cartographie, domaine PL, évolution, gènes de résistance, interaction plante-nématode, *Meloidogyne*, NBS-LRR, NLR, *Prunus*, TNL

Title: The *Ma* resistance locus to root-knot nematodes in *Prunus*: Structural originality and evolution within the NBS-LRR gene family in plants**Abstract**

Root-knot nematodes (RKNs), *Meloidogyne spp.*, are extremely polyphagous pests that severely challenge plants worldwide and especially perennials. The specific genetic resistance of plants mainly relies on genes of the family of NBS-LRR receptors (or NLRs grouping TNL, CNL and RNL subfamilies) that are pivotal factors for control of pests and pathogens. In *Prunus spp.*, the *Ma* plum and *RMja* almond genes, both located on the chromosome 7, express different spectra for resistance (R) to RKNs. The *Ma* gene confers resistance to all RKNs tested and displays a peculiar TNL structure due to a C-terminal region designated post-LRR (PL) region and comprising five repeated domains. In this context, this thesis work has characterized the originality of this PL domain within *Prunus* and other land plants and has revealed the genetic relationship between the *Ma* and *RMja* genes.

We first studied the frequency, distribution and structural characteristics of i) TNL genes and ii) their PL domains in the peach genome, the reference genome for Rosaceae. Among the 195 predicted TNLs from the eight peach chromosomes, it was possible to retrieve the PL domains among two thirds of the genes. We could use consensus sequences defined on the basis of 124 PL domains to design specific motifs, and we found that the use of these motifs significantly increased the numbers of PL domains and correlative TNL genes detected in diverse Dicot genomes. Based on PL signatures, a genome mining approach showed that TNL genes with multiple PL domains are rare in peach and the other plants that have been tested. The five-PL domain pattern is probably unique to *Ma* and its orthologues within *Prunus* and closely related genera from the Rosaceae and was probably inherited from their common ancestor in the order Rosales. We showed that the PL domain, present in the large majority of TNLs among Angiosperms, are specific to this gene subfamily.

We then investigated whether TNLs and their associated PL domains were present in Gymnosperms through their most economically important representatives, the conifers. Indeed, the origin, evolution and distribution of NBS-LRR sequences and of the contingents of TNLs, CNLs and RNLs have been unclear in the species of this taxonomic group, partly due to their large and complex genomes. Consequently, we identified and analysed a resource of over 3800

expressed NBS-LRR sequences from reference transcriptomes of seven conifer species. None of the PL motifs previously defined in Angiosperms were detected from the entire dataset of predicted peptides. Conversely, we found complete PL motifs in the early diverged Angiosperm *Amborella trichopoda*, and a few partial ones in the ancient Gymnosperm *Ginkgo biloba*. These observations suggest an early partial acquisition in seed plants and an adaptive expansion in flowering plants. In complement, we showed that conifers and Rosaceae have numerous RNLs and TNLs. By enlarging our study to other land plant genomes, we uncovered an average ratio of 1:10 between RNLs and TNLs numbers.

We finally carried out a high-resolution mapping and localised precisely the *RMja* gene in almond. Using a BAC library, *RMja* mapped to the *Ma* resistance cluster and the *Ma* orthologue is by far the best candidate. This co-localisation opens the way for comparison of sequence polymorphism within three orthologous regions of the *Ma* locus, i.e. plum (complete R spectrum), almond (incomplete R spectrum) and peach (null R spectrum). We showed that the *Ma* TNL cluster has evolved *Ma* orthologues with a unique conserved structure. Our results suggest that the polymorphism contained in the PL-domain repeats might underlie differential resistance interactions with RKNs and an original immune mechanism in woody perennials. In these immune processes for recognition and signalling, other components such as RNLs or receptor lectin kinases might be involved. This work paves the way for future comparative and functional approaches aiming to unravel the molecular determinants involved in the resistance to RKNs in *Prunus*.

Keywords: evolution, gene mapping, *Meloidogyne*, NBS-LRR, NLR, plant-nematode interaction, PL domain, *Prunus*, resistance genes, TNL

REMERCIEMENTS

Je remercie Daniel Esmenjaud pour toutes ces années de travail ensemble, sa disponibilité et pour ses belles qualités humaines. Merci de m'avoir fait profiter de votre connaissance des plantes et des nématodes. Je remercie Pierre Abad pour la confiance qu'il m'a témoignée en me permettant de poursuivre ce projet de thèse et pour son soutien concernant les projets menés hors du site de Sophia Antipolis. Je remercie tous les membres de l'équipe IPN présente ou passée : Roger, Ulysse, Chantal, Michaël, Etienne, Nathalie, Joffrey, Bruno, Corinne, ... et tous les autres. Je remercie plus particulièrement Ariane Fazari pour son aide précieuse et son amitié. Je remercie également mes collègues de l'ISA qui ont contribué directement ou indirectement à ce projet : Emeline Deleury, Benoît Industri, Laurent Lapeyre, Michel Ponchet, Frédéric Potier, ...

Merci à l'équipe de culture in vitro passée et présente : Christine Poncet et Lydia Ottenwaelder pour leur précieux soutien, Rolande Poupet et Annie Henri mes collègues et amies. Ce travail est également dédié à Claudine Andréoli et Jean-Paul Onesto qui ont beaucoup compté dans mon évolution professionnelle.

Merci à Henri Duval, pour son expertise en matière d'arbres fruitiers et tous ceux de l'INRA d'Avignon, du CNRGV et de MGX qui ont participé activement à ce travail.

Je remercie particulièrement les équipes de l'Université d'Oxford et de l'Université de Laval avec lesquelles j'ai eu la chance de collaborer. Un grand merci à John MacKay et Geneviève Parent au contact desquels j'ai énormément appris et dont l'humilité, la bonne humeur et l'excellence scientifique m'ont beaucoup marqué.

Merci à Jean-Luc Poëssel et Ignazio Verde pour les discussions particulièrement intéressantes autour des *Prunus*, des génomes et des gènes de résistance.

Enfin, je dédie ce travail à mes parents et grands-parents, ainsi qu'à Julia sans qui rien de tout ceci n'aurait été possible et qui m'accompagne et me soutient depuis plus de 10 ans. Merci de croire en moi.

A Julia...

TABLE DES MATIERES

RESUME	II
ABSTRACT	IV
REMERCIEMENTS	6
TABLE DES MATIERES	10
LISTE DES ABREVIATIONS	14
LISTE DES TABLEAUX, FIGURES ET ANNEXES	16
PUBLICATIONS	18
INTRODUCTION GENERALE	20
CHAPITRE 1 : CONTEXTE GENERAL	24
1.1 LA REPONSE IMMUNE DES PLANTES : UN SYSTEME A DOUBLE DETENTE	24
1.2 LES NBS-LRRs	26
1.2.1 LA STRUCTURE DES NBS-LRRs	26
1.2.2 LES DIFFERENTS TYPES DE PERCEPTION	33
1.2.3 LES RESEAUX DE NBS-LRRs ET LE 'RESISTOSOME' COMME NOUVEAUX PARADIGMES	36
1.2.4 LES VOIES DE SIGNALISATION ET LA LOCALISATION DES NBS-LRRs	38
1.2.5 LE NIVEAU D'EXPRESSION ET LA REGULATION DES NBS-LRRs	40
1.2.6 L'EVOLUTION DES NBS-LRRs	43
1.3 LES NEMATODES	50
1.3.1 LES NEMATODES PHYTOPARASITES	50
1.3.2 LES NEMATODES A GALLES DU GENRE <i>MELOIDOGYNE</i>	52
1.3.3 LE CYCLE DE VIE ET L'ANATOMIE DES <i>MELOIDOGYNE</i>	52
1.3.4 LA REPRODUCTION ET L'ADAPTATION DES <i>MELOIDOGYNE</i>	54
1.3.5 LES EFFECTEURS DE NEMATODES : MANIPULATEURS DE CELLULES	54
1.4 LE GENRE <i>PRUNUS</i>	56
1.4.1 CLASSIFICATION ET POSITIONNEMENT DES <i>PRUNUS</i>	56
1.4.2 AIRE D'ORIGINE	59
1.4.3 INTRA ET INTER-COMPATIBILITES	59
1.4.4 DONNEES GENOMIQUES	60
1.5 <i>PRUNUS</i> – <i>MELOIDOGYNE</i> : MODELE POUR L'INTERACTION PLANTE PERENNE LIGNEUSE-NEMATODE A GALLES	60
1.5.1 REACTION COMPATIBLE	60
1.5.2 LA LUTTE CONTRE LES NEMATODES A GALLES EN VERGER	61
1.5.3 REACTION INCOMPATIBLE <i>PRUNUS</i> - <i>MELOIDOGYNE</i>	61
1.5.4 LES GENES DE RESISTANCE AUX <i>MELOIDOGYNE</i> CHEZ LES <i>PRUNUS</i>	64

1.5.5	LE GENE <i>MA</i> : UN GENE PARTICULIER	66
1.5.6	OUVERTURE	68
CHAPITRE 2 : MATERIEL ET METHODES		72
2.1	MATERIEL VEGETAL	72
2.2	MATERIEL ANIMAL	73
2.3	EXTRACTIONS D'ACIDES NUCLEIQUES	73
2.3.1	EXTRACTION D'ADN DE <i>PRUNUS</i>	73
2.3.2	EXTRACTION D'ARN TOTAUX DE <i>PRUNUS</i>	74
2.3.3	SYNTHESE DES ADN COMPLEMENTAIRES (ADNC) ET CLONAGE POUR SEQUENÇAGE	74
2.4	EXPRESSION DU GENE <i>MA</i>	75
2.5	GENOTYPAGE	75
2.6	PHENOTYPAGE	76
2.7	ANALYSE DE SEQUENCES	76
2.8	DONNEES ADDITIONNELLES ET BASES DE DONNEES	77
CHAPITRE 3 : LES TNLS DANS LE GENOME DU PECHER ET LA CARACTERISATION DU DOMAINE POST-LRR		80
3.1	INTRODUCTION	80
3.2	ARTICLE PUBLIE DANS BMC GENOMICS	81
CHAPITRE 4 : LES NBS-LRRS CHEZ LES CONIFERES ET TRAIT EVOLUTIFS DE CETTE FAMILLE MULTIGENIQUE CHEZ LES PLANTES TERRESTRES		100
4.1	INTRODUCTION	100
4.2	ARTICLE PUBLIE DANS SCIENTIFIC REPORTS	103
CHAPITRE 5 : CARACTERISATION DU GENE <i>RMJA</i> CHEZ L'AMANDIER		120
5.1	INTRODUCTION	120
5.2	ARTICLE PUBLIE DANS FRONTIERS IN PLANT SCIENCE	122
CHAPITRE 6 : DONNEES COMPLEMENTAIRES RELATIVES AUX MECANISMES ET DETERMINANTS DE LA RESISTANCE CONFERE PAR LE GENE <i>MA</i>		140
6.1	INTRODUCTION	140
6.2	DONNEES D'EXPRESSION DU GENE <i>MA</i> CHEZ LE CLONE P.2175 (<i>P. CERASIFERA</i>) ET EN SYSTEME HETEROLOGUE	140
6.3	EXPLOITATION DE LA DIVERSITE DU PRUNIER MYROBOLAN	144
CHAPITRE 7 : CONCLUSION GENERALE ET PERSPECTIVES		152
7.1	CONCLUSION GENERALE	152
7.1.1	ORIGINALITE DU DOMAINE PL AU SEIN DES NBS-LRRS	152
7.1.2	IMPLICATION DU (DES) DOMAINE(S) PL(S) DANS LA RESISTANCE	153
7.1.3	ORIGINE ET EVOLUTION DES NBS-LRRS ET DU DOMAINE PL	156

7.2	PERSPECTIVES	159
7.2.1	EXPLOITATION DES SEQUENCES ALLELIQUES ET DES DONNEES D'EXPRESSION	159
7.2.2	MODIFICATION DE LA REGION PL CHEZ <i>Ma</i> (APPROCHE MECANISTIQUE)	159
7.2.3	RECHERCHE DES PARTENAIRES MOLECULAIRES DU GENE <i>Ma</i>	160
7.2.4	UTILISATION DES TNLS MODELES AFIN DE COMPRENDRE LE ROLE DU DOMAINE PL	160
7.2.5	APPLICATIONS	161
REFERENCES BIBLIOGRAPHIQUES		164
ANNEXE 1		176
ANNEXE 2		178
ANNEXE 3		180
ANNEXE 4		184
ANNEXE 5		190
ANNEXE 6		194
ANNEXE 7		196
ANNEXE 8		198

LISTE DES ABBREVIATIONS

Bs2	Bacterial Spot disease resistance in pepper
Bs4	Bacterial Spot disease resistance in tomato
CC	Coiled-coil
CNL	CC-NBS-LRR
EDS1	Enhanced Disease Susceptibility 1
ETI	Effector-Triggered Immunity
HR	Hypersensitive Response
HSP90	Heat Shock Protein 90kDa
ID	Integrated Domain
LRR	Leucine-Rich Repeat
MAPK	Mitogen-Associated Protein Kinase
Mi-1.2	Root-knot nematode resistance in tomato
miRNA	Micro RNA
NB-ARC	Nucleotide-Binding adaptor shared by APAF-1, R proteins, and CED-4
NBS	Nucleotide-Binding Site
NDR1	Non-race specific Disease Resistance
NLL	NBS LRR Linker
NLR	NOD-Like Receptors
NMD	Nonsense Mediated Decay
NOD	Nucleotide Oligomerization Domain
NRC	NLR Required for Cell death
PAD4	PhytoAlexin Deficient 4
PCD	Programmed Cell Death
PTI	Pathogen-associated molecular pattern Triggered Immunity
Pto	Resistance to <i>Pseudomonas syringae</i> in tomato
R1, R8	Resistance to <i>Phytophthora infestans</i>
RAR1	Required for Mla12 Resistance 1
RIN4	RPM1-INTERACTING protein 4
RLK	Receptor Lectin Kinase
RLP	Receptor Like-Protein
RNL	RPW8-NBS-LRR
ROS	Reactive Oxygen Species
Rpi-blb2	Late blight resistance protein <i>Solanum bulbocastanum</i>
RPS4	Resistance to <i>Pseudomonas syringae</i>
RPW8	Resistance to Powdery mildew 8
RRS1	Resistance to <i>Ralstonia Solanacearum</i> 1
SA	Salicylic Acid
SAG101	Senescence-associated carboxylesterase 101
SGT1	Suppressor of G2 allele of <i>skp1</i>
siRNA	Small Interfering RNA
SRFR1	SUPPRESSOR OF <i>rps4</i> -RLD
STAND	Signal-Transduction ATPase with Numerous Domains

Sw5b	Protein spotted wilt resistance-5
TIR	Toll/Interleukin-1 Receptor
TMV	<i>Tobacco mosaic virus</i>
TNL	TIR-NBS-LRR
VIGS	Virus-Induced Gene Silencing

LISTE DES TABLEAUX, FIGURES ET ANNEXES

TABLEAU

Tableau 1 - Productions et surfaces fruitières cultivées mondiales des principales espèces de <i>Prunus</i> . Données FAO – 2016.....	56
---	----

FIGURES

Figure 1 - Représentation schématique des deux voies principales de défense, PTI et ETI, induites chez la plante pour faire face aux agents pathogènes.	25
Figure 2 - Structure tridimensionnelle, prédite par le serveur i-tasser, du domaine NBS de la protéine Ma.....	28
Figure 3 - Structures tridimensionnelles, prédites par le serveur i-tasser, des domaines LRR et TIR de la protéine Ma.	30
Figure 4 - Deux systèmes de surveillance entre NBS-LRRs et protéines kinases illustrant les stratégies de leurre et de garde.	34
Figure 5 - Mécanisme de défense engagé par la paire de TNLs RPS4/RRS1 incluant une stratégie de leurre intégré.	36
Figure 6 - Réseau fonctionnel entre TNLs et RNLs dans le déclenchement de la réponse immune chez les plantes.	37
Figure 7 - Représentation schématique des voies de signalisation des NBS-LRRs.	39
Figure 8 - Annotation structurale de différents TNLs caractérisés avec leur rôle identifié.	42
Figure 9 - Origine et organisation structurale des NBS-LRRs au cours de l'évolution des plantes.....	43
Figure 10 - Relation entre le nombre de NBS-LRRs pour une espèce donnée et la taille de son génome.	45
Figure 11 - Variation dans les estimations de NBS-LRRs et distribution des NBS-LRRs dans divers génomes de plantes.....	46
Figure 12 - Mécanismes génétiques intervenant dans l'évolution des gènes de résistance de type NBS-LRR.	48
Figure 13 - Quelques conséquences possibles sur le phénotype exprimé par les plantes suite aux mécanismes génétiques mis en jeu par les NBS-LRRs.	49
Figure 14 - Représentation schématique du phylum Nematoda et position des nématodes à galles du genre <i>Meloidogyne</i>	51
Figure 15 - Cycle de vie de <i>Meloidogyne incognita</i> sur racines de <i>Prunus cerasifera</i>	53
Figure 16 - Position des <i>Prunus</i> dans la classification simplifiée des plantes terrestres et leurs routes évolutives.	58
Figure 17 - L'interaction <i>Meloidogyne-Prunus</i>	63
Figure 18 - Spectres (A) et caractéristiques (B) de la résistance conférée par les gènes identifiés vis-à-vis des <i>Meloidogyne</i> chez les <i>Prunus</i>	65
Figure 19 - Annotation fonctionnelle de la protéine prédite Ma et de quelques TNLs de référence. ...	67
Figure 20 - Distribution des TNLs chez 56 génomes et transcriptomes du règne végétal.	102

Figure 21 - Logigramme décrivant la stratégie de localisation et identification du gène <i>RMja</i> chez l'amandier <i>P. dulcis</i> cv. 'Alnem 1' .	121
Figure 22 - Expression du gène <i>Ma</i> chez le prunier myrobolan P.2175 et en système hétérologue chez la tomate.	142
Figure 23 – Carte illustrant la diversité géographique de la collection de pruniers myrobolans phénotypés vis-à-vis de <i>Meloidogyne</i> spp.	144
Figure 24 - Méthodologie d'analyse de la diversité allélique du gène <i>Ma</i> .	145
Figure 25 - Polymorphisme observé au niveau des domaines PL3, PL4 et PL5 au sein de quelques variétés de la collection allélique du prunier myrobolan et de variétés d'autres espèces de <i>Prunus</i> diploïdes.	148
Figure 26 - Diagramme regroupant les données et hypothèses concernant le domaine PL chez les TNLs.	158

ANNEXES

ANNEXE 1 – Liste des différents marqueurs moléculaires (amorces) utilisés dans le marquage fin du gène *RMja*.

ANNEXE 2 – Liste des logiciels utilisés dans les différentes études présentées dans ce manuscrit.

ANNEXE 3 – Liste détaillée des bases de données dont sont issues les données additionnelles utilisées dans les différentes études du manuscrit.

ANNEXE 4 – Liste détaillée des tailles de génomes, des effectifs de NBS-LRRs et de leurs taux de clustérisation et références associées.

ANNEXE 5 – Alignement de la zone allant du domaine LRR (bleu) au domaine PL5 chez *Ma*, *RMja* et deux isoformes détectées chez *Prupe.7G65400*.

ANNEXE 6 – Alignement des domaines PL de *RPS4*, *N* et *Ma* avec leur structure secondaire prédite (S2ndP).

ANNEXE 7 – Alignement de quelques signatures Post-LRR (motifs 1, 2 et 3) chez *Amborella trichopoda*, *Ginkgo biloba* et cinq TNLs de référence (*N*, *Bs4*, *RPS4*, *Gro1-4* et *Ma*) chez les Dicotylédones.

ANNEXE 8 – Identification de signatures caractéristiques du motif 1 du domaine PL chez des protéines kinases des plantes archaïques *Physcomitrella patens*, *Sphagnum fallax*, *Amborella trichopoda* et *Aquilegia caerulea*.

PUBLICATIONS

Ce manuscrit comprend des éléments issus des travaux publiés ci-dessous.

Saucet SB, Van Ghelder C, Abad P, Duval H and Esmenjaud D (2016) **Resistance to root-knot nematodes *Meloidogyne* spp. in woody plants.** *New Phytologist Tansley Review*. **211**:41-56. doi: 10.1111/nph.13933.

Van Ghelder C and Esmenjaud D (2016) **TNL genes in peach: insights into the post-LRR domain.** *BMC Genomics*. 2016 Apr 30;17:317. doi: 10.1186/s12864-016-2635-0.

Van Ghelder C, Esmenjaud D, Callot C, Dubois E, Mazier M and Duval H. (2018) ***Ma* orthologous genes in *Prunus* spp. shed light on a noteworthy NBS-LRR cluster conferring differential resistance to root-knot nematodes.** *Frontiers in Plant Sciences*. **9**:1269. doi: 10.3389/fpls.2018.01269.

Duval H., Van Ghelder C., Portier U., Confolent C., Meza P. and Esmenjaud D. (2019) **New data completing the spectrum of the *Ma*, *RMia*, and *RMja* genes for resistance to root-knot nematodes (*Meloidogyne* spp.) in *Prunus*.** *Phytopathology* **109**(4):615-622. doi: 10.1094/PHYTO-05-18-0173-R.

Van Ghelder, C. Parent G.J., Rigault, P., Prunier, J., Giguère, I., Caron, S., Stival Sena, J., Deslauriers, A., Bousquet, J., Esmenjaud, D. and MacKay, J. (2019) **The large repertoire of conifer NLR resistance genes includes drought responsive and highly diversified RNLs.** *Scientific Reports* **9** :11614. doi: 10.1038/s41598-019-47950-7.

INTRODUCTION GENERALE

Les plantes sont constamment agressées par de nombreux stress d'origine biotique engendrés par divers ravageurs ou agents pathogènes. Parmi les ravageurs figurent les nématodes, vers ronds microscopiques qui sont des parasites telluriques obligatoires. L'un des groupes les plus nuisibles comprend les nématodes à galles du genre *Meloidogyne*, parmi lesquels trois espèces principales, *M. incognita*, *M. arenaria* et *M. javanica*, originaires des régions chaudes de la planète, constituent une grave menace pour l'agriculture par leur capacité à se développer sur plus de 5500 plantes différentes (Goodey *et al.*, 1965). A ces espèces majeures en expansion, notamment en raison du changement climatique et des échanges internationaux croissants, s'ajoutent des espèces émergentes invasives comme l'espèce tropicale *M. enterolobii*.

Avec l'abandon progressif des molécules chimiques du fait de leur toxicité, la lutte contre les nématodes devra reposer désormais sur des méthodes alternatives respectueuses de l'environnement telles que les méthodes culturales, la lutte biologique ainsi que l'utilisation de plantes antagonistes et de variétés résistantes. Etant donnée la très grande polyphagie des nématodes à galles, les variétés résistantes sont particulièrement prometteuses car les autres alternatives sont difficiles à mettre en œuvre.

La résistance naturelle trouve son origine dans l'immunité innée des plantes aux bio-agresseurs. En effet, celle-ci est portée au niveau cellulaire par des complexes protéiques capables de reconnaître spécifiquement des molécules du « non-soi » produites par les agents pathogènes ou ravageurs, et de déclencher une réponse de la plante susceptible de moduler drastiquement l'issue de l'infection. Hormis les défenses physiques mises en jeu par la plante, on trouve deux grands types de récepteurs protéiques impliqués dans cette reconnaissance et induisant la réponse immune de la plante : les récepteurs membranaires, qui agissent le plus souvent par reconnaissance de molécules génériques produites par les bio-agresseurs, et les récepteurs cytoplasmiques, qui effectuent une surveillance plus ciblée et spécifique. Cette dernière catégorie est un constituant majeur de la réponse immune des plantes et est principalement représentée par la grande famille des gènes de résistance (gènes *R*) de type NBS-LRR (nucleotide-binding site – leucine-rich repeat). Avec l'essor du séquençage de nouvelle génération et la disponibilité de nombreux génomes et transcriptomes de plantes, des efforts conséquents ont été consacrés à la localisation et à l'identification des gènes *R*.

Dans un contexte d'interaction continue entre la plante et ses bioagresseurs, tout particulièrement chez les plantes pérennes, la durabilité des gènes *R* est une composante

primordiale à évaluer. C'est notamment le cas chez les espèces pérennes fruitières à noyau du genre *Prunus* qui sont fortement touchées par les attaques des espèces prédominantes de *Meloidogyne*. Face à ces ravageurs, la stratégie de pyramidage de gènes *R* dans un même fond génétique est mise en œuvre afin d'éviter un contournement des résistances. Cette stratégie s'appuie sur les gènes *Ma*, *RMja* et *RMia* qui ont été identifiés, respectivement, chez le prunier myrobolan (*P. cerasifera*), l'amandier (*P. dulcis*) et le pêcher (*P. persica*), et qui portent des spectres de résistance différents. Le gène *Ma* appartenant à la famille des NBS-LRRs confère une résistance totale à l'ensemble des espèces de *Meloidogyne* testées à ce jour (7 espèces et plus de 30 isolats), résistance qui n'est affectée ni par une température élevée, ni par une pression forte d'inoculum. Cette résistance *Ma* à large spectre non contournée à ce jour est une ressource essentielle et la connaissance des bases moléculaires qui sous-tendent ses propriétés remarquables est un enjeu important.

Ce travail de thèse se propose donc de caractériser l'originalité structurale et/ou fonctionnelle du gène *Ma* et de contribuer à identifier les déterminants mis en jeu dans le spectre de résistance aux nématodes à galles chez les plantes du genre *Prunus*. Les objectifs plus spécifiques de cette étude sont :

- La caractérisation fine d'un nouveau domaine protéique qui détermine l'originalité structurale du gène *Ma*.
- L'étude de la distribution de ce domaine au sein des plantes terrestres (Embryophytes) et dans l'évolution de la famille des NBS-LRRs.
- La localisation et l'identification chez l'amandier du gène *RMja* porté par le même chromosome que *Ma* mais à spectre plus étroit.
- Une contribution à l'identification des déterminants de la résistance portée par *Ma* ainsi que des partenaires moléculaires potentiellement impliqués dans la réponse de la plante à l'infection.

CHAPITRE 1

CONTEXTE GENERAL

Chapitre 1 : CONTEXTE GENERAL

1.1 La réponse immune des plantes : un système à double détente

Les plantes doivent faire face à un éventail extrêmement large de menaces biotiques. La diversité des agents pathogènes au sens large ainsi que la pression continue qu'ils exercent sont autant de facteurs impactant les espèces végétales, notamment pérennes. Malgré tout, le succès de l'infection d'une plante par un pathogène (terme qui par commodité désignera pour nous aussi bien un agent pathogène qu'un ravageur), appelé aussi réaction compatible, reste un événement rare. En complément des barrières physiques ou chimiques mises en place par les plantes, regroupées sous le terme de résistance non-hôte (Heath, 2000a), les plantes possèdent un système de défense inductible permettant de détecter le « non-soi ». L'absence de cellules immunitaires spécialisées et d'un système adaptatif comme chez les animaux oblige les plantes à doter chaque cellule d'un système de reconnaissance et de défense (immunité innée) (Meunier & Broz, 2017). Celui-ci peut être décrit en deux processus interconnectés (Jones & Dangl, 2006). Un premier niveau de défense, nommé 'Pathogen-associated molecular pattern (PAMP) Triggered Immunity' ou PTI, implique le déclenchement de la réponse immune de la plante suite à la détection de molécules génériques associées aux pathogènes ou microbes tels que des flagellines, peptidoglycanes, lipopolysaccharides ou chitines (Zhang & Zhou, 2010). Cette reconnaissance s'opère via des récepteurs cellulaires majoritairement transmembranaires, comme les 'Receptor Lectin Kinases' (RLKs) ou les 'Receptor-Like Proteins' (RLPs) qui déclencheront une cascade de signalisation intracellulaire, menée par des 'Mitogen-Associated Protein Kinases' (MAPKs) et aboutissant à la réponse immune de la plante (Ali *et al.*, 2007) (**Figure 1**). En réponse à ce système, les pathogènes ont développé des molécules, ou effecteurs, capables de contourner, voire de désamorcer, ce premier niveau de défense cellulaire et de favoriser le développement du pathogène au sein de son hôte (Jones & Dangl, 2006).

Parallèlement, les plantes possèdent un second processus de défense codé par un arsenal de gènes de résistance (gènes *R*) qui détectent, la présence de pathogènes spécifiques, et induisent la réponse de la plante. Cette réponse intracellulaire nommée 'Effector Triggered Immunity' ou ETI, impliquera la détection directe ou indirecte d'effecteurs également spécifiques (Chisholm *et al.*, 2006). L'ETI pourra fortement stimuler les voies associées à la PTI (synthèse de dérivés oxygénés (ROS) et d'acide salicylique (SA), activation de cascades MAPK et de kinases dépendantes du Ca²⁺) mais aussi utiliser des mécanismes compensatoires permettant une meilleure résilience du système (Cui *et al.*, 2017). Le déclenchement de cette réaction s'accompagne le plus souvent d'une réponse dite d'hypersensibilité (Hypersensitive Response ; HR) associée à une mort cellulaire programmée (Programmed Cell Death ;

PCD) localisée (Heath, 2000b). Les récepteurs intracellulaires impliqués dans l'ETI appartiennent à la grande famille de gènes NBS-LRR ('Nucleotide-Binding Site – Leucline Rich Repat'), nommés d'après les deux principaux domaines protéiques qui les composent. On trouvera également les termes de NB-LRRs ou de NLRs (pour NOD-Like Receptors) provenant de leur équivalent dans le règne animal. Nous utiliserons le terme de NBS-LRRs afin de différencier ces récepteurs chez les plantes des récepteurs NLR chez les animaux.

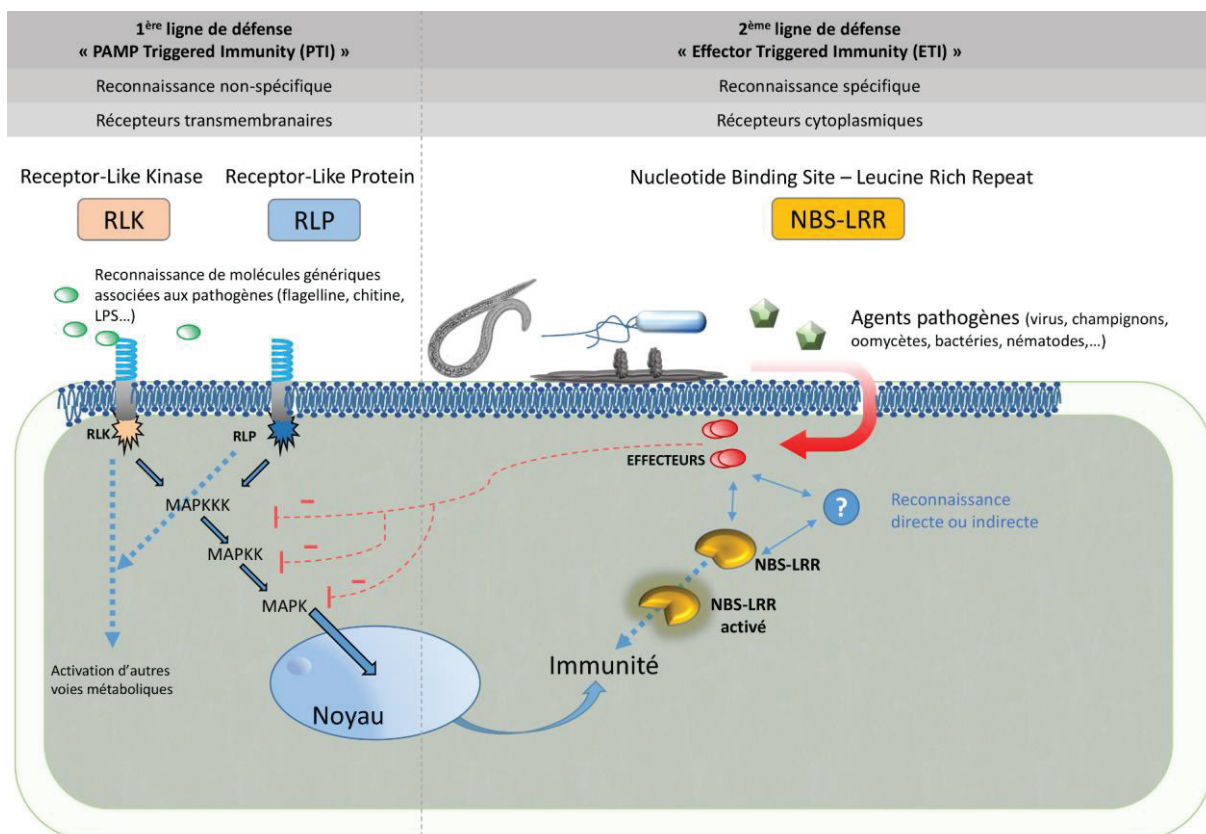


Figure 1 - Représentation schématique des deux voies principales de défense, PTI et ETI, induites chez la plante pour faire face aux agents pathogènes.
Adaptée de (Chisholm et al., 2006)

1.2 Les NBS-LRRs

1.2.1 La structure des NBS-LRRs

Le mécanisme général de l'ETI implique une reconnaissance exclusive directe ou indirecte du non-soi, une activation contrôlée et ponctuelle, et le déclenchement d'une voie de signalisation aboutissant à la réponse immune de la plante. En tant qu'acteurs majeurs de cette voie, les NBS-LRRs peuvent remplir toutes ces fonctions. Pour ce faire, ils s'appuient sur une structure très conservée et modulaire composée de divers domaines protéiques qui ont par ailleurs permis leur classification en trois sous-familles formant des clades indépendants (Shao *et al.*, 2014). S'ajoutant au domaine LRR en position C-terminale et au domaine NBS en position centrale, c'est principalement la nature du domaine situé en position N-terminale qui définit ces trois sous-familles. On trouve ainsi les '**T**oll-**I**nterleukin-1 **R**eceptors' ((TIR)-NBS-LRRs ou TNLs), les '**C**oiled-**C**oils' ((CC)-NBS-LRRs ou CNLs) et les '**R**esistance to **P**owdery milde**W**8s' ((RPW8)-NBS-LRRs ou RNLs) (Meyers *et al.*, 1999; Xiao *et al.*, 2001). Pour comprendre le mode de fonctionnement de ces acteurs majeurs de la résilience végétale, il faut tout d'abord examiner les différents domaines protéiques qui les composent.

1.2.1.1 Le domaine 'Nucleotide Binding Site' (NBS)

Le domaine NBS, tout comme le domaine NOD (**N**ucleotide-binding **O**ligomerization **D**omain) aussi appelé NACHT, chez les métazoaires, appartient à la famille des NTPases de type STAND ('**S**ignal **T**ransduction **A**TPases with **N**umerous **D**omains') (Leipe *et al.*, 2004). Le NBS, aussi appelé NB-ARC pour **N**ucleotide **B**inding-human **A**PAP-1 (**A**poptotic **P**rotease-**A**ctivating **F**actor **1**), plant **R** proteins, and nematode **CED-4** (**C**ell **D**Eath protein 4), a pour fonction principale de fixer et hydrolyser l'ADP/ATP (Tameling *et al.*, 2002). Cette fonction induit réversiblement une modification de conformation des NBS-LRRs et permet de faire passer la protéine d'un stade actif à inactif et vice-versa afin d'assurer la transduction du signal. Cet équilibre entre les formes actives et inactives pourrait être un facteur déterminant du déclenchement de la réaction de défense. Ainsi tout élément intervenant pour favoriser le déplacement de l'équilibre vers la forme active de la protéine tendrait à promouvoir le déclenchement de la réponse immune (Zhang *et al.*, 2017b). En plus d'une fonction de régulation, ce domaine possède également des propriétés d'oligomérisation.

Ce domaine, placé en position centrale et composé d'environ 350 acides aminés, est constitué de 8 motifs organisés en 3 sous-domaines (Meyers *et al.*, 1999, 2003; van Ooijen *et al.*, 2008). On trouve de la région N-terminale vers la région C-terminale (**Figure 2**):

- † Le sous-domaine NB comprenant les motifs P-loop (aussi appelé kinase ou Walker A) fixant l'ATP/ADP (Takken *et al.*, 2006), RNBS-A, Walker B (kinase 2) requis pour l'hydrolyse de l'ATP (revue dans (Hanson & Whiteheart, 2005), et le motif RNBS-B (kinase 3) interagissant avec la base purique.
- † Le sous-domaine ARC1 se compose du motif RNBS-C et du motif hydrophobe GLPL.
- † Le sous-domaine ARC2 contenant le motif RNBS-D et le motif MHD. La mutation des résidus H ou D dans ce dernier motif a la particularité de déclencher l'auto-activation chez de nombreux NBS-LRRs (Bendahmane *et al.*, 2002), probablement en favorisant la fixation de l'ATP (forme active) au détriment de l'ADP (forme inactive) (Williams *et al.*, 2011).

Les motifs soulignés ci-dessus (P-loop, Walker B, RNBS-B et GLPL) sont conservés chez tous les NBS-LRRs alors que les autres varient en fonction de la sous-famille considérée (Meyers *et al.*, 2003). Certains de ces motifs sont impliqués dans des interactions intra-domaines, d'autres dans des interactions avec le domaine LRR afin de garantir la stabilité et une activation raisonnée de la protéine (Slootweg *et al.*, 2013). Jusqu'à très récemment aucune structure 3D de NBS de plante n'avait été résolue et leur prédiction pouvait être approximée *in silico* via les modèles APAF-1 ou CED-4 (van der Biezen & Jones, 1998) comme illustré avec le domaine NBS de la protéine Ma (

Figure 2). La structure tridimensionnelle de ce domaine majeur commence à être mieux caractérisée chez les plantes (Steele *et al.*, 2019; Wang *et al.*, 2019a).

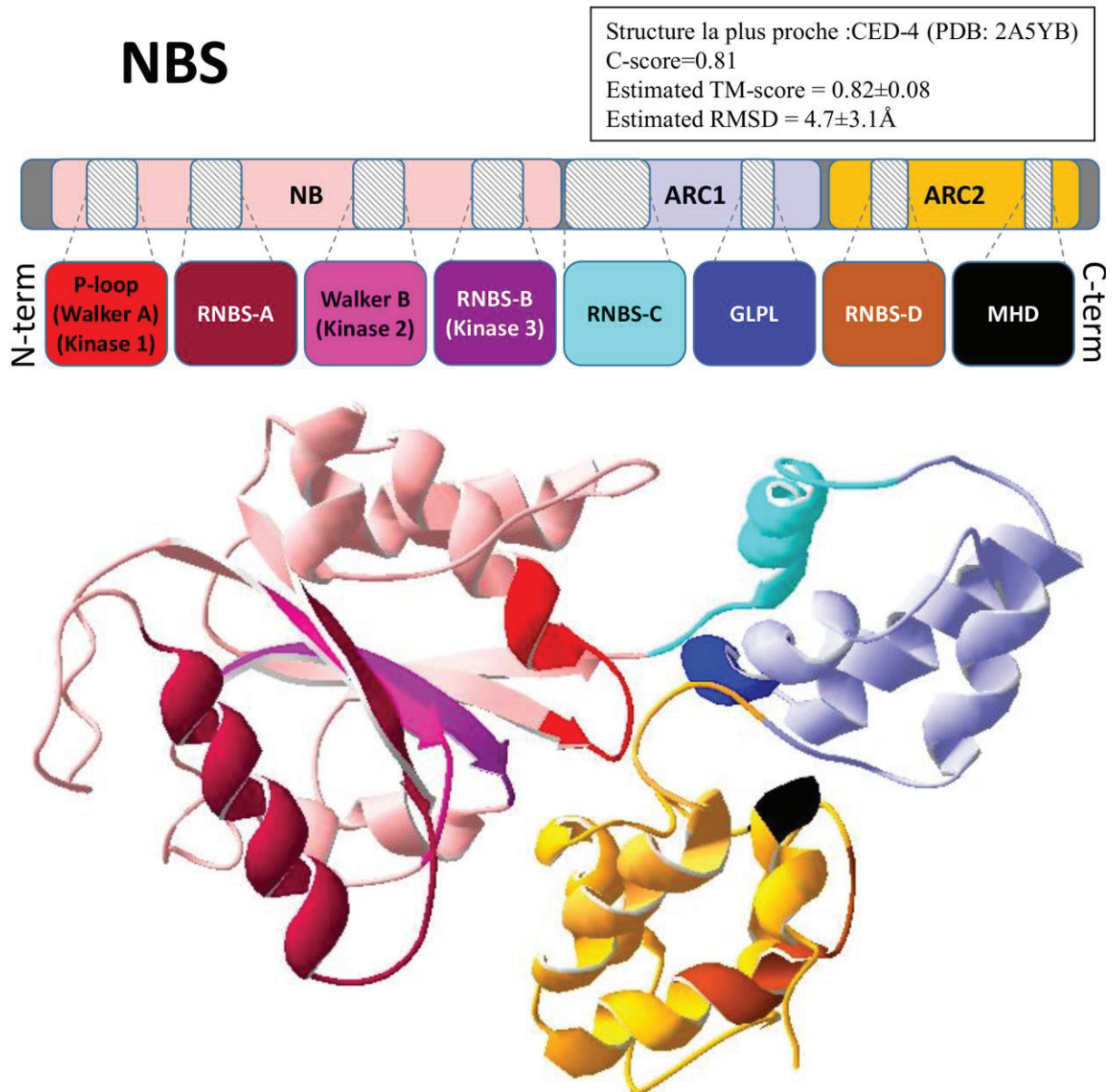


Figure 2 - Structure tridimensionnelle, prédite par le serveur i-tasser, du domaine NBS de la protéine Ma. Les différents sous-domaines et motifs d'un domaine NBS (en haut). Prédiction *in silico* du domaine NBS de Ma basé sur la structure de CED-4 (PDB: 2A5YB) avec les positions des différents motifs et de la molécule d'ATP/ADP (en bas). Les scores de prédiction sont indiqués dans l'encadré. Le C-score est le score de confiance pour estimer la qualité de la prédiction du modèle. Il se situe entre -5 et +2. Plus il est élevé, meilleure est la prédiction. Les TM-score et RMSD sont des standards pour mesurer la similarité structurelle entre deux structures.

1.2.1.2 Le domaine 'Leucine-Rich Repeat' (LRR)

Ce domaine, présent chez d'autres familles de protéines comme les récepteur kinases, est localisé en position C-terminale chez les NBS-LRRs. Il est composé de répétitions de motifs polymorphes d'environ 25 résidus possédant des signatures du type LxxLxxLxxLxL (Kobe & Deisenhofer, 1994; Jones & Jones, 1997). Comme beaucoup de domaines versatiles impliquant des motifs répétés, le domaine LRR a depuis longtemps été associé à des fonctions d'interaction avec des ligands comme c'est le cas chez de nombreux récepteurs membranaires (Deisenhofer & Kobe, 1995). Plusieurs exemples ont montré l'implication du domaine LRR de protéines NBS-LRR (L6 et RPP1) dans des interactions directes avec des effecteurs (Dodds *et al.*, 2006; Krasileva *et al.*, 2010). Les nombreuses études structurales relatives à ce domaine, notamment chez les récepteur kinases contenant un domaine LRR extracellulaire (Hohmann *et al.*, 2017), ont révélé une structure tridimensionnelle caractéristique dite en forme de « fer à cheval » qui s'appuie sur une alternance de feuillets β et d'hélices α comme illustré avec la structure prédite du domaine LRR de la protéine Ma (**Figure 3**). Cette structure, théorisée sous la forme E4C5 (alternance d'environ 4 résidus par brins β et 5 résidus en structure coil) (Mondragón-Palomino *et al.*, 2002), contient des résidus impliqués dans des interactions intramoléculaires avec le domaine NBS maintenant la protéine en position inactive et une large partie exposée aux solvants, dans laquelle les résidus jouent un rôle prépondérant dans les interactions avec les ligands exogènes (Ade *et al.*, 2007; Van Ooijen *et al.*, 2008). Certains exemples montrent que la modification d'un seul résidu du domaine LRR peut compromettre la reconnaissance d'un effecteur ou améliorer le spectre de résistance conféré par un NBS-LRR (Bakker *et al.*, 2014; Segretin *et al.*, 2014). Par ailleurs la plasticité de ce domaine, destinée à accroître ses capacités de reconnaissance, est en ligne avec la sélection positive (Darwinienne) à laquelle il est soumis. En effet, contrairement au domaine NBS soumis à une forte sélection négative ($Ka/Ks \ll 1$) en vue de conserver des motifs cruciaux, le domaine LRR présente fréquemment un ratio $Ka/Ks > 1$ (Mondragón-Palomino *et al.*, 2002).

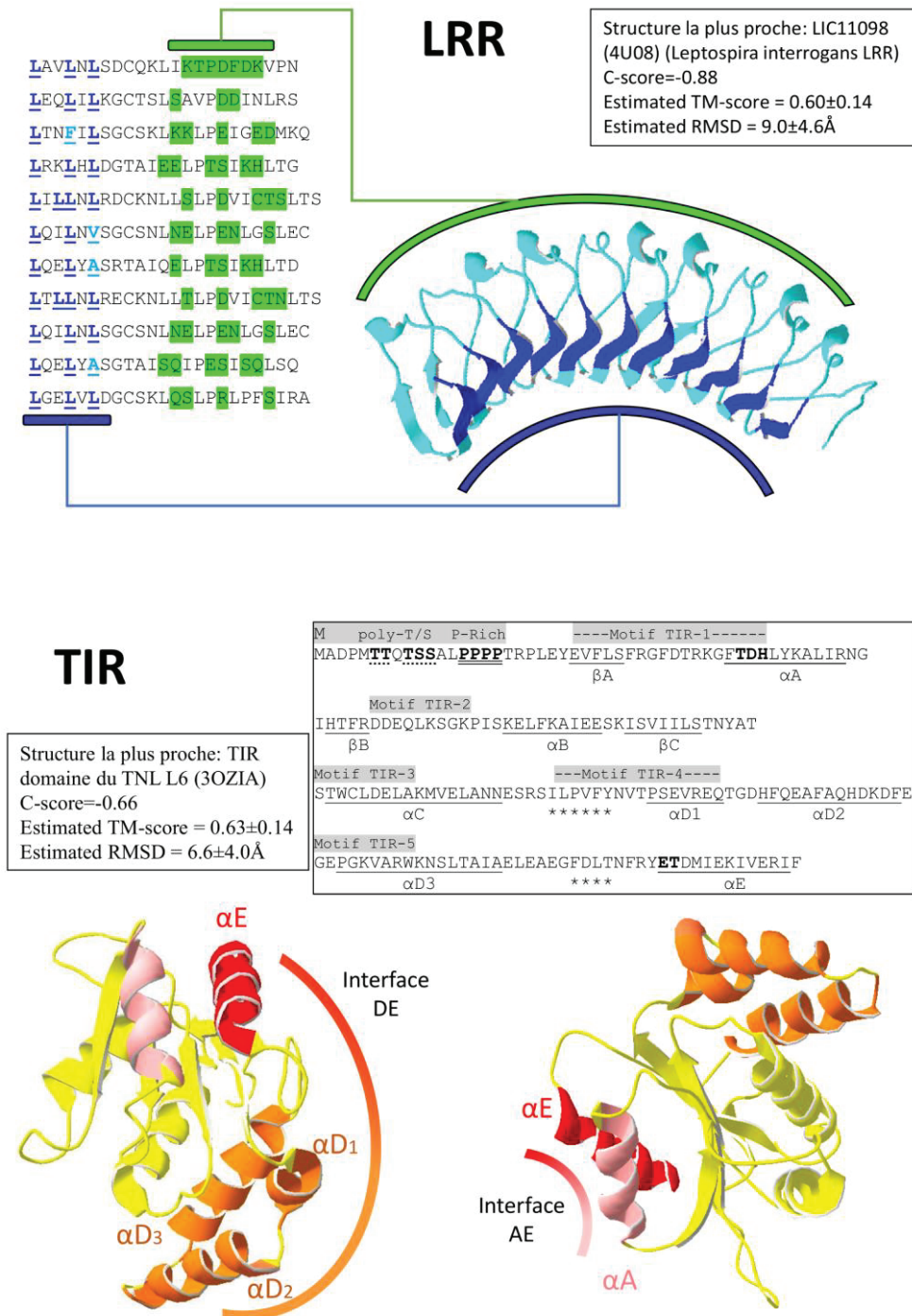


Figure 3 - Structures tridimensionnelles, prédites par le serveur i-tasser, des domaines LRR et TIR de la protéine Ma.

Les différents motifs et résidus impliqués dans les feuillets β et hélices α du domaine LRR sont identifiés en bleu foncé et vert, respectivement (en haut). Prédiction *in silico* du domaine LRR de Ma basée sur la structure du domaine LRR de la protéine LIC11098 (PDB :4U08). Prédiction *in silico* du domaine TIR de Ma basée sur la structure du domaine TIR de la protéine L6 (PDB :3OZIA) (en bas). La figure indique la position des interfaces DE et AE, identifiées comme importantes pour les processus de dimérisation (résidus des hélices αA et αE en gras et soulignés) et d'activation de l'immunité (Bernoux *et al.*, 2011; Williams *et al.*, 2014). Dans l'encadré, séquence du domaine TIR de Ma avec les résidus impliqués dans les différents motifs (en gris) et structures secondaires (soulignés). (*) indique l'emplacement de feuillets β présents chez L6 mais non prédits pour le domaine TIR de Ma. Les scores de prédiction sont indiqués dans les encadrés. Le C-score est le score de confiance pour estimer la qualité de la prédiction du modèle. Il se situe entre -5 et +2. Plus il est élevé, meilleure est la prédiction. Les TM-scores et RMSDs sont des standards pour mesurer la similarité structurelle entre deux structures.

1.2.1.3 Le domaine 'Toll/Interleukin-1 Receptor' (TIR)

Le domaine TIR fut identifié chez le gène *N* responsable de la résistance au *Tobacco mosaic virus* (TMV) chez *Nicotiana tabacum* par homologie avec les domaines cytoplasmiques des récepteurs Toll et Interleukine-1 (IL-1R) chez, respectivement, la drosophile et l'homme (Heguy *et al.*, 1992; Whitham *et al.*, 1994). Tout comme les IL-1R, le domaine TIR est susceptible de déclencher une cascade moléculaire induisant une réaction de défense. Ainsi, le domaine TIR seul des TNLs L6, RPS4 ou SNC1 est capable d'induire une mort cellulaire après une dimérisation essentielle au processus (Bernoux *et al.*, 2011; Williams *et al.*, 2014; Zhang *et al.*, 2017a). Tout comme pour les domaines NBS et LRR, une étude conduite sur RPP1 a montré l'importance des interactions inter-domaines entre le domaine TIR et le domaine NBS. En effet celles-ci inhiberaient la dimérisation des NBS-LRRs et l'auto-activation synonyme de mort cellulaire spontanée (Schreiber *et al.*, 2016). Dans la majorité des cas, ce domaine est composé d'une suite de résidus S ou T (en gras et soulignés en pointillés dans l'encadré (**Figure 3**)) puis de résidus proline (en gras et soulignés deux fois dans l'encadré (**Figure 3**)) dans sa partie la plus N-terminale, suivie de 3 à 5 motifs conservés (Meyers *et al.*, 1999, 2003). Certains TNLs présenteront aussi, ou en lieu et place des résidus S/T, un peptide signal ou un domaine transmembranaire reflétant leur possible localisation dans la cellule. D'un point de vue structural, ce domaine est constitué de diverses interfaces secondaires formées par des hélices α (interfaces $\alpha A-\alpha E$ et $\alpha D-\alpha E$) qui sont impliqués dans les fonctions de dimérisation (**Figure 3**). L'étude des structures 3D des domaines TIR des NBS-LRRs L6, RPS4/RRS1, RPP1, SNC1 et RPV1 a permis d'identifier les interfaces et certains acides aminés cruciaux pour les processus de dimérisation et/ou d'auto-activation (Bernoux *et al.*, 2011; Williams *et al.*, 2014, 2016; Zhang *et al.*, 2017a). Certains domaines TIR de TNLs atypiques comme RRS1, dont le TIR n'est pas capable d'activer la voie de signalisation par lui-même, possèdent des motifs moins conservés et une absence d'hélices $\alpha D1$ et $\alpha D2$ (Williams *et al.*, 2014).

1.2.1.4 Le domaine Coiled-Coil (CC)

Ce domaine structurel possède une séquence très peu conservée et de taille variable (150-250 acides aminés) (Meyers *et al.* 2003 ; Bai *et al.* 2002). Même si un motif EDVID a pu être identifié, les diverses signatures composant ce domaine semblent dépendantes de la taxonomie et/ou du mode de fonctionnement. Le CNL modèle Mi 1.2, responsable de la résistance aux nématodes à galles chez la tomate, possède deux sous-domaines dans sa partie N-terminale dont le premier régulerait négativement le domaine LRR inactivant ainsi les voies de transduction (Hwang & Williamson, 2003). Comme le domaine TIR, certains exemples ont montré que le domaine CC peut être capable de se dimériser. Cependant le déclenchement d'une HR par la seule action d'un domaine CC semble être moins généralisé comme illustré chez les modèles Rx ou RPS5 qui en sont incapables. Une étude récente menée sur l'ensemble des domaines CC des CNLs chez *Arabidopsis* a mis en lumière trois groupes majeurs de séquences avec pour point commun quatre hélices α (H1a et b et H2a et b) impliquées dans les processus

de dimérisation. Cette étude a également mis en lumière les processus préférentiels d'hétérodimérisation dont font l'objet ces domaines, dans un réseau complexe, pour aboutir à différents phénotypes chez la plante (Wróblewski *et al.*, 2018). Par ailleurs, le domaine CC pourrait également interagir directement avec des kinases cibles d'effecteurs comme les domaines CC des CNLs Prf et ZAR1 avec respectivement, les kinases Pto et ZED1 (Mucyn *et al.*, 2006; Lewis *et al.*, 2013). Les CC sont ainsi directement impliqués dans un rôle de surveillance d'éléments de la PTI, ciblés par des effecteurs.

1.2.1.5 Le domaine RPW8

Le domaine RPW8 ou 'resistance to powdery mildew 8' a été identifié chez les deux protéines d'*Arabidopsis* RPW8.1 et RPW8.2, conférant une résistance à large spectre à l'oïdium. De structure inconnue ce domaine semblerait posséder un module d'ancrage à la membrane plasmique (Xiao *et al.*, 2001). Le domaine RPW8 est à lui seul capable de déclencher la réponse de la plante. L'identification de ces deux protéines contenant uniquement ce domaine a permis par la suite d'identifier certains CNLs comme présentant en position N-terminale, une homologie avec le domaine RPW8. Jusqu'à récemment les études phylogénétiques consacrées aux NBS-LRRs et basées sur le domaine NBS, pointaient un clade de TNLs et plusieurs clades de CNLs. Il s'avère qu'un de ces clades de CNLs regroupait les RNLs, formant ainsi un groupe distinct (Shao *et al.*, 2014). Encore aujourd'hui, certaines études continuent d'attribuer un lien entre domaines CC et RPW8 en nommant ces derniers CC_R et ce malgré des structures, des séquences et des positions phylogénétiques distinctes (Wróblewski *et al.*, 2018). Deux gènes modèles aux profils différents servent à illustrer la diversité observée chez les RNLs. Le gène ADR1 ('activated disease resistance 1') impliqué dans la tolérance à la sécheresse qui se retrouve chez toutes les plantes en peu d'exemplaires (Chini *et al.*, 2004) et le gène NRG1 ('N requirement gene 1') qui a, quant à lui, été associé à la résistance au *Tobacco mosaic virus* (TMV) (Peart *et al.*, 2005). Les séquences homologues à NRG1, en plus grand nombre, ne se retrouvent que dans certains groupes taxonomiques (Collier *et al.*, 2011).

1.2.1.6 Les domaines additionnels et les structures atypiques

Chez certains groupes taxonomiques, on peut trouver en position N-terminale des NBS-LRRs, des domaines remplissant des fonctions d'activation de cascade moléculaire ou d'interaction avec l'ADN du type 'BED zinc finger DNA-binding', 'hydrolase', 'kinase' ou 'Solanaceae'. On peut également identifier des domaines atypiques inclus à différentes positions dans les structures de NBS-LRRs et qui sont regroupés sous le terme de domaines intégrés (IDs). On pourra citer les domaines liés aux facteurs de transcription comme le domaine WRKY chez le TNL RRS1 d'*Arabidopsis* (Deslandes *et al.*, 2002), liés aux oxydo-réductions comme le domaine HMA ('heavy metal-associated') chez les CNLs RGA5 et Pik-1 du riz (Césari *et al.*, 2014) ou encore le domaine kinase particulièrement bien représenté. Selon

les estimations, entre 3,5% (Kroj *et al.*, 2016) et 10% des NBS-LRRs chez les plantes à fleurs possèderaient un ID dont la nature peut être très variable puisque 265 domaines différents ont pu être détectés (Sarris *et al.*, 2016). Certains modèles particulièrement bien étudiés attribuent le plus souvent à ces IDs, une fonction de leurre pour tromper les effecteurs, qui, s'associant avec eux plutôt qu'avec leurs cibles privilégiées, provoqueront leur détection et donc une réponse de la plante. Il s'agit donc ici d'un accroissement des capacités de reconnaissance d'un NBS-LRR standard par acquisition d'une nouvelle diversité, quitte à perdre certaines fonctionnalités. En effet le domaine TIR de RRS1 ne peut induire par lui-même le déclenchement de la réaction de HR. Il doit s'appuyer sur l'intervention d'un autre NBS-LRR doté d'une structure plus conventionnelle, RPS4. Ce fonctionnement dit « en paire » permettra le déclenchement de la voie de signalisation (Le Roux *et al.*, 2015).

En plus de ces structures particulières, certains exemples de NBS-LRRs tronqués (ne possédant qu'un ou deux domaines canoniques) ont été identifiés comme importants dans divers processus de résistance comme RPW8.1 et RPW8.2 (formés du seul domaine RPW8), RBA1 (constitué du seul domaine TIR) et TN2 (structure en TIR-NBS). Ceci souligne l'importance d'étudier également ces candidats atypiques dit « tronqués ».

Enfin les TNLs ont la particularité de disposer de résidus s'étendant bien au-delà du LRR et qui ne semblent pas être présents chez les CNLs ou les RNLs. Bien que des signatures caractéristiques aient été détectées chez *Arabidopsis* (Meyers *et al.*, 2003), ces séquences comprenant environ 200 acides aminés sont restées identifiées comme une extension C-terminale située après le domaine LRR et éventuellement nommée « domaine C-terminal ». L'étude du gène *Ma*, responsable de la résistance aux *Meloidogyne* chez les *Prunus* a mis en lumière ces modules en nommant ces séquences, domaine Post-LRR (PL) (Claverie *et al.*, 2011).

1.2.2 Les différents types de perception

La composante qui nécessite le plus de diversité est la reconnaissance du non-soi. De nombreux processus sont mis en jeu par les NBS-LRRs pour percevoir les pathogènes (revue dans (Cesari, 2017)). Le mode de perception le plus simple est la reconnaissance directe. Ce type d'interaction entre un effecteur et un NBS-LRR représente bien le modèle « gene for gene » théorisé par Flor (Flor, 1955) et peut être illustré par les modèles *L* (TNLs chez le lin) / AvrL (effecteur du champignon pathogène *Melampsora lini*) ou *RPPI* (TNL chez *Arabidopsis*) / ATR1 (effecteur de l'oomycète *Hyaloperonospora arabidopsis*) (Dodds *et al.*, 2006; Krasileva *et al.*, 2010). Ce modèle favorise une coévolution hôte-pathogène et nécessite une diversité extrêmement importante pour répondre à l'immense diversité des effecteurs. Par conséquent, la reconnaissance directe n'est sans doute pas le mode de fonctionnement le plus répandu et reste probablement un mode opportuniste.

A l'inverse, le mode de reconnaissance indirecte permet à un nombre restreint de gènes *R* de détecter une grande variété d'effecteurs. Pour ce faire, des modifications apportées par différents effecteurs à une (des) protéine(s) dite(s) « de garde » permettent à une poignée de NBS-LRRs de percevoir la présence du pathogène (McHale *et al.*, 2006; van der Hoorn & Kamoun, 2008). La protéine de garde RIN4 ('RPM1-Interacting protein 4') en est la parfaite illustration puisque différentes modifications apportées par plusieurs effecteurs sont surveillées par les CNLs RPM1 et RPS2 (Lee *et al.*, 2015). En surveillant certains complexes clefs, les NBS-LRRs peuvent ainsi contrôler un ensemble d'attaques potentielles.

Par ailleurs, et afin de préserver les cibles les plus importantes pour la cellule, des cibles alternatives trompant les effecteurs existent (van der Hoorn & Kamoun, 2008). Ces leurres ciblés par les effecteurs n'induisent pas d'actions particulières mais sont reconnus par des NBS-LRRs qui activeront la voie ETI. On trouve ainsi des protéines-leurres non fonctionnelles telles que les pseudo-kinases RKS1 et ZED1 qui sont sous la surveillance du CNL ZAR1 (**Figure 4**). Une fois modifiées par les effecteurs bactériens AvrAC (*Xanthomonas* sp.) et HopZ1a (*Pseudomonas syringae*), ZAR1 détectera cette modification et déclenchera la réponse de la plante (Wang *et al.*, 2015). Ces exemples illustrent l'intrication qui existe entre acteurs de la PTI et de l'ETI afin de produire une réponse aux menaces.

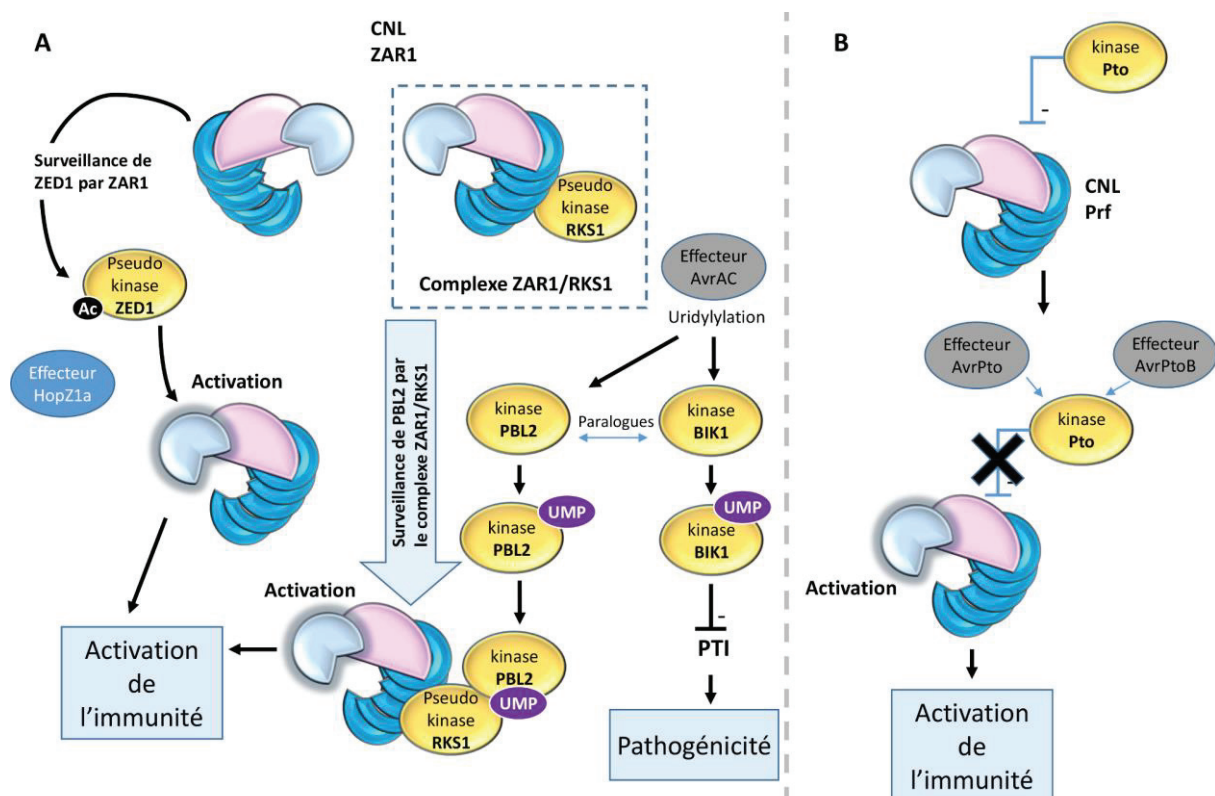


Figure 4 - Deux systèmes de surveillance entre NBS-LRRs et protéines kinases illustrant les stratégies de leurre et de garde.

(A) Réseau de surveillance entre le CNL ZAR1 et les kinases ZED1, PBL2 et RKS1. (B) Interaction entre le CNL Prf et la kinase Pto, codés par des gènes co-localisés dans le génome de la tomate. Les domaines CC, NBS et LRR sont représentés respectivement en gris, rose et bleu. Basé sur (Mucyn *et al.*, 2006; Wang *et al.*, 2015).

La stratégie de leurre est aussi une capacité acquise intrinsèque de certains NBS-LRRs ayant incorporé un domaine surnuméraire intégré (ID), comme évoqué en 1.2.1.6, mimant une cible potentielle d'effecteurs. Dans la plupart des cas, ces domaines sont non fonctionnels mais conservent les résidus essentiels permettant à l'effecteur de se fixer ou d'induire une modification qui sera postérieurement détectée. L'effecteur PopP2 de *Ralstonia solanacearum* a la capacité d'acétyler les facteurs de transcription de type WRKY entraînant leur dissociation avec l'ADN et en conséquence de réprimer l'immunité (**Figure 5**). Cependant, PopP2 va également acétyler le domaine WRKY intégré au TNL RRS1. RRS1 est alors associé à l'ADN via ce domaine mais également organisé en hétérodimère avec l'autre TNL RPS4. L'acétylation du domaine WRKY, qui joue ici le rôle de leurre, libérera le complexe RRS1/RPS4, ce qui déclenchera l'immunité dépendante de RPS4, RRS1 jouant un rôle de régulateur négatif de RPS4 (Le Roux *et al.*, 2015). Cette stratégie fonctionnant en paire de NBS-LRRs, couple un NBS-LRR modifié, pourvu d'un domaine cible d'effecteur, dit « NBS-LRR sensor », à un deuxième NBS-LRR de structure conventionnelle que l'on peut qualifier de « NBS-LRR helper ». Ce mode de fonctionnement, incluant la régulation négative entre les deux NBS-LRRs, pourrait s'expliquer par le déséquilibre structurel provoqué par l'acquisition de ce nouveau domaine chez le « NBS-LRR sensor » ne permettant plus à la protéine de garantir une stabilité entre états d'inactivation et d'activation indispensable à la viabilité de la cellule. La fonction d'activation a donc été transférée au deuxième élément de la paire. Il est intéressant de noter que cette relation fonctionnelle s'accompagne d'une co-localisation dans le génome où chaque couple de 'TNL sensor / TNL helper', en effectif d'au moins 8 chez *Arabidopsis*, se positionne en anti-sens et à courte distance, et que chaque TNL d'une paire est très différent de l'autre (Narusaka *et al.*, 2009).

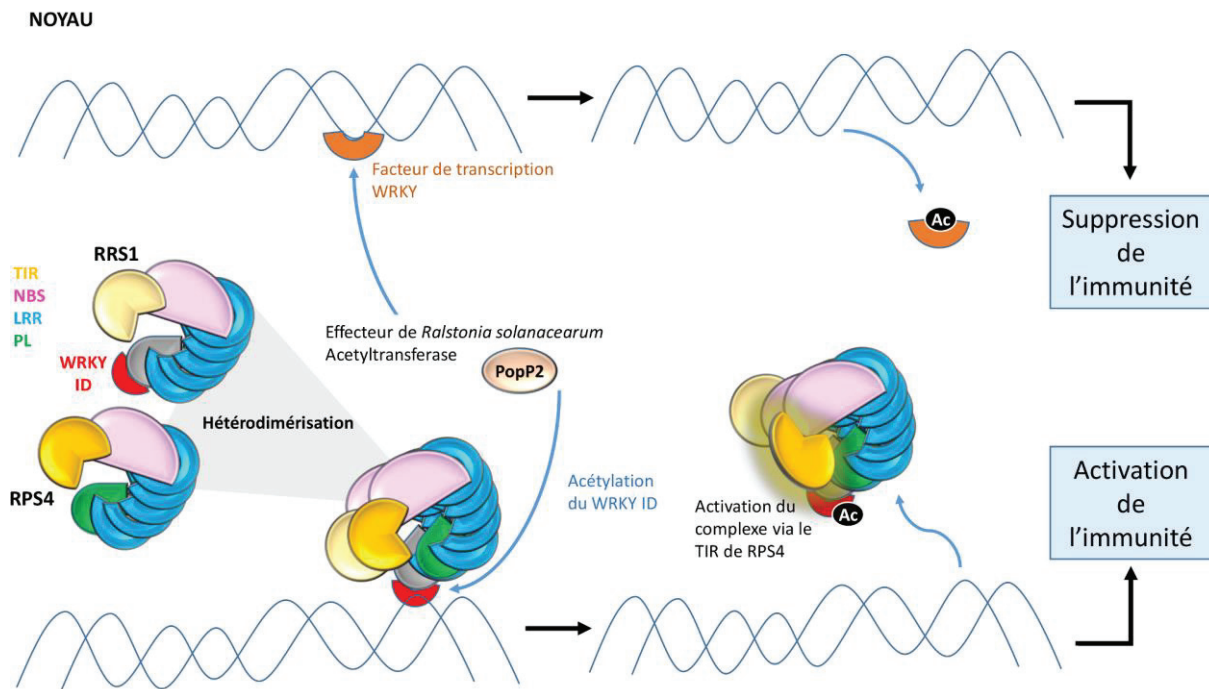


Figure 5 - Mécanisme de défense engagé par la paire de TNLs RPS4/RRS1 incluant une stratégie de leurre intégré.

Les facteurs de transcription WRKY se fixent sur l'ADN. Adapté de (Le Roux *et al.*, 2015).

Par le jeu de la sélection, la liaison génétique est favorisée entre des facteurs impliqués dans une réponse commune à un pathogène ou s'inscrivant dans un processus lié. Ce phénomène de co-localisation est également identifié entre le CNL Prf qui se retrouve physiquement lié avec un cluster de récepteur kinases contenant la kinase Pto dans le génome de la tomate. Prf surveille Pto qui se trouve être ciblée par les effecteurs de *Pseudomonas syringae*. En retour, Prf est susceptible de déclencher la réponse immune de la plante (**Figure 4**) (Mucyn *et al.*, 2006). Le champ des modes de perception et d'activation semble être très large et reste à explorer à l'image de ce qui a été décrit chez d'autres acteurs de la défense de plantes revue par (Kourelis & van der Hoorn, 2018).

1.2.3 Les réseaux de NBS-LRRs et le 'résistosome' comme nouveaux paradigmes

Récemment, l'idée d'une coopération plus large entre NBS-LRRs a été mise en évidence chez les Solanaceae (Rathjen & Dodds, 2017). En effet, un réseau composé de différents CNLs connus et agissant de concert pour induire une réponse chez cette famille botanique, a été identifié. Dans ce système des CNLs, conférant des résistances à divers pathogènes tels que Rpi-blb2 (oomycète *Phytophthora infestans*), Mi-1.2 (nématodes à galles, certains pucerons et aleurodes), Sw5b (virus TSWV), R8 (*P. infestans*), R1 (*P. infestans*), Prf (*Pseudomonas syringae*), ou Bs2 (*Xanthomonas campestris*), perçoivent la présence des pathogènes et activent d'autres CNLs appelés NRC-1 à NRC-4

(‘NBS-LRR required for cell death’) qui déclencheront le programme de mort cellulaire (Wu *et al.*, 2017). Ces NRCs agissent avec une redondance fonctionnelle et sont restreints au clade des Astéridés contenant notamment les Solanaceae. Cette étude montre que certains groupes taxonomiques ont pu développer des voies de signalisation particulières en spécialisant certains acteurs de l’ETI.

D’autre part, Peart et ses collaborateurs ont montré que le TNL N coopère avec le RNL NRG-1 (‘N requirement gene 1’) en tabac pour conférer une résistance au ‘*Tobacco mosaic virus*’ (Peart *et al.*, 2005). Plus récemment, il a été montré que NRG-1 était aussi essentiel aux résistances médiées par les TNLs Roq1 (résistance à *Xanthomonas* spp.) et RPP1 (résistance à *Pseudomonas*) mais pas pour les résistances induites par les CNLs Bs2 et RPS2. Par ailleurs NRG-1 agirait en aval de ces TNLs percevant la menace et même en aval de la pierre angulaire activant la voie de signalisation associée aux TNLs, EDS1 (‘enhanced disease susceptibility 1’) (Qi *et al.*, 2018). Curieusement le RNL ADR1 agirait comme régulateur positif sur les résistances médiées à la fois par des TNLs (RPP1, RPP4) mais aussi des CNLs (RPS2 et Rx2) (Bonardi *et al.*, 2011). Là encore, ce type de fonctionnement est taxonomiquement restreint puisque à la fois les TNLs et les RNLs de type NRG1, ne se retrouvent que dans certains groupes phylétiques (Collier *et al.*, 2011). Une étude fonctionnelle récente a mis en lumière une partie importante du réseau incluant les deux RNLs ADR1 et NRG1, agissant en aval de nombreux TNLs (Roq1, RRS1/RPS4, RPP1, RPP2, RPP4, WRR4A, N, CHS3/CSA1, SNC1 et CHS1/SOC3) et de quelques CNLs (RPM1, RPS2, Rx2), pour induire une réponse de la plante (Figure 6) (Castel *et al.*, 2019).

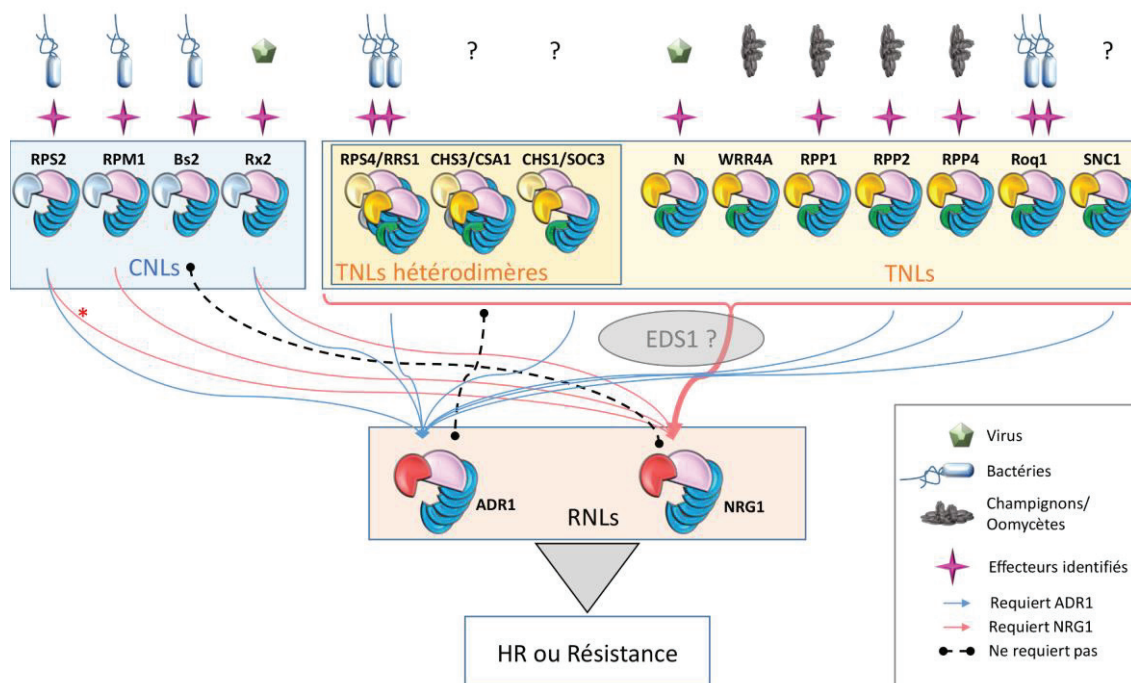


Figure 6 - Réseau fonctionnel entre TNLs et RNLs dans le déclenchement de la réponse immunitaire chez les plantes.

Les RNLs NRG1 et ADR1 perçoivent, en aval, le signal déclenché par les CNLs, les TNLs et potentiellement EDS1. *Données contradictoires entre les deux études. Adapté de (Qi *et al.*, 2018; Castel *et al.*, 2019).

Comme évoqué dans le paragraphe 1.2.1, plusieurs domaines sont impliqués dans la dimérisation des NBS-LRRs, notamment le domaine NBS, processus essentiel au déclenchement de la réponse immune de la plante. Chez les animaux, un processus plus large d'oligomérisation des protéines NACHT-LRR a été montré notamment avec NLRC4, une CARD-NACHT-LRR chez la souris. Le complexe formé, nommé 'inflammasome', regroupe entre autres 10 monomères de NLRC4. Se peut-il que de telles structures existent dans le système immunitaire des plantes ? En se basant sur le complexe ZAR1-RKS1-PBS2 (CNL-pseudo kinase-kinase leurre), les auteurs ont, dans un premier temps, réussi à révéler la structure du couple ZAR1-RKS1 montrant la seule implication du domaine LRR dans leur interaction (Wang *et al.*, 2019b). En ajoutant la kinase PBS2 préalablement uridylylée par l'effecteur AvrAC, ils ont obtenu la formation du complexe ZAR1-RKS1-PBS2 uridylylé grâce à un changement de conformation enclenché par l'expulsion de l'ADP lié au domaine NBS de ZAR1 (Wang *et al.*, 2019b). Enfin, l'ajout de dATP leur a permis de mettre en évidence par Cryo EM, un complexe pentamérique (Wang *et al.*, 2019a). Dans ce pentamère, tous les domaines de ZAR1 sont impliqués dans de multiples interactions, et les régions N-terminales des cinq domaines CC se réorganisent pour former une structure en forme d'entonnoir (Wang *et al.*, 2019a) qui pourrait créer des pores aux niveaux des membranes. Ces études majeures ont permis de lever le voile sur ce processus en confortant un peu plus les similitudes observées entre règnes animal et végétal. Il reste aujourd'hui à comprendre si les TNLs pourraient être engagés dans de telles structures sachant les différences qui existent entre leurs régions N- et C-terminales et celles des CNLs.

1.2.4 Les voies de signalisation et la localisation des NBS-LRRs

Les voies de signalisation déclenchées par les NBS-LRRs sont différentes selon les sous-familles. Deux voies de défense initiées par les trois sous-familles coexistent au sein des plantes. En effet, TNLs et RNLs activeraient en aval une voie impliquant les mêmes facteurs (Xiao *et al.*, 2005) alors que les CNLs agirait sur d'autres composants. Ainsi le composant NDR1 ('Non-race specific Disease Resistance-1') serait requis par de nombreux CNLs pour induire une réponse de la plante (Aarts *et al.*, 1998). Cette protéine, présentant une similarité avec les intégrines, serait directement liée à l'initiation des cascades MAPK de la PTI mais serait également impliquée dans la perception de la menace (Knepper *et al.*, 2011).

Les TNLs et RNLs dépendent eux aussi d'un composant crucial : EDS1 ('enhanced disease susceptibility 1'). Tout comme NDR1, EDS1 est un régulateur positif de l'immunité basale. Cette lipase interagirait à la fois avec des TNLs (RPS4, SNC1 ou RPS6), des régulateurs négatifs de l'immunité (SRFR1) (Bhattacharjee *et al.*, 2011) et des RNLs (NRG1) (Qi *et al.*, 2018) mais également avec d'autres lipases ('Phytoalexin deficient 4' - PAD4 et 'Senescence Associated Gene 101' - SAG101)

(Wagner *et al.*, 2013) (**Figure 7**). En s'associant soit à PAD4, soit à SAG101, EDS1 joue un rôle de connecteur entre les TNLs activés par la présence d'effecteurs et les voies de signalisation de la défense et de la mort cellulaire programmée. Le complexe EDS1/PAD4 promeut également l'expression des gènes permettant la synthèse de l'acide salicylique (SA) et l'accumulation de celui-ci induira l'expression de EDS1 et de PAD4 dans une boucle positive contribuant au phénomène de résistance (Cui *et al.*, 2017). Il est intéressant de noter que, contrairement à EDS1 et PAD4, SAG101 est absent des mêmes groupes taxonomiques (Monocotylédones, certains Astérides) que les RNLs du type NRG1 et que les TNLs (Wagner *et al.*, 2013). Enfin un autre complexe a été identifié comme essentiel pour les NBS-LRRs en général. Ce complexe, composé de la protéine chaperonne HSP90 ('Heat Shock Protein 90kD) et de ses deux co-chaperonnes SGT1 ('Suppressor of G2 allele of skp1') et RAR1 ('Required for Mla12 resistance'), contribuerait à la stabilisation, la maturation et l'activation des NBS-LRRs (Kadota *et al.*, 2010). SGT1 serait également impliqué dans un large éventail de processus incluant la régulation et l'éventuel adressage des NBS-LRRs au protéasome.

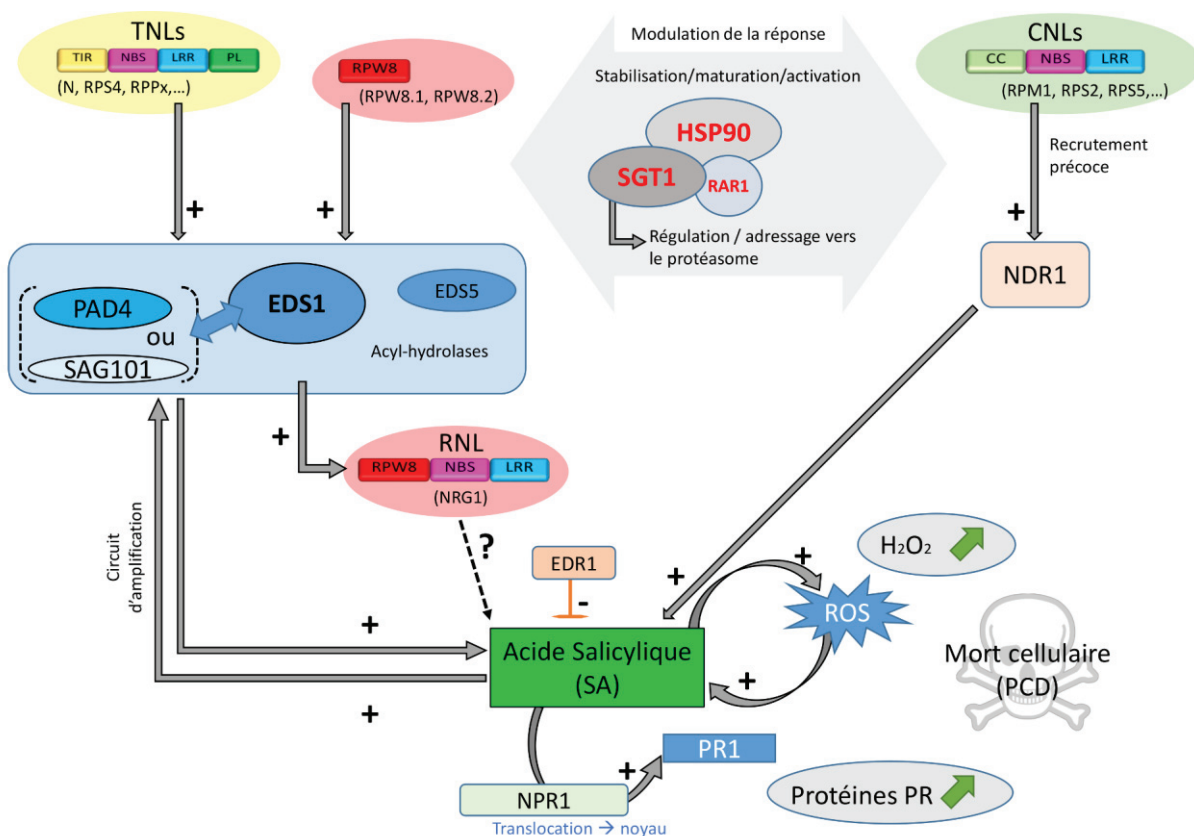


Figure 7 - Représentation schématique des voies de signalisation des NBS-LRRs.

NDR1 (Non race-specific Disease Resistance), RAR1 (Required for Mla Resistance), SGT1 (Suppressor of G2 allele of SKP1), EDS1 (Enhanced Disease Susceptibility), PAD4 (PhytoAlexin Deficient 4), SAG101 (Senescence Associated Gene 101), NPR1 (Non expressor of PR genes 1); +/- : régulation positive/négative. Adapté de (Hammond-Kosack & Parker, 2003; Caldwell & Michelmore, 2009).

Les NBS-LRRs sont des récepteurs cytoplasmiques qui peuvent émettre un signal à partir de divers compartiments. Ainsi de nombreux NBS-LRRs possèdent des signaux d'adressage vers des organelles ou nécessitent une translocation essentielle à leur action. Certains NBS-LRRs peuvent être par exemple adressés à la membrane plasmique (comme le CNL RPM1) (Gao *et al.*, 2011), au réticulum endoplasmique ou encore à l'appareil de Golgi (les TNLs RPP1 ou L6) (Michael Weaver *et al.*, 2006; Takemoto *et al.*, 2012). La double localisation nucléo-cytoplasmique est documentée pour de nombreux exemples tels N, Rx1, SNC1 ou RPS4 (Burch-Smith *et al.*, 2007; Cheng *et al.*, 2009; Slootweg *et al.*, 2010; Sohn *et al.*, 2014). Dans la plupart de ces exemples, la reconnaissance de l'effecteur s'opère dans le cytosol puis la relocalisation et l'accumulation du NBS-LRR dans le noyau sont des étapes cruciales pour le déclenchement de la HR (Slootweg *et al.*, 2010). L'étude des séquences permet de prédire leur localisation basale mais seules les études *in planta* ont permis de révéler précisément leur localisation en absence et en présence des pathogènes qu'ils contrôlent.

1.2.5 Le niveau d'expression et la régulation des NBS-LRRs

La plupart des NBS-LRRs sont faiblement transcrits de façon constitutive. En présence d'un pathogène, ils peuvent alors être fortement induits très localement (McHale *et al.*, 2006). Par exemple, le niveau d'expression du gène *PsoRPM2* ('Prunus sogdiana Resistance Protein to Meloidogyne incognita 2') après inoculation avec le nématode à galles *M. incognita* conditionne directement le succès ou l'échec du processus de résistance à ce parasite (Zhu *et al.*, 2017). Le niveau d'expression des NBS-LRRs peut également varier en fonction du tissu considéré. Nombre d'espèces d'Angiospermes semblent avoir un niveau d'expression des NBS-LRRs plus élevé dans les racines que dans les pousses alors que les Brassicaceae (*Arabidopsis* inclus) semblent avoir un ratio inversé reflétant une possible adaptation à la symbiose (Munch *et al.*, 2018). Les facteurs de transcription jouent évidemment un rôle important dans le niveau d'expression des NBS-LRRs. La plupart des NBS-LRRs présentent dans leur région promotrice un enrichissement en sites de fixation pour facteurs de transcription de type « adaptation aux stress biotiques », « HSP (Heat Shock Protein) », « réponse au stress hydrique » ou « réponse à l'acide abscissique ».

Par ailleurs, l'épissage alternatif est un élément prépondérant de la régulation post-transcriptionnelle dans la réponse qu'apportent les NBS-LRRs aux différents stress (Yang *et al.*, 2014). Un nombre important de gènes de résistance sont concernés par ce phénomène et notamment les TNLs. En effet, les TNLs diffèrent des CNLs et RNLs par une structure exon/intron plus séquentielle avec fréquemment un exon codant pour un domaine (**Figure 8**), favorisant ainsi l'apparition de transcrits alternatifs par exons alternatifs ou par rétention d'introns (Mastrangelo *et al.*, 2012) comme identifié chez *N* ou *BS4* (Dinesh-Kumar & Baker, 2000; Schornack *et al.*, 2004). Ainsi le ratio entre le transcrit tronqué du gène *N* et son

transcrit pleine longueur se trouve modifié lors de l'infection et est essentiel au bon fonctionnement de la résistance (Jordan *et al.*, 2002). Dans ce système la présence des deux isoformes est capitale afin de produire une résistance complète (Dinesh-Kumar & Baker, 2000). Par ailleurs les pathogènes influent sur la production des différentes isoformes de NBS-LRRs. En effet, une étude menée chez *Brachypodium distachyon* après infection avec le *Panicum mosaic virus* a mis en lumière des phénomènes d'épissage alternatif, dus au virus, chez plus de 100 gènes impliqués dans l'immunité dont une dizaine de NBS-LRRs (Mandadi & Scholthof, 2015). La régulation de la quantité des différentes isoformes de NBS-LRRs est fortement influencée par le mécanisme de dégradation des ARNm non-sens (NMD) (Rayson *et al.*, 2012). L'altération de ce mécanisme a pour conséquence l'apparition d'une auto-immunité largement provoquée par la régulation aberrante des TNLs (Riehs-Kearnan *et al.*, 2012). Les NBS-LRRs sont aussi régulés par les miRNA et siRNA qui cibleront préférentiellement les domaines/motifs conservés (motif P-loop du domaine NBS – voir paragraphe 1.2.1.1). Ces ARNs non-codants interviendront dans le maintien du niveau basal d'expression des NBS-LRRs, le retour à la normale après la phase d'infection, mais aussi dans l'aide à l'installation des organismes symbiotiques en sous-régulant les NBS-LRRs (Zhai *et al.*, 2011). Similairement, les niveaux protéiques de NBS-LRRs restent faibles à l'état basal et sous forte régulation post-traductionnelle. Pour être structurellement repliés correctement, les NBS-LRRs interagissent avec le complexe HRS (HSP90-SGT1-RAR1) évoqué en 1.2.4 (Kadota *et al.*, 2010). Cette plaque tournante garantit leur repliement, la production de protéines stables inactives évitant les phénomènes d'auto-immunité néfastes pour la plante mais joue aussi un rôle dans la balance nucléocytoplasmique de ces protéines (Zhang *et al.*, 2004). Les NBS-LRRs peuvent subir diverses modifications post-traductionnelles (acétylation, sumoylation) et être ubiquitinylés pour être dirigés vers le protéasome afin d'y être dégradés. La protéine 'constitutive expresser of PR genes 1' (CPR1) joue un rôle essentiel dans le renouvellement des NBS-LRRs (Gou *et al.*, 2012).

	Gène	Espèce source	Résistances connues
	<i>Ma</i>	<i>Prunus cerasifera</i>	Nématodes (<i>Meloidogyne spp.</i>)
	<i>Rdr1</i>	<i>Rosa multiflora</i>	Champignons (<i>Diplocarpon rosae</i>)
	<i>Bs4</i>	<i>Lycopersicon esculentum</i>	Bactéries (<i>Xanthomonas campestris</i>)
	<i>Gro1-4</i>	<i>Solanum tuberosum</i>	Nématodes (<i>Globodera rostochiensis</i>)
	<i>Y-I</i>	<i>Solanum tuberosum</i>	Virus (PVY)
	<i>N</i>	<i>Nicotiana tobaccum</i>	Virus (TMV)
	<i>RPS4</i>	<i>Arabidopsis thaliana</i>	Bactéries (<i>Pseudomonas syringae</i>)
	<i>RAC1</i>	<i>Arabidopsis thaliana</i>	Oomycètes (<i>Albugo candida</i>)
	<i>CSA1</i>	<i>Arabidopsis thaliana</i>	Adaptation développemental à la lumière
	<i>WRR4</i>	<i>Arabidopsis thaliana</i>	Oomycètes (<i>Albugo candida</i>)
	<i>RPP5</i>	<i>Arabidopsis thaliana</i>	Oomycètes (<i>Peronospora parasitica</i>)
	<i>RLM1</i>	<i>Arabidopsis thaliana</i>	Champignons (<i>Melampsora lini</i>)
	<i>RJ2</i>	<i>Glycine max</i>	Restraint la nodulation par <i>Sinorhizobium fredii</i>
(*)	<i>M</i>	<i>Linum usitatissimum</i>	Champignons (<i>Melampsora lini</i>)
(*)	<i>L6</i>	<i>Linum usitatissimum</i>	Champignons (<i>Melampsora lini</i>)
(*)	<i>RRS1</i>	<i>Arabidopsis thaliana</i>	Bactéries (<i>Ralstonia solanacearum</i>)
(*)	<i>DAR4</i>	<i>Arabidopsis thaliana</i>	Adaptation au stress thermique (froid)

Figure 8 - Annotation structurale de différents TNLs caractérisés avec leur rôle identifié.

Chaque exon codant pour un domaine protéique donné est identifié avec une couleur différente. Les introns sont indiqués avec une ligne bleue (séparant des exons codant pour des domaines protéiques différents) ou noire en 'Λ' (séparant des exons codant pour un même domaine protéique). *indique les TNLs ne disposant pas de domaine PL (vert).

1.2.6 L'évolution des NBS-LRRs

1.2.6.1 Aux origines des NBS-LRRs

Les NBS-LRRs trouvent leur origine très tôt dans la lignée végétale. Une étude récente conduite sur des transcriptomes d'algues vertes (Charophytes ; Coleochaetales ; Streptophyte algae), ancêtres des plantes terrestres (Embryophytes), a mis en évidence des TNLs et CNLs tels qu'on les connaît chez les plantes dites évoluées (Gao *et al.*, 2018). L'antériorité présumée des CNLs sur les TNLs, suggérée notamment par leur structure plus séquentielle, reste aujourd'hui spéculative du fait que des séquences des deux sous-familles ont été identifiées chez les algues vertes. Avant cette étape et chez les organismes plus anciens tels que les bactéries et archéobactéries, les différents domaines (TIR, NBS, LRR) n'existaient que sous la forme de domaines indépendants (Yue *et al.*, 2012). Ainsi, il semble que la fusion des domaines NBS-LRRs, se soit opérée avant la conquête terrestre des plantes (**Figure 9**).

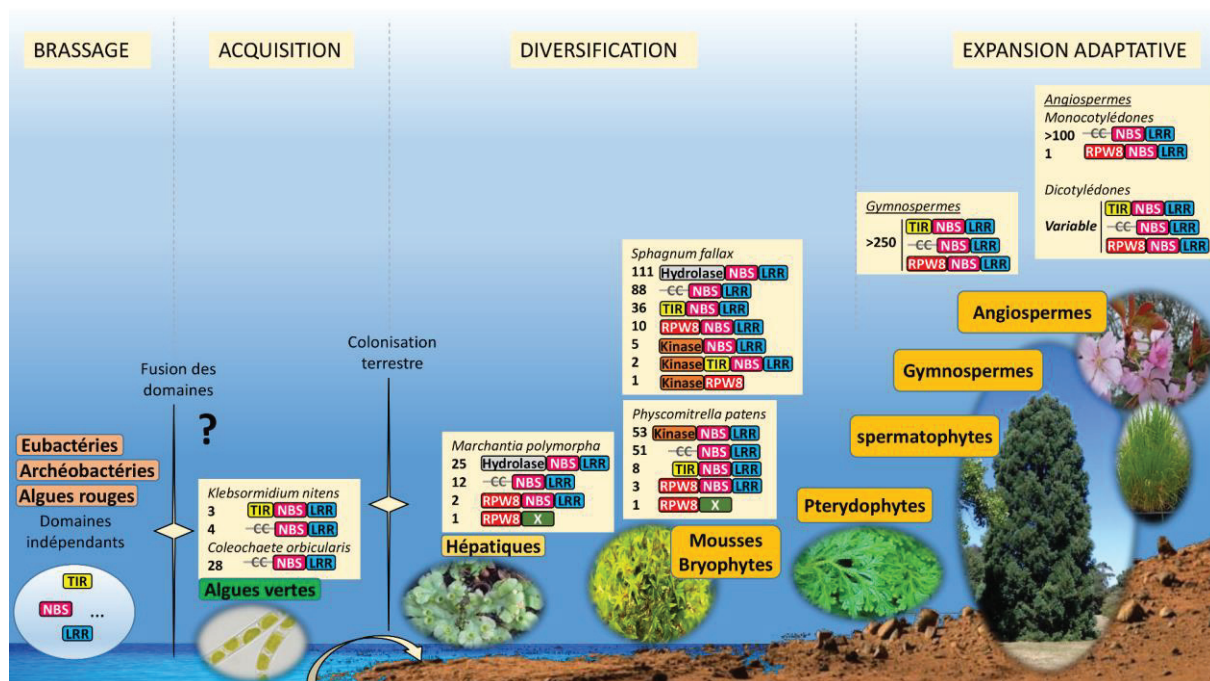


Figure 9 - Origine et organisation structurale des NBS-LRRs au cours de l'évolution des plantes.

L'annotation fonctionnelle des NBS-LRRs pour chaque groupe taxonomique/espèce est indiquée avec les effectifs respectifs dans les encarts jaunes. Les références associées sont mentionnées dans le texte et/ou en Annexe 4.

L'étude de génomes de plantes terrestres anciennes, telles que l'hépatique *Marchantia polymorpha* et les mousses *Sphagnum fallax* et *Physcomitrella patens*, a mis en lumière une diversité importante de NBS-LRRs. En plus des trois sous-familles (TNLs, RNLs et CNLs), des hydrolase-NBS-LRRs et des kinase-NBS-LRRs ont été identifiées en nombre important, indiquant une expansion des possibilités de défense essentiellement dans la fonction d'activation de la voie de défense. Cette diversification a pu

jouer un rôle important dans la colonisation du milieu terrestre (Xue *et al.*, 2012; Gao *et al.*, 2018). Ainsi, différentes stratégies comprenant divers récepteurs immuns semblent avoir été testées avant que les spermatophytes incluant Gymnospermes et Angiospermes ne basent l'essentiel de leur arsenal de défense sur les trois sous-familles principales, TNLs, RNLs et CNLs.

L'apparition des NBS-LRRs pourrait avoir résulté d'une intégration de blocs préexistants remplissant chacun une fonction bien déterminée et initiée par une structure de type NTPase-TPR (Urbach & Ausubel, 2017). Ce résultat, associant un domaine d'activité nucléoside tri-phosphatase (NTPase) tels les domaines NBS (chez les plantes) et NACHT/NOD (chez les animaux) avec un domaine impliqué dans les interactions, constitué de répétitions TPR ('**T**etratrico**P**eptide **R**epeat'), indique que les NBS-LRRs et les NACHT-LRRs n'ont pas évolué à partir d'un ancêtre commun (Urbach & Ausubel, 2017). En effet, il s'agirait plutôt d'un phénomène d'évolution convergente entre les plantes et les animaux avec également l'ajout de domaines de signalisation N-terminaux différents chez les plantes (TIR, RPW8, ...) et chez les animaux (CARD, BIR, PYD, ...) (Meunier & Broz, 2017). L'objectif commun de ces assemblages est de pouvoir distinguer exclusivement les molécules du non-soi (rôle des domaines répétés), passer à un état d'activation pour une durée limitée (rôle des domaines NTPase) et déclencher une PCD ciblée (rôle des domaines activateurs) (Urbach & Ausubel, 2017). Par ailleurs, les autres composants de l'immunité tels que les récepteur lectine kinases ainsi que les éléments caractéristiques de la voie de signalisation des NBS-LRRs (NDR1, HSP90, EDS1, PAD4, SAG101, SGT1, RAR1 et WRKY) ont également déjà été mis en place avant la séparation procaryotes/eucaryotes (Yue *et al.*, 2012).

1.2.6.2 Distribution des NBS-LRRs à travers les génomes de plantes

Pour répondre aux défis imposés par les pathogènes, les plantes terrestres disposent d'un nombre important de NBS-LRRs approchant fréquemment 1% de l'ensemble des gènes prédits pour un génome donné. Dès le début des années 2000, les NBS-LRRs ont fait l'objet de travaux de recensement et de classification afin de les identifier et de mieux comprendre leur évolution (Meyers *et al.* 2003). Depuis, des dizaines d'études ont annoté cette famille de gènes à travers de nombreux génomes d'Angiospermes. A l'image du nombre global de gènes codant des protéines au sein d'un génome donné, le nombre de NBS-LRRs n'est pas proportionnel à la taille des génomes, ni à leur ploïdie (Levy & Feldman, 2002; Borrelli *et al.*, 2018). Ainsi, on observe une différence modérée entre les espèces à petits génomes (ex : *A. thaliana* ou le peuplier) et les « giga-génomes » de certaines Monocotylédones et des Gymnospermes (**Figure 10**). Même si des variations existent entre les groupes taxonomiques, les effectifs de NBS-LRRs se situent dans une fourchette large entre quelques unités et un millier par génome (**Figure 10**). Quelques espèces font cependant exception, comme le pommier ou le blé commun, illustrés par, respectivement, les points bleu et jaune les plus élevés (**Figure 10**), qui ont connu une expansion massive de cette famille.

Par ailleurs, le nombre de NBS-LRRs contenus dans les génomes de plantes ne semble pas corrélé au degré d'évolution de l'organisme considéré puisque même chez les bryophytes on peut détecter des séquences de cette famille de gènes en effectifs comparables (243 et 123 chez, respectivement, *S. fallax* et *P. patens*) à ce que l'on retrouve chez les Angiospermes évoluées (Xue *et al.*, 2012).

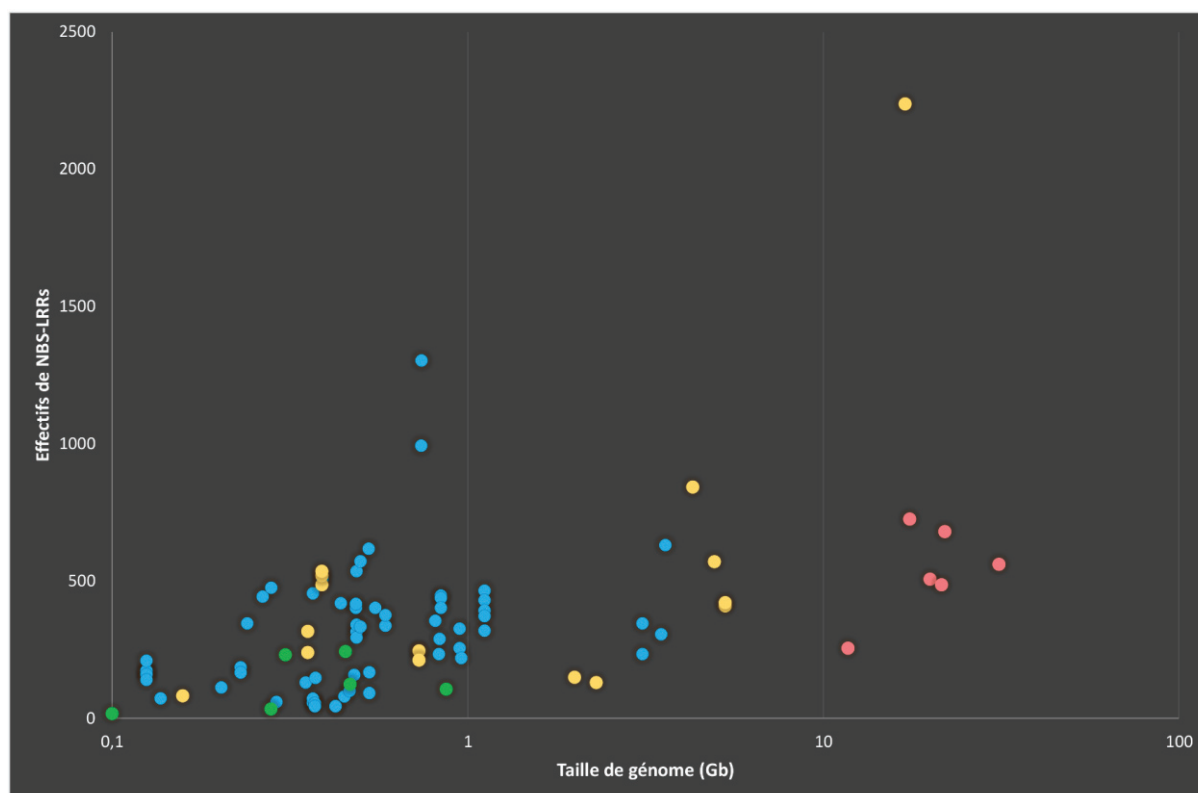


Figure 10 - Relation entre le nombre de NBS-LRRs pour une espèce donnée et la taille de son génome.

Les Bryophytes et plantes archaïques sont représentées en vert, les Angiospermes Dicotylédones en bleu, les Monocotylédones en jaune et les Gymnospermes en rouge. Les effectifs et leurs références sont détaillés dans l'Annexe 4.

Il est à noter que le dénombrement de cette famille au sein d'une même espèce peut montrer des différences significatives en fonction des études. Certaines espèces d'intérêt fondamental et/ou évolutif, analysées au cours d'études indépendantes, ont livré des effectifs variables et ceci même chez l'espèce modèle *A. thaliana* (Meyers *et al.*, 2003; Guo *et al.*, 2011; Hofberger *et al.*, 2014; Gao *et al.*, 2018) (**Figure 11A**). Des facteurs tels que (i) le choix du matériel biologique utilisé pour le séquençage (différents clones), (ii) le type de données analysées (transcriptomes/génomes complets ou partiels), (iii) les caractéristiques des données (complexité des génomes, qualité des assemblages), (iv) la méthodologie de détection et d'annotation (prédictions de gènes), sont en mesure d'expliquer de tels biais. L'inclusion ou non de séquences tronquées (contenant seulement un ou deux domaines caractéristiques des NBS-LRRs) peut également contribuer à ces variations. Des études ciblées ont montré que des séquences contenant un ou deux domaines seulement pouvaient jouer un rôle

particulièrement important : c'est le cas par exemple de RPW8.1 et RPW8.2, constitués du seul domaine RPW8, qui confèrent une résistance à l'oïdium (Xiao *et al.*, 2001), de RBA1 qui ne comporte qu'un domaine TIR et qui est capable de reconnaître l'effecteur bactérien de type III, HopBA1, et d'induire une mort cellulaire (Nishimura *et al.*, 2017) ou encore de TN2 associant seulement un domaine TIR et un domaine NBS impliqués dans la surveillance de cibles potentielles d'effecteurs (Zhao *et al.*, 2015). Ces résultats encouragent donc à réaliser des études les plus exhaustives possibles et par conséquent d'intégrer ces séquences tronquées.

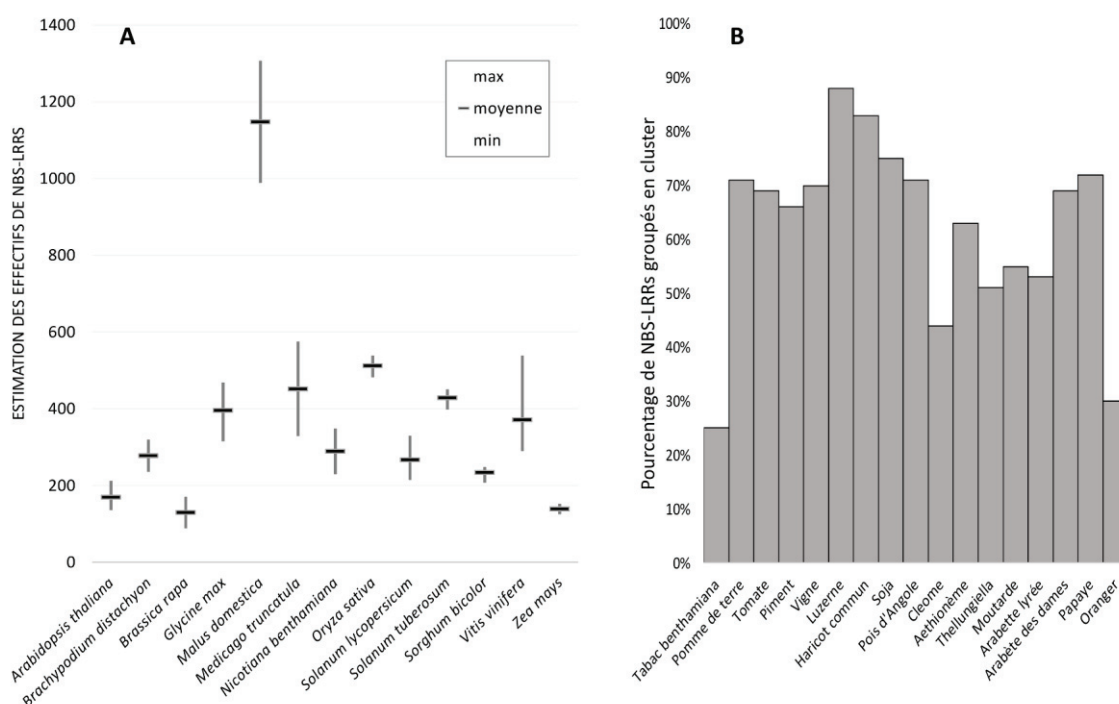


Figure 11 - Variation dans les estimations de NBS-LRRs et distribution des NBS-LRRs dans divers génomes de plantes.

(A) Estimation des effectifs de NBS-LRRs d'une espèce donnée en fonction des études. (B) Pourcentage de NBS-LRRs organisés en cluster pour un génome donné. Les effectifs et leurs références sont détaillés dans l'Annexe 4.

Ces études ont mis également en lumière des ratios extrêmement variables entre TNLs et CNLs selon les génomes de plantes considérés. Certains groupes phylétiques ont une répartition équilibrée ou légèrement en faveur des TNLs (Brassicaceae ou Rosaceae), d'autres une prépondérance des CNLs (Solanaceae) jusqu'au cas extrême des groupes ne disposant plus de TNLs (Monocotylédones) (**Annexe 4**).

Hormis l'intérêt évolutif d'étudier la distribution génomique des NBS-LRRs, la cartographie de ces gènes est d'un intérêt majeur pour les sélectionneurs et pour l'utilisation effective de ces sources de résistance en arboriculture, sylviculture et agriculture en général. Globalement, la répartition des NBS-

LRRs n'est pas homogène à travers les différents chromosomes d'un génome (Meyers *et al.*, 2003). Même si certains NBS-LRRs se retrouvent isolés (singletons), ils sont essentiellement organisés en clusters de taille variable, certains clusters pouvant contenir plusieurs dizaines de NBS-LRRs (Baumgarten *et al.*, 2003). Les cytochromes P450 ou les RLKs sont d'autres exemples de grandes familles disposant d'une distribution génomique similaire (Schuler & Werck-Reichhart, 2003; Shiu & Bleecker, 2003). Ce taux de 'clustérisation' des NBS-LRRs reste assez élevé chez de nombreuses espèces (**Figure 11B**) mais son évaluation précise reste largement dépendante de la qualité des assemblages et de l'annotation des génomes. On peut noter une similarité importante entre les séquences de gènes appartenant à un même cluster. Ceci a pour conséquence, le plus souvent, d'associer le positionnement de ces séquences dans la phylogénie et leur localisation physique dans le génome. La théorie de mort et naissance des NBS-LRRs (Michelmore *et al.*, 1998) prévoit que le transfert d'un NBS-LRR peut s'effectuer à longue distance par duplication segmentale ou crossing-over et que ce NBS-LRR va éventuellement se dupliquer en tandem en cas d'expansion adaptative et former un néo-cluster. Différents mécanismes sont à l'origine de ces expansions et contractions des effectifs et de la diversité des NBS-LRRs.

1.2.6.3 Expansions et contractions des NBS-LRRs

La famille des NBS-LRRs est extrêmement dynamique. Son expansion chez certaines espèces comme le pommier (*Malus domestica*) s'explique probablement par (i) une duplication globale récente de son génome en comparaison avec ses cousins les *Prunus*, puis sans doute par des duplications locales par rapport au poirier (*Pyrus x bretschneideri*), une espèce encore plus proche (Jia *et al.*, 2015). La sélection imposée notamment par la domestication d'espèces telles que le pommier ou le poirier ou par l'exposition à une grande variété de pathogènes pourrait expliquer la rétention d'un nombre important de NBS-LRRs (Jia *et al.*, 2015). Hormis les duplications de génomes entiers, différents modes de duplications sont identifiés comme responsables de l'expansion de cette famille (Baumgarten *et al.*, 2003). On trouve des duplications en tandem, ectopiques ou segmentales en fonction de l'échelle à laquelle s'effectue la duplication (**Figure 12**) (Leister, 2004). Deux études menées chez trois génomes de légumineuses et trois Solanacées ont montré que la majorité des duplications de NBS-LRRs s'opèrent à courte distance (duplication locale en tandem) puis, dans une moindre mesure, par duplication ectopique et, plus rarement, par duplication segmentale (Shao *et al.*, 2014; Qian *et al.*, 2017).

Ces processus peuvent avoir des conséquences directes sur le phénotype. Ils peuvent être générateurs de diversité par réarrangement de domaines et notamment du domaine LRR mais aussi par intégration de nouveaux domaines (**Figure 13**). La proximité physique des clusters de NBS-LRRs avec d'autres protéines liées aux voies de défense combinée avec le processus de duplication locale des NBS-LRRs favorise l'intégration de domaines liés aux mécanismes de défense comme les domaines kinase, WRKY, HMA, WD40, BED zinc finger...

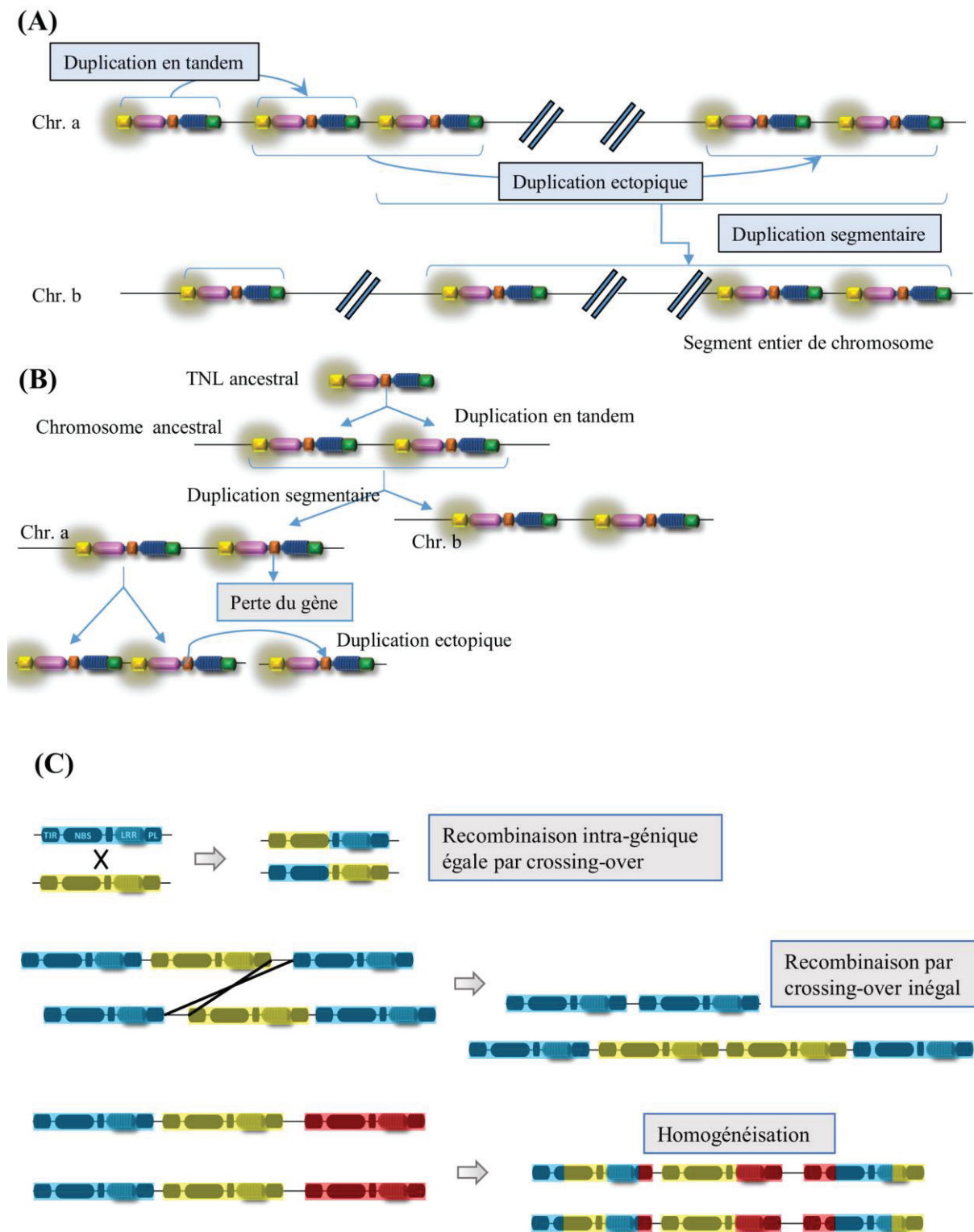


Figure 12 - Mécanismes génétiques intervenant dans l'évolution des gènes de résistance de type NBS-LRR. (A) et (B) Les trois types de duplications se produisant chez les NBS-LRRs d'un point de vue spatial (A) et évolutif (B). (C) La recombinaison chez les NBS-LRRs : créatrice de diversité mais aussi responsable de pertes de gènes. Adapté de (Leister, 2004; Shao *et al.*, 2014; Cardoso-Moreira *et al.*, 2016)

Ces duplications partielles ou complètes sont génératrices de diversité mais aussi de pseudogénéisation (apparition de décalage de cadre de lecture et de codons stop prématures par exemple) (**Figure 13**).

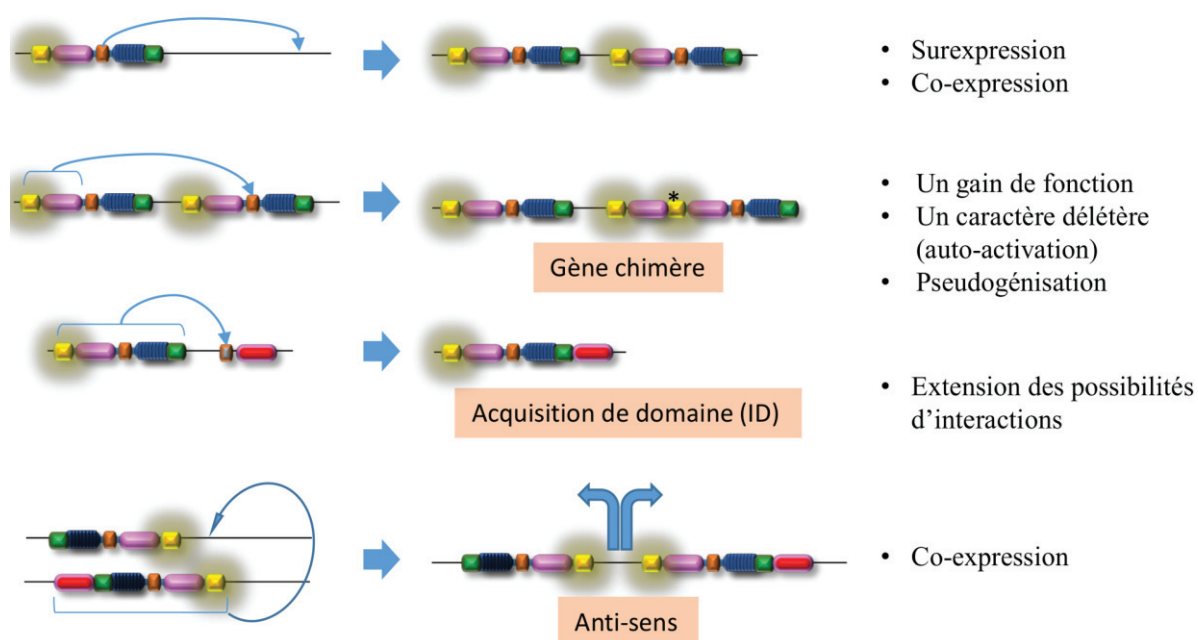


Figure 13 - Quelques conséquences possibles sur le phénotype exprimé par les plantes suite aux mécanismes génétiques mis en jeu par les NBS-LRRs.

L'isolement d'un NBS-LRR issu de recombinaison (singleton) favorisera une nouvelle spécificité de résistance du fait qu'il n'est pas soumis à des échanges inter-géniques favorisant l'homogénéisation (**Figure 12C**). Ceci se réalise d'autant plus facilement que la plante en retire un avantage adaptatif en réponse à un stress donné. En contrepartie ces gènes ou clusters peuvent être perdus, par divers mécanismes génétiques (crossing-over inégal, réparation homologue ou non-homologue) pour des raisons de sélection ou de surexpression néfaste pour la plante (Luo *et al.*, 2012). Par leur mobilité, les éléments transposables jouent également un rôle important dans le processus de « mort » des NBS-LRRs en s'insérant fréquemment dans les régions régulatrices ou directement dans un NBS-LRR provoquant une pseudogénéisation (Duval *et al.*, 2014). Ainsi ces divers mécanismes peuvent conduire au gain et à la perte rapide de centaines de NBS-LRRs entre deux espèces proches et leur ancêtre commun. C'est le cas notamment chez le haricot (*Phaseolus vulgaris*) et le soja (*Glycine max*) qui ont divergé il y a 19 Ma (Shao *et al.*, 2014). De même, bien que les deux espèces du genre *Oryza*, *O. japonica* et *O. indica*, n'aient divergé que depuis 0.1 Ma, 34% de leurs loci NBS-LRR ne sont déjà plus présents que dans l'une d'elles (Wei *et al.*, 2012; Luo *et al.*, 2012).

Au final, la modularité, le dynamisme génétique, les effectifs ainsi que l'étendue de leur réseau font des NBS-LRRs une pierre angulaire de la réponse des plantes aux différents stress (biotiques et abiotiques) exerçant une pression constante sur les plantes. Parmi ces menaces biotiques, les nématodes parasites des plantes représentent un groupe de pathogènes particulièrement diversifié et agressif.

1.3 Les nématodes

Les nématodes (phylum Nematoda) sont des vers ronds non segmentés ayant colonisé la plupart des milieux. La majorité d'entre eux sont parasites et représentent une part importante du règne animal et une diversité qui est largement sous-évaluée, avec plus de 26 000 espèces connues (Hugot *et al.*, 2001). Ce groupe taxonomique possède son organisme modèle de référence (*C. elegans*) ainsi que des représentants parasitant la plupart des espèces du règne animal jusqu'aux plus évoluées, mais aussi du règne végétal, dits nématodes phytoparasites.

1.3.1 Les nématodes phytoparasites

Les nématodes phytoparasites causent des pertes massives pour l'agriculture mondiale. Celles-ci, sont estimées à plus 150 milliards de dollars par an (Nicol *et al.*, 2011; Singh *et al.*, 2015). Certaines espèces sont également impliquées dans des interactions avec d'autres agents pathogènes tels que les virus (Demangeat *et al.*, 2010), les bactéries (Rubio-Cabetas *et al.*, 2001) ou les insectes (Mota *et al.*, 2006) accentuant l'impact des nématodes phytoparasites.

Environ 4100 espèces de nématodes phytoparasites sont actuellement décrites (Decraemer & Hunt, 2006) et appartiennent majoritairement aux ordres Tylenchida et Dorylaimida (**Figure 14**). Plusieurs stratégies de parasitisme, qui ciblent le plus souvent le système racinaire, peuvent être décrites selon la localisation et la mobilité du nématode lors de son interaction avec la plante. Les ectoparasites demeurent à l'extérieur des tissus végétaux et resteront mobiles pour ponctionner les cellules à l'aide de leur stylet. Ce groupe comprend notamment les nématodes vecteurs de virus (genres *Xiphinema*, *Longidorus* et *Trichodorus*) qui peuvent transmettre à la plante une charge virale lors de leur nutrition. Les semi-endoparasites sont partiellement engagés dans les tissus végétaux avec une ponte se produisant à l'extérieur de la plante (*Tylenchulus semipenetrans*). Enfin, les endoparasites réalisent la majeure partie de leur cycle de vie à l'intérieur des tissus végétaux. Parmi ceux-ci on trouve les migrants, comme les espèces du genre *Pratylenchus*, et les sédentaires tels que les nématodes à kyste (*Globodera* spp. et *Heterodera* spp.) et les nématodes à galles du genre *Meloidogyne*.

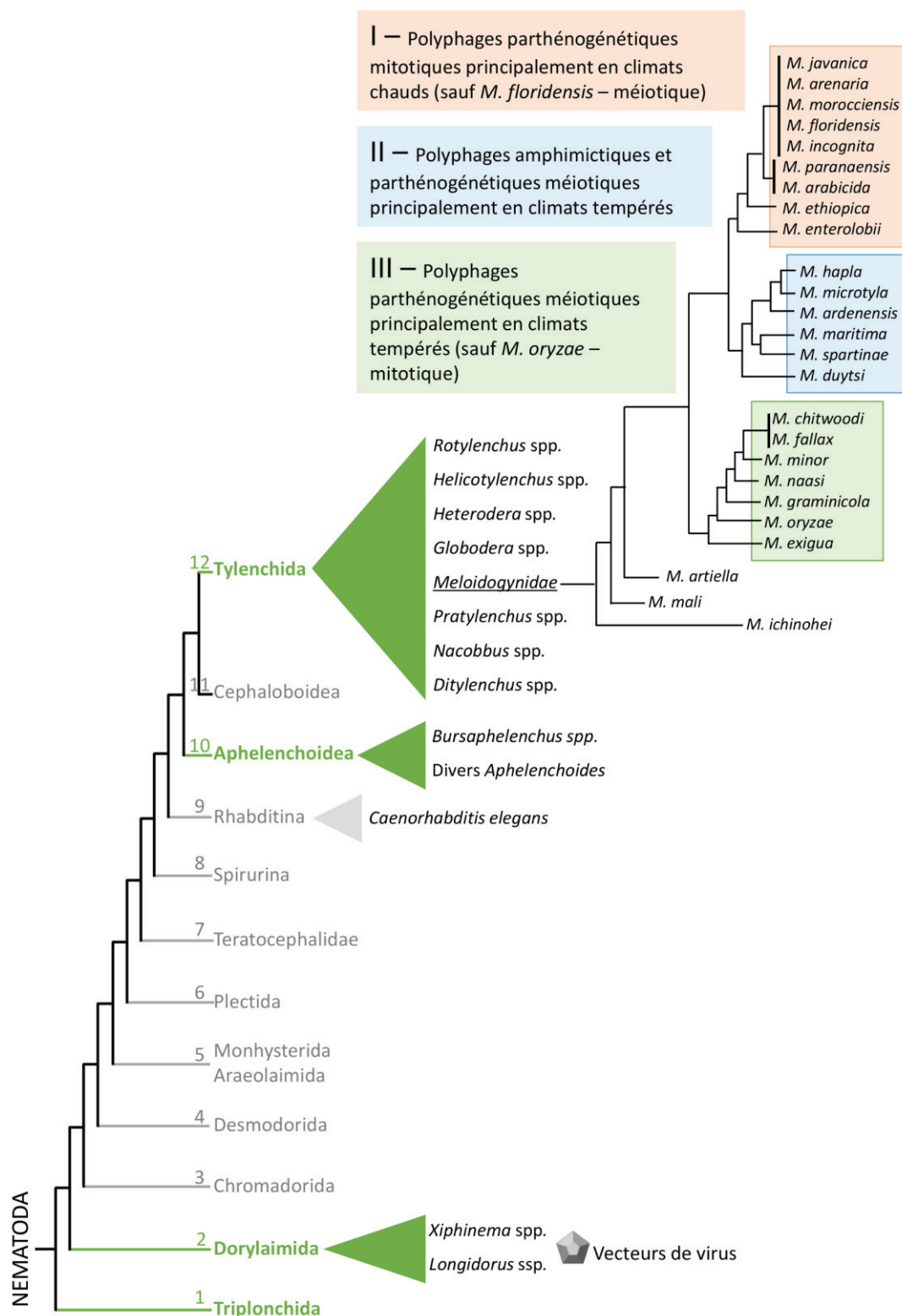


Figure 14 - Représentation schématique du phylum Nematoda et position des nématodes à galles du genre *Meloidogyne*.

Parmi les 12 clades, ceux contenant les nématodes phytoparasites sont indiqués en vert et illustrés avec quelques genres majeurs (adapté à partir de (Quist *et al.*, 2015)). Les traits de vie et reproductifs des *Meloidogyne* sont indiqués pour les trois clades principaux (I à III). Adapté de (Holterman *et al.*, 2009)

1.3.2 Les nématodes à galles du genre *Meloidogyne*

Les espèces endoparasites de nématodes sont les espèces les plus nuisibles à l'agriculture à l'échelle de la planète. Parmi celles-ci, les nématodes à galles du genre *Meloidogyne* se placent en tête par leur impact économique et l'intérêt scientifique qui leur est porté (Jones *et al.*, 2013). La distribution mondiale des nématodes à galles et leur grande polyphagie expliquent en grande partie cette importance. Dès 1965, plus de 5500 espèces de plantes susceptibles d'être parasitées par ce groupe de nématodes avaient été répertoriées (Goodey *et al.*, 1965). La plupart des espèces végétales cultivées sont en fait touchées par ces parasites. A ce jour 98 espèces appartenant au genre *Meloidogyne* sont décrites (Jones *et al.*, 2013) parmi lesquelles quatre espèces, *M. incognita*, *M. arenaria*, *M. javanica* et *M. hapla* (**Figure 14**) sont responsables à elles seules de 90% des dégâts engendrés par ces nématodes au niveau mondial (Barker *et al.*, 1985). *M. incognita*, dont le génome a été séquencé en 2008 (Abad *et al.*, 2008), est l'espèce modèle du genre. Avec les modifications climatiques et les échanges commerciaux accrus, des espèces décrites plus récemment telles que *M. enterolobii*, *M. floridensis* et *M. ethiopica* sont en émergence. Les nématodes à galles contiennent également les espèces de quarantaine, *M. chitwoodi* et *M. fallax*, uniquement présentes dans les régions septentrionales sur les espèces annuelles.

1.3.3 Le cycle de vie et l'anatomie des *Meloidogyne*

En plus de leur polyphagie, des paramètres inhérents à leur cycle de vie sont déterminants pour expliquer l'impact de ces parasites. En effet, à 25-28°C, ce cycle peut être complété en moins d'un mois chez *M. incognita*. A l'issue de son cycle, chaque femelle est capable de produire environ un millier d'œufs (Castagnone-Sereno *et al.*, 2013) qui sont inclus dans une masse gélatineuse disposée à la surface de la racine ou dans la racine elle-même (**Figure 12**). Lorsque les conditions environnementales sont favorables, chaque œuf libère directement une larve de 2^{ème} stade (J2) car la mue J1-J2 a lieu dans l'enveloppe de l'œuf. Ces J2s sont filiformes et mobiles dans le sol (phase exophyte) et, attirées par les exsudats racinaires, elles pénètrent à l'intérieur de leur hôte. Cette pénétration a lieu préférentiellement au niveau de l'apex des racines en croissance (début de la phase endophyte). Une fois à l'intérieur, les larves migrent entre les cellules en direction de l'extrémité apicale de la racine grâce à l'utilisation d'enzymes de dégradation (étape 1 – **Figure 15**). Après avoir atteint la zone méristématique terminale, les J2s effectuent un demi-tour pour remonter le long du cylindre central et trouver un site favorable dans celui-ci (étapes 2 et 3 – **Figure 15**). Les larves vont alors se fixer définitivement (étape 4 – **Figure 15**) et induire leur site nourricier constitué de cellules géantes (Khallouk *et al.*, 2011). Le nématode injecte des molécules majoritairement produites dans ses glandes œsophagiennes (deux glandes sub-ventrales et une glande dorsale) via son stylet. Elles induisent des modifications des parois (migration) puis des cellules du site nourricier (fixation) de son hôte. En effet par cette action du nématode, la machinerie cellulaire végétale est activée et génère, à la suite de divisions nucléaires sans cytotélière,

des cellules géantes, multi-nucléées, et particulièrement enrichies en nutriments qui favorisent son développement et aboutissent à la formation de galles. Dans ce processus, les *Meloidogyne* détournent amplement le métabolisme cellulaire à leur profit pour s'alimenter. Les larves J2 se développent et évoluent en stades juvéniles 3 (J3) puis 4 (J4). La mue de ces J4s aboutit au stade adulte femelle pyriforme (**Figure 15**). Ces femelles produisent alors une masse d'œufs qui, une fois libérés, démarrent un nouveau cycle (Abad *et al.*, 2008). Des mâles sont présents et leur effectif est d'autant plus élevé que les larves J2 sont confrontées à des conditions défavorables de développement (très forte densité rendant difficile la nutrition ou mauvais état des racines nourricières de la plante hôte). Les mâles restent filiformes, ne s'alimentent pas et quittent la racine. Les variations du sex-ratio reflètent une adaptation aux conditions du milieu (Triantaphyllou, 1973).

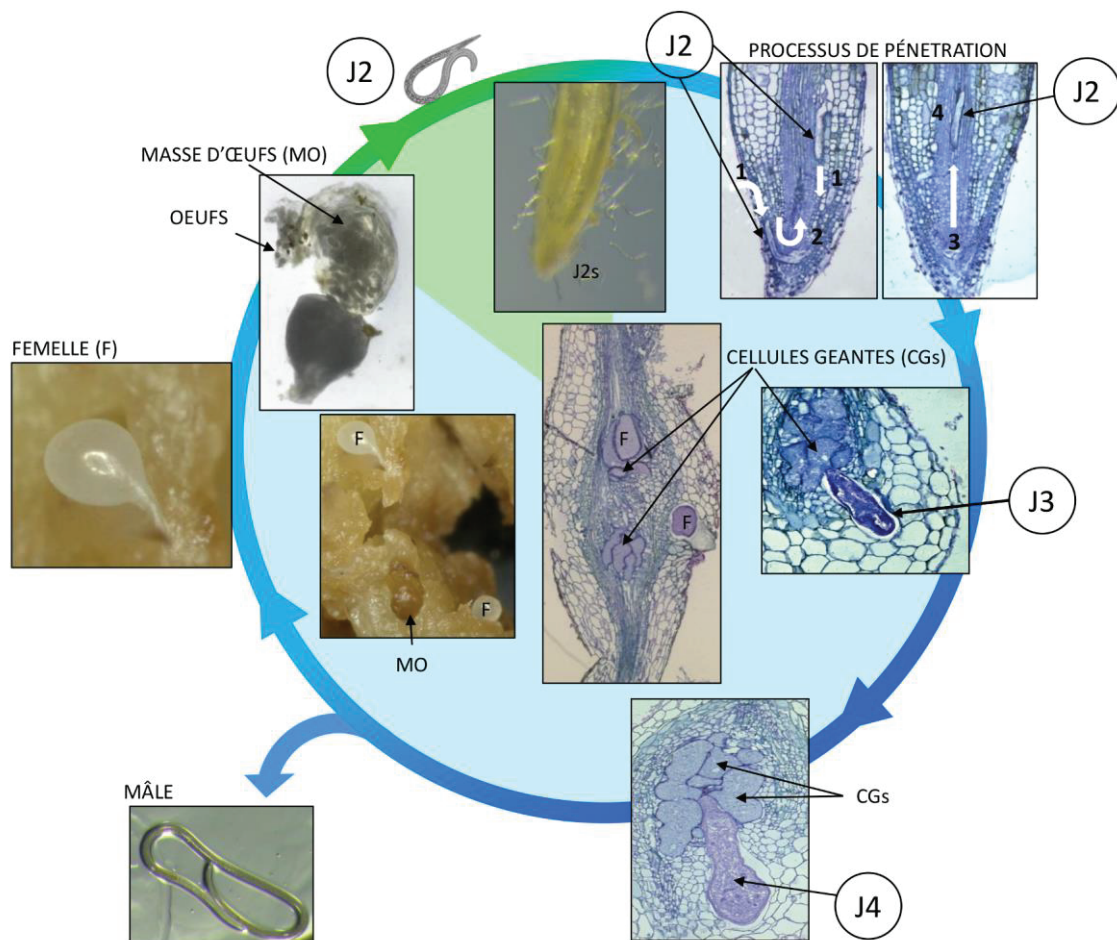


Figure 15 - Cycle de vie de *Meloidogyne incognita* sur racines de *Prunus cerasifera*.

La phase exophyte est illustrée en vert et la phase endophyte en bleu. La larve juvénile mobile de stade 2 (J2) migre en 4 étapes : elle pénètre au niveau de l'apex pour descendre (1) vers son extrémité et effectuer un demi-tour (2) avant de remonter le long du cylindre central (3) et d'atteindre l'emplacement du futur site nourricier (4). Les stades juvéniles 3 (J3) et 4 (J4) fixés sont visibles en coupe racinaire avec les cellules géantes. Représentation adaptée de (Abad *et al.*, 2008; Khallouk *et al.*, 2011).

1.3.4 *La reproduction et l'adaptation des Meloidogyne*

Parmi les nématodes à galles, on trouve des espèces parthénogénétiques mitotiques strictes (=clonales) dont les trois plus importantes sont, *M. incognita*, *M. arenaria* et *M. javanica* (**Figure 14**). Étonnamment, ce mode de reproduction qui pourrait être considéré comme peu générateur de diversité, ne semble pas s'opposer à l'adaptation à une large gamme d'hôte, une distribution géographique étendue ou à une capacité à contourner diverses sources de résistances végétales (Castagnone-Sereno & Danchin, 2014). Une étude menée sur *M. incognita* a montré que la majeure partie du génome est composée de paires de segments homologues mais très divergents qui ont probablement évolué indépendamment en l'absence de recombinaison sexuée, ce qui pourrait conférer à ce génome une plasticité que son mode de reproduction tend à exclure (Abad *et al.*, 2008; Castagnone-Sereno & Danchin, 2014).

A l'inverse, une espèce diploïde telle que *M. hapla* qui pourtant se reproduit par parthénogenèse méiotique (en absence de mâles) ou sexuée, modes qui génèrent une diversité génétique plus élevée (Thomas *et al.*, 2012), présente une répartition géographique et une gamme d'hôte plus restreinte que les espèces précédemment citées.

Enfin on trouve également chez les nématodes à galles, des espèces amphimictiques telles que *M. carolinensis* ou *M. megalyta* qui complètent l'éventail des possibilités de mode de reproduction.

Par ailleurs un autre mécanisme évolutif, indépendant du mode de reproduction, a contribué à l'adaptation de ces nématodes aux plantes. Il s'agit d'évènements de transferts horizontaux de gènes codant notamment pour des enzymes de dégradation de la paroi végétale, survenus probablement entre des champignons et bactéries et l'ancêtre commun du groupe des Tylenchida (Danchin *et al.*, 2010).

1.3.5 *Les effecteurs de nématodes : manipulateurs de cellules*

Afin de pouvoir parasiter et manipuler les cellules de leurs hôtes, les *Meloidogyne* s'appuient sur l'utilisation de molécules sécrétées appelées effecteurs. Ces effecteurs présentent une diversité importante reflétant la grande variabilité des cibles et des fonctions qui sont visées. Les effecteurs de nématodes phytoparasites, et particulièrement des nématodes à galles, font l'objet depuis plusieurs années d'efforts importants d'identification et de caractérisation fonctionnelle. Les effecteurs sécrétés dans la cellule hôte agissent de façon redondante et coopérative afin de (1) faciliter l'invasion, (2) contrer les défenses de la plante en ciblant les composants de la réponse immunitaire de la plante, le trafic de vésicules ou la signalisation hormonale et (3) favoriser la croissance et le développement du parasite en manipulant les fonctions cellulaires de la plante (Abad & Williamson, 2010). Les effecteurs sont produits dans divers compartiments sécrétoires du nématode parmi lesquels les trois glandes œsophagiennes sont de première importance. Leurs activités respectives semblent liées au stade de

l'infection : schématiquement les deux glandes sub-ventrales agissent dans les phases précoces et la glande dorsale plus tardivement pour la maintenance parasitaire dans le temps. Les premiers effecteurs identifiés chez *Meloidogyne* impliquaient des fonctions de dégradation de la paroi cellulaire (cellulases) (Rosso *et al.*, 1999), puis des familles telles que les calréticulines et 14-3-3 ont été identifiées (Jaubert *et al.*, 2002, 2004). Plus récemment et avec le développement de technologies haut-débit plusieurs centaines de protéines sécrétées ont été identifiées dans le genre *Meloidogyne* (revue par (Truong *et al.*, 2015)) et ont permis une estimation du 'secrétome' de *M. incognita* à environ 500 protéines (Bellaïfiore *et al.*, 2008). Parmi toutes ces protéines sécrétées, seul un nombre restreint présente des domaines protéiques connus. Cependant leur localisation ainsi que des données de silencing ou de sur-expression en plante ont permis de suggérer dans quel processus ils pouvaient être impliqués (revue par (Truong *et al.*, 2015)). Parmi les effecteurs les mieux caractérisés on retrouve Mi-CRT, une calréticuline ciblant la paroi cellulaire (apoplasme) des cellules géantes. L'effecteur Mi-EFF1 est lui retrouvé préférentiellement dans les noyaux des cellules géantes et semble être produit uniquement dans les stades fixés de *M. incognita*. Plus récemment, des effecteurs contenant des domaines lectine de type C, capables de lier des sucres en présence de calcium, ont été identifiés comme importants pour le parasitisme notamment en modulant la réponse immunitaire de la plante (Danchin *et al.*, 2013). L'effecteur Mj-FAR-1 chez *M. javanica* semble promouvoir l'induction des galles et la formation des cellules géantes du site nourricier chez la tomate et ainsi permettre le développement complet du nématode (Iberkleid *et al.*, 2013). Même si l'on commence à disposer d'une image moins parcellaire des processus dans lesquels sont impliqués ces effecteurs, le défi reste d'identifier précisément leurs cibles. Nombre d'entre elles ont déjà été identifiées chez des pathogènes moins évolués tel que le modèle bactérien *P. syringae* sur *A. thaliana* qui illustre bien la diversité des actions combinées réalisées par les effecteurs (clivage, protection, phosphorylation, ou uridylylation de protéines hôtes) (Lee *et al.*, 2015). En revanche les couples cibles/effecteurs identifiés chez les nématodes tels que elf18/Mi-CRT, scarecrow-like TF/16D10 ou TIP2/8D05 sont beaucoup plus rares (Huang *et al.*, 2006; Jaouannet *et al.*, 2013; Xue *et al.*, 2013).

Ainsi de nombreux effecteurs sont engagés dans le parasitisme d'une grande diversité d'espèces végétales. Parmi celles-ci, on trouve les espèces pérennes du genre *Prunus* d'une grande importance économique et que leur statut de plante ligneuse ne préserve pas des attaques des *Meloidogyne*.

1.4 Le genre *Prunus*

En arboriculture, les variétés tout comme les porte-greffes sont impactés de façon récurrente par les pathogènes des cultures. Les différentes espèces du genre *Prunus* qui, au niveau mondial, revêtent un intérêt majeur en matière d'ornement mais surtout en matière de production fruitière, ne dérogent pas à ce constat.

1.4.1 Classification et positionnement des *Prunus*

Le genre *Prunus* L. (Rosaceae) compte environ 200 espèces (Chin *et al.*, 2014) dont les productions fruitières les plus importantes économiquement sont les pêches, prunes, amandes, abricots et cerises. Les *Prunus* font partie des 30 cultures mondiales les plus importantes en valeurs cumulées de surfaces cultivées comme de volumes de production (**Tableau 1**).

Production en milliers de tonnes en 2016							
	Monde	Afrique	Amérique	Asie	Europe	Océanie	1 ^{ers} producteurs
Pêches / Nectarines	24976	860	2159	17499	4373	84	Chine - 14469
Prunes / Prunelles	12051	399	963	8033	2636	20	Chine - 6676
Abricots	3881	514	102	2409	844	12	Turquie - 730
Cerises / griottes	3696	21	585	1502	1568	20	Turquie - 792
Amandes	3214	278	2035	510	318	73	USA - 2003
Surfaces cultivées en milliers d'hectares en 2016							
	Monde	Afrique	Amérique	Asie	Europe	Océanie	1 ^{ères} surfaces cultivées
Prunes / Prunelles	2651	55	82	2120	391	3	Chine - 1961
Amandes	1866	454	389	350	652	21	Espagne - 545
Pêches / Nectarines	1640	77	159	1133	259	12	Chine - 839
Cerises / griottes	654	7	86	243	315	3	Turquie - 107
Abricots	568	72	8	380	99	9	Turquie - 124

Tableau 1 - Productions et surfaces fruitières cultivées mondiales des principales espèces de *Prunus*. Données FAO – 2016.

Les espèces du genre *Prunus* sont des Angiospermes Dicotylédones appartenant au clade des Rosidae, à l'ordre des Rosales et à la famille des Rosaceae (**Figure 16**). Au sein des Rosaceae, le genre *Prunus* n'est pas définitivement positionné mais côtoie de nombreux genres d'intérêt agronomique comme le pommier (*Malus*), le poirier (*Pyrus*), ou le fraisier (*Fragaria*) (Potter *et al.*, 2007) (**Figure 16**). Différentes études basées sur l'ADN plastidial et nucléaire ribosomique (ITS) montrent que, contrairement aux Rosaceae, le genre *Prunus* est un groupe monophylétique (Chin *et al.*, 2014). Son

nombre de chromosomes de base est égal à 8 ($2n=2x=16$) mais la ploïdie de certaines espèces peut varier jusqu'à $2n=22x=176$ chez *P. laurocerasus* (Watkins, 1976).

La classification établie par (Rehder, 1940), admise par de nombreuses études (revue par (Chin *et al.*, 2014)), divise le genre *Prunus* en cinq sous-genres :

- † Les *Amygdalus* qui comprennent notamment des espèces diploïdes comme le pêcher (*Prunus persica*) et l'amandier (*P. dulcis*) et les espèces sauvages (*P. davidiana* et *P. mira*)
- † Les *Prunophora* ou *Prunus sensu stricto* regroupant les pruniers hexaploïdes (*P. domestica*, *P. insititia*), tétraploïdes (*P. spinosa*), ou diploïdes (*P. cerasifera*, *P. salicina*) mais aussi les abricotiers (*P. armeniaca*, *P. mume*, *P. brigantina*)
- † Les *Cerasus* avec les merisiers (*P. avium*) et le cerisier acide (*P. cerasus*)
- † Les *Laurocerasus* et (v) les *Padus*.

Une liste non exhaustive des différentes espèces de *Prunus* est présentée en **Figure 16**.

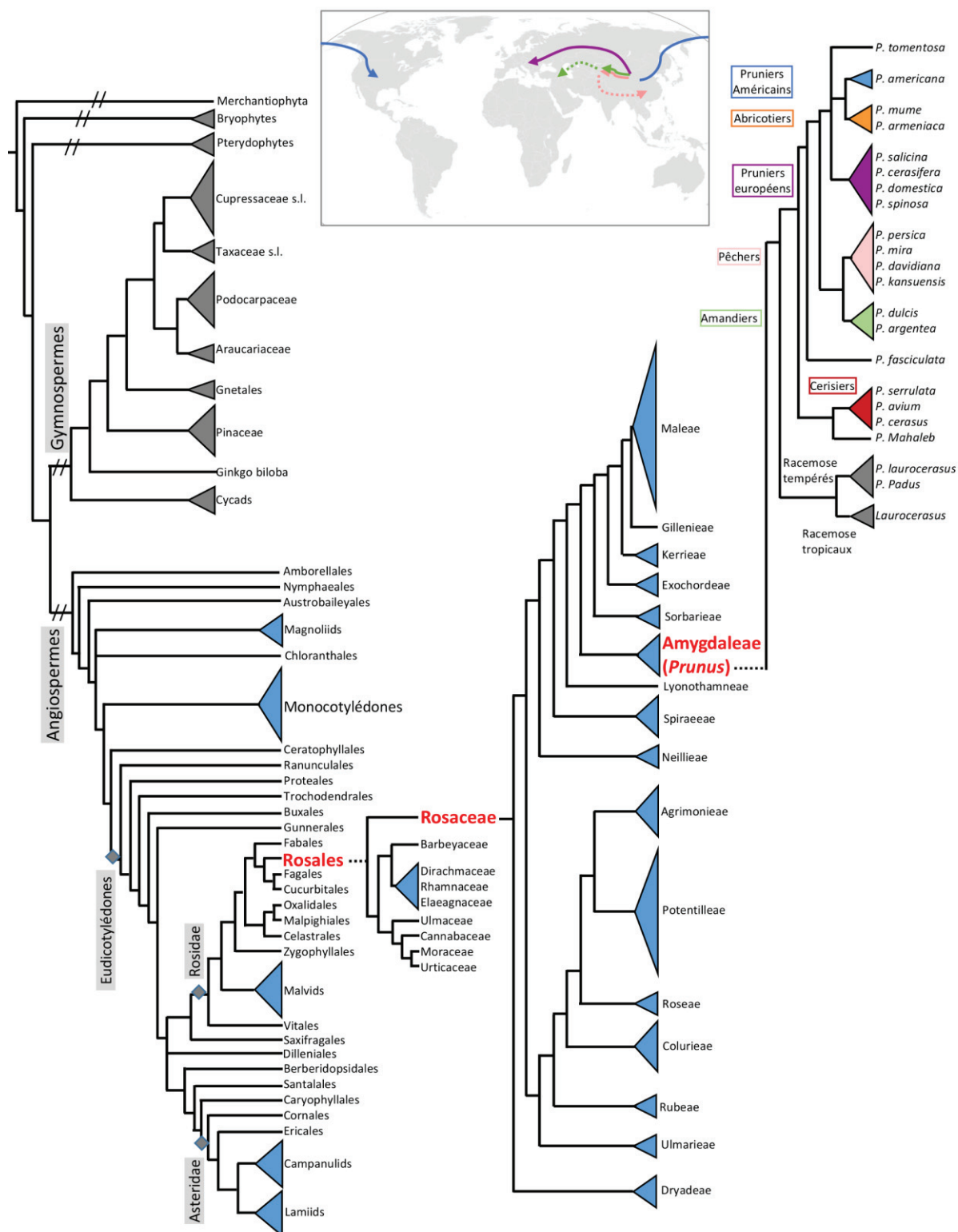


Figure 16 - Position des *Prunus* dans la classification simplifiée des plantes terrestres et leurs routes évolutives.

Classification combinée simplifiée des plantes terrestres basée sur (Lu *et al.*, 2014) pour les Gymnospermes, (APG IV *et al.*, 2016) pour les Angiospermes, (Potter *et al.*, 2007) pour les Rosaceae et (Chin *et al.*, 2014) pour les *Prunus*. L'encadré illustre les routes putatives de migration suivies par différentes lignées de *Prunus* à partir de l'aire d'origine hypothétique (Asie orientale). Les lignées de pruniers eurasiens et américains sont représentées respectivement en mauve et bleu, les amandiers en vert et les pêchers en rose – basé sur (Chin *et al.*, 2014).

1.4.2 Aire d'origine

Les Rosaceae ont probablement évolué à partir du continent nord-américain il y a 90 millions d'années (Ma) (pendant la période du Crétacé) (Crepet *et al.*, 2004; Potter *et al.*, 2007). Des proto-*Prunus* ont vraisemblablement été dispersés depuis l'actuelle région de Colombie britannique vers l'Asie orientale via la bande terrestre de Béring. Le genre *Prunus* actuel serait apparu il y a environ 61 Ma (tôt dans la période Eocène) en Asie orientale. L'Eurasie héberge les plus vieux fossiles d'endocarpes de *Prunus* (55 Ma - *P. wutuensis*) et le maximum de diversité observée dans ce genre, ce qui est en accord avec leur aire putative d'origine (Chin *et al.*, 2014) (**Figure 16**). Des événements de dispersion et de spéciation expliqueraient la distribution actuelle du genre à partir de cette aire. Le sous-genre *Amygdalus* (pêchers et amandiers) s'est disséminé de l'Asie orientale vers l'Asie occidentale il y a 49 Ma, précédant une diversification et une évolution séparée par le relief himalayen en Asie occidentale (Moyen-Orient) pour les amandiers et en Asie orientale pour les pêchers (Watkins, 1976). Les pruniers européens (*P. spinosa*, *P. cerasifera* et *P. domestica*) ont pour leur part migré à partir de l'Asie orientale plus récemment (31 Ma) (Chin *et al.*, 2014) (**Figure 16**). Aujourd'hui l'aire de répartition du prunier myrobolan (*P. cerasifera*) s'étend de l'Asie centrale au sud de l'Europe, avec un pic de concentration et de diversité autour de la Mer Caspienne (Crimée, Krymsk, Géorgie, Arménie, Iran) (Eremin, 1978).

1.4.3 Intra et inter-compatibilités

La production des fruits à noyau est basée sur l'emploi de variétés présentant les caractéristiques nutritives/gustatives sélectionnées, greffées sur des porte-greffes leur fournissant vigueur et adaptation aux conditions telluriques. Les porte-greffes font l'objet d'une sélection et les améliorateurs peuvent s'appuyer sur les capacités d'hybridation et de compatibilité au greffage interspécifiques que possèdent la plupart des *Prunus*. Par ailleurs, l'hybridation interspécifique se produit également à l'état naturel notamment entre pêchers et amandiers (*Amygdalus*) ou entre espèces différentes de pruniers (*Prunophora*).

Certains facteurs freinent les possibilités offertes aux sélectionneurs et améliorateurs. Ainsi les amandiers, cerisiers et certains pruniers sont autostériles par auto-incompatibilité gamétophytique ou parfois par stérilité mâle. Les croisements intraspécifiques et le fort niveau d'hétérozygotie permettent toutefois d'exploiter une variabilité importante. Ceci est particulièrement vrai pour les pruniers myrobolans d'un point de vue adaptatif et morphologique. A l'inverse, les pêchers et certaines variétés de pruniers et d'abricotiers sont des autogames préférentiels, ce qui induit une diversité génétique moindre. Les programmes de sélection des porte-greffes s'intéressent aux capacités de vigueur, de résilience face aux conditions telluriques abiotiques (sécheresse, asphyxie ou calcaire) et biotiques (maladies et ravageurs du sol). Vis-à-vis de ces derniers facteurs, les nématodes à galles sont un enjeu majeur. Dans la perspective d'obtenir des porte-greffes polyvalents, les capacités de certains *Prunus*

comme le prunier myrobolan à se croiser facilement avec des pêchers, amandiers ou abricotiers permettent d'envisager de cumuler un maximum de caractères favorables et d'assurer notamment une aptitude plus diversifiée au greffage (Dosba *et al.*, 1994).

1.4.4 Données génomiques

Chez les Rosaceae d'intérêt agronomique et ornemental, les génomes de diverses espèces de *Fragaria*, *Malus*, *Pyrus*, *Rosa* et *Prunus*, ont été obtenus et sont accessibles sur la plateforme 'Genome Database for Rosaceae' (GDR ; <https://www.rosaceae.org/>) (Jung *et al.*, 2019). Chez les *Prunus*, le génome du pêcher (*P. persica* cv. 'Lovell') fut le premier à être séquencé (Verde *et al.*, 2013). De par sa taille relativement modeste (265 Mb), les nombreux transcriptomes disponibles et les efforts de re-séquençage profond (version 2.0.a.1 (Verde *et al.*, 2013)), il est le génome de référence pour les Rosaceae et dispose d'un assemblage de très bonne qualité. Par ailleurs le pêcher, étant autogame préférentiel et faisant l'objet d'une sélection importante, est fortement homozygote. Contrairement à d'autres Rosaceae des genres *Malus* ou *Pyrus*, le pêcher n'a pas subi de récente duplication de génome, ce qui en fait une pierre angulaire pour les études évolutives (Jia *et al.*, 2015). Plus récemment d'autres génomes du genre ont été assemblés : l'abricotier du japon (*P. mume* (Zhang *et al.*, 2012)), le merisier (*P. avium* (Shirasawa *et al.*, 2017)), un cerisier ornemental (*P. yedoensis* (Baek *et al.*, 2018)), l'amandier (*P. dulcis* (Pere Arús & Tyler Alioto, com. pers., 2019)) et le prunier domestique (*P. domestica* (Zhebentyayeva *et al.*, 2019)). La diversité de ces ressources et la synténie existant chez les *Prunus* ont permis une progression rapide vers la découverte et la caractérisation des sources de résistance exploitable pour lutter contre les nématodes à galles.

1.5 *Prunus* – *Meloidogyne* : modèle pour l'interaction plante pérenne ligneuse-nématode à galles

1.5.1 Réaction compatible

Dans la cadre de la réaction compatible entre le nématode à galles et sa plante-hôte, la larve J2 induit au niveau du cylindre central l'apparition de cellules géantes. Ces cellules géantes (entre 5 et 7) de ploïdie pouvant être supérieure à 64 (Huang & Maggenti, 1969) présentent un cytoplasme extrêmement riche en organites, synonyme de métabolisme hyperactif. D'un point de vue macroscopique ces symptômes se traduisent par l'apparition de galles sur les racines. Les *Prunus* infectés par les nématodes à galles présentent des caractéristiques similaires, tant au niveau cellulaire qu'au niveau des symptômes

externes, à celles décrites chez la tomate qui est un excellent hôte (Khallouk *et al.*, 2011). En effet, chez les *Prunus*, les galles peuvent aussi atteindre des tailles étonnamment importantes malgré la lignification des tissus racinaires. Ces déformations provoquent une perturbation de la circulation de la sève brute et élaborée ainsi que de l'absorption de l'eau et des nutriments. Les conséquences en sont essentiellement un retard de croissance (diamètre du collet, hauteur des arbres, poids de la partie aérienne) (Duval *et al.*, 2019), une moindre résistance à la sécheresse, un jaunissement et une perte des feuilles et une baisse de production de fruits (**Figure 17**).

1.5.2 La lutte contre les nématodes à galles en verger

Les traits de vie des nématodes à galles (endoparasitisme, polyphagie, multiplication exponentielle, ...) en font des ravageurs extrêmement difficiles à contrôler. Les nématicides chimiques sont les molécules les plus toxiques utilisées en agriculture et la plupart d'entre elles ont été retirées du marché en Europe. La recherche s'est orientée vers des méthodes alternatives telles que les rotations culturales, l'utilisation de plantes à effet antagoniste ou la lutte biologique (organismes prédateurs). Ces méthodes, qui ont montré des résultats très contrastés chez les espèces annuelles et/ou maraîchères (Djian-caporalino *et al.*, 2019), paraissent peu adaptées au contexte de l'arboriculture fruitière. De ce fait chez les *Prunus* comme chez d'autres espèces pérennes, la résistance (R) variétale issue de sources naturelles reste la meilleure alternative. Les facteurs génétiques de résistance (facteurs R) étant relativement rares chez les plantes, leur utilisation doit être raisonnée (Djian-Caporalino *et al.*, 2014) afin d'en optimiser la durabilité. Vis-à-vis des *Meloidogyne*, des cas de contournement du gène de résistance *Mil-2* de la tomate ont été observés très tôt (Prot, 1984; Castagnone-Sereno *et al.*, 1994; Kaloshian *et al.*, 1996)). Les arbres présents en verger pendant de nombreuses années sont particulièrement exposés au risque d'apparition de lignées virulentes de ces nématodes (Saucet *et al.*, 2016). L'étude approfondie de ces facteurs R et des possibilités permettant d'améliorer leur durabilité, tel le cumul de plusieurs sources de résistance dans un même clone (pyramidage), revêt donc un intérêt crucial dans une approche de gestion de ces ravageurs à long terme.

1.5.3 Réaction incompatible *Prunus* - *Meloidogyne*

La notion de réaction incompatible entre le parasite et son hôte renvoie à la présence d'un gène de résistance chez l'hôte et à celle d'un gène d'avirulence correspondant chez le pathogène. Ainsi, certains clones de *Prunus* possèdent des facteurs génétiques ou gènes *R* induisant une résistance qui empêche le développement des *Meloidogyne*. En présence du gène *Ma* du prunier myrobolan *P. cerasifera* P. 2175, les larves mobiles J2 pénètrent dans la racine selon le même itinéraire de migration que celui décrit dans l'interaction compatible entre le nématode et un prunier myrobolan sensible, comme le clone *P. cerasifera* P. 2032, mais échouent dans l'établissement du site nourricier, ce qui entraîne leur mort par

incapacité de se nourrir normalement (Voisin *et al.*, 1999). En réponse la plante induit une réaction de type hypersensible (HR-like) qui se traduit par la mort cellulaire des cellules entourant la partie antérieure du nématode (Khallouk *et al.*, 2011). Les polyphénols sécrétés (**Figure 17**) suggèrent la présence de lignine, de callose ou de protéines de dégradation et de défense. Une étude comparée de la dynamique de pénétration et développement de *M. arenaria* sur P.2175/P.2032 précise cela en montrant que, jusqu'à 2 jours post-inoculation, il n'y a aucune différence de comportement du nématode entre pruniers sensible et résistant. En revanche, après 4 jours, les plants résistants sont indemnes de larves fixées J3 et ont un effectif de J2 endophytes en diminution (Voisin *et al.*, 1999). Le plant résistant, indemne de galles mais subissant des nécroses d'hypersensibilité, réagit également en induisant l'apparition de nouveaux apex racinaires latéraux en amont des zones de migration du nématode (Khallouk *et al.*, 2011). C'est à partir de cette période entre 2 et 4 jours post-inoculation que le processus de résistance s'exprime. Chez les *Prunus*, la résistance aux *Meloidogyne* est conférée par quelques gènes majeurs caractérisés par des spectres d'action différents.

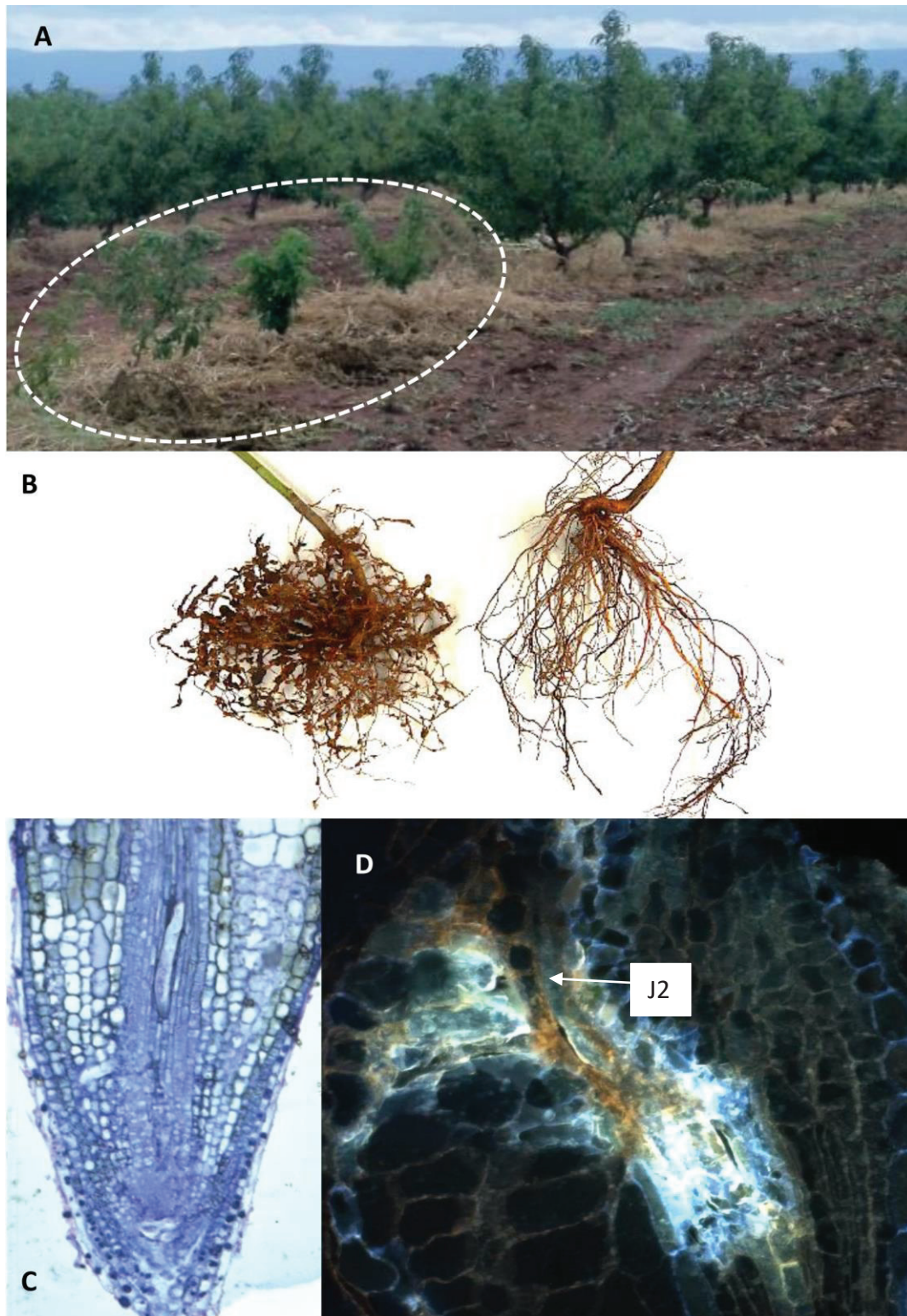


Figure 17 - L'interaction *Meloidogyne-Prunus*.

(A) Retard de croissance de plants de *Prunus* sensibles dans une parcelle au Maroc contenant un mélange des espèces *M. incognita*, *M. javanica* et *M. arenaria* (D. Esmenjaud, com. pers.). (B) Illustration des réactions compatible et incompatible induites respectivement sur *Prunus* sensible (S, gène *Ma* absent, à gauche) et résistant (R, gène *Ma* présent, à droite). (C) Coupe racinaire colorée au bleu de toluidine réalisée sur apex racinaire de *Prunus* résistant *Ma*, 3 jours après inoculation avec *M. incognita*. On observe notamment une larve J2 bloquée au stade filiforme au niveau du cylindre central. (D) Coupe racinaire non colorée observée sous UV réalisée sur racine résistante *Ma* après inoculation avec *M. incognita*. La réponse locale HR-like de la plante est visible en brun-orangé autour de la larve J2. (Photos C et D extraites de (Khallouk *et al.*, 2011)).

1.5.4 Les gènes de résistance aux *Meloidogyne* chez les *Prunus*

Les premières sources de résistance aux nématodes à galles ont été identifiées dans des cultivars asiatiques de pêchers (*P. persica*) dans les années 60 (Sharpe *et al.*, 1969). Plus tard, cette résistance sera attribuée à la présence d'un gène dominant, *RMia*, conférant une résistance totale aux espèces *M. incognita* et *M. arenaria* (Claverie *et al.*, 2004a; Esmenjaud *et al.*, 2009). Ce gène n'a pas d'effet ou un effet très limité (population dépendant) sur *M. javanica* (Nyczepir & Esmenjaud, 2007; Duval *et al.*, 2019). Il est localisé sur le groupe de liaison 2 (LG2) et appartient très probablement à la sous-famille des TNLs (Duval *et al.*, 2014). Un pêcher chinois (*P. kansuensis*) abrite également sur le LG2, un gène de résistance à *M. incognita* nommé *PkMi* (Cao *et al.*, 2014). Cependant son action vis-à-vis des autres *Meloidogyne* ainsi que sa relation avec le gène *RMia* sont inconnues.

Au sein du sous-genre *Amygdalus*, les accessions 'Alnem' d'amandier (*P. dulcis*) portent le gène majeur dominant *RMja* conférant une résistance totale à *M. javanica* (Kochba & Spiegel-Roy, 1975). Contrairement aux gènes de résistance des pêchers précédemment évoqués, le gène *RMja*, localisé sur le LG7, ne confère aucune résistance à *M. incognita*. Les gènes de résistance des sources pêchers et amandiers ne sont donc pas liées (Van Ghelder *et al.*, 2010).

Plusieurs résistances ont également été identifiées chez les pruniers. Détecté chez le prunier myrobolan (*P. cerasifera*), le gène *Ma* a fait l'objet de nombreux travaux ces dernières années car il est le plus prometteur. Il confère une résistance totale à plus de 30 espèces et isolats de *Meloidogyne* (Esmenjaud *et al.*, 1994, 1997; Duval *et al.*, 2019). Localisé sur le LG7, *Ma* a été le deuxième gène de résistance des plantes aux *Meloidogyne* cloné (Claverie *et al.*, 2011) après le gène *Mi-1.2* de la tomate (Milligan *et al.*, 1998). Il appartient à la sous-famille des TNLs mais avec une structure originale qui sera analysée de façon plus approfondie. A ce jour aucune espèce de *Meloidogyne* n'a contourné l'action du gène *Ma*. Un autre gène, *Rjap*, identifié chez le prunier japonais (*P. salicina*) présente également un spectre de résistance large et une localisation sur le LG7 (Claverie *et al.*, 2004b). Même si aucun marquage fin n'a été réalisé sur ce dernier, il est très probable qu'il soit l'orthologue du gène *Ma*. Enfin un prunier myrobolan sauvage chinois (*P. sogdiana*) porte le gène *PsoRPM2* induisant une résistance à *M. incognita* (Zhu *et al.*, 2017). Les différents spectres de résistance et les caractéristiques des gènes majeurs identifiés chez les *Prunus* vis-à-vis des espèces de *Meloidogyne* sont résumés sur la **Figure 18**.

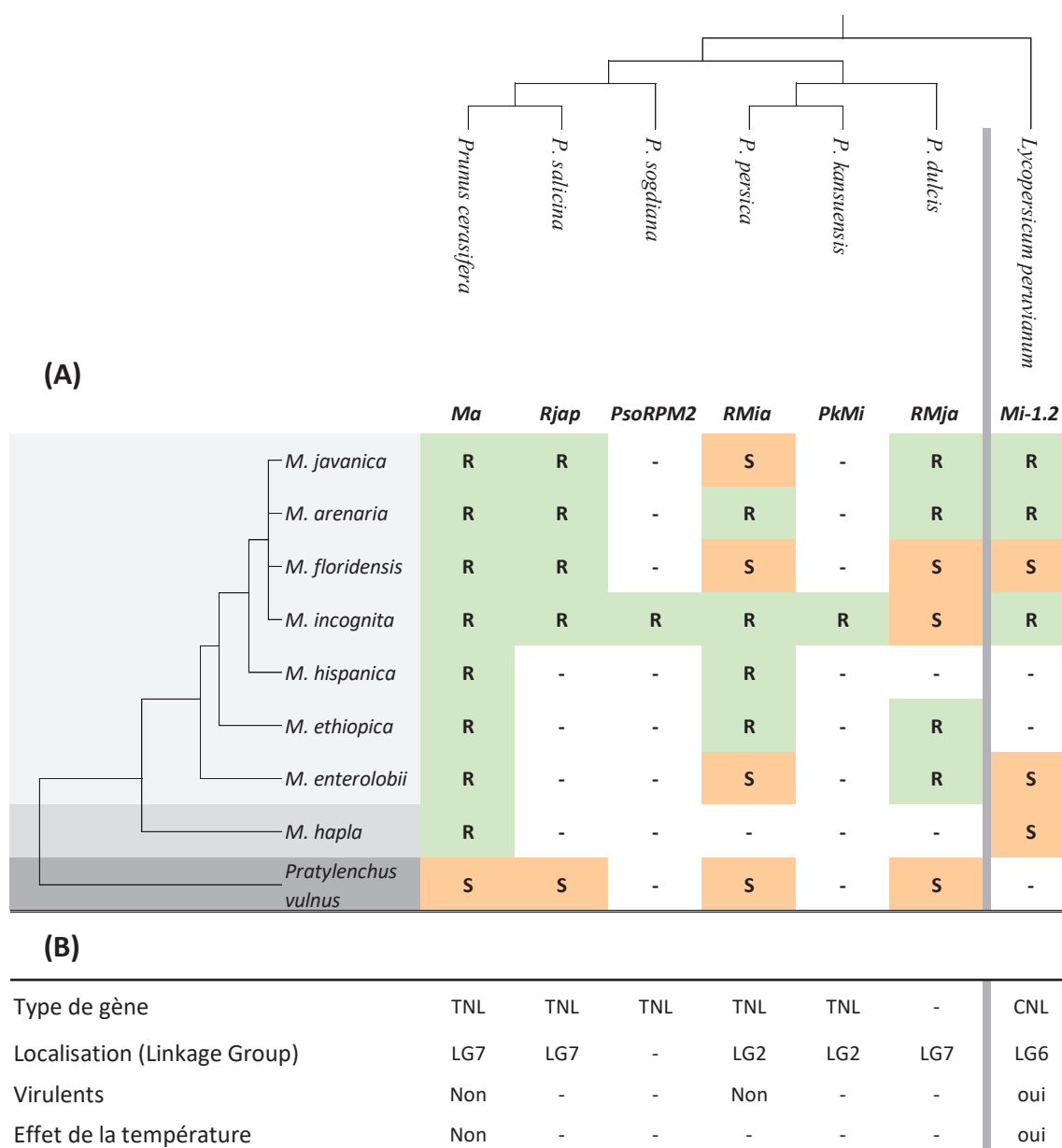


Figure 18 - Spectres (A) et caractéristiques (B) de la résistance conférée par les gènes identifiés vis-à-vis des *Meloidogyne* chez les *Prunus*.

'R' Résistant, 'S' Sensible, '-' inconnu. Le gène *Mi-1.2* de la tomate est utilisé comme gène de référence externe. *Pratylenchus vulnus* est utilisé comme représentant d'un genre externe aux *Meloidogyne*. Les références bibliographiques sont données dans le texte.

1.5.5 Le gène *Ma* : un gène particulier

1.5.5.1 Propriétés biologiques

Le gène *Ma* confère une résistance totale vis-à-vis des espèces de nématodes à galles testées. Il est actuellement le gène qui a le spectre le plus large vis-à-vis des *Meloidogyne*, notamment par rapport au gène de référence *Mi-1.2* (ou *Mi*) de la tomate (**Figure 18A**) (Milligan *et al.*, 1998). Le gène *Ma* est à dominance complète (aucune différence entre homozygotes et hétérozygotes) et son activité n'est pas affectée par les hautes températures (Esmenjaud *et al.*, 1997), contrairement au gène *Mi-1.2* (Dropkin, 1969). A la différence de ce dernier également (Castagnone-Sereno *et al.*, 1994), *Ma* n'est à ce jour contourné par aucune espèce de nématodes à galles et sa durabilité a notamment été montrée vis-à-vis de *M. incognita* sur une durée de 4 ans (Khallouk *et al.*, 2013) (**Figure 18B**). Il agit très probablement lors d'une étape critique de la relation plante-nématode à galles, comme l'instauration du site nourricier, et permet le déclenchement d'une réponse forte et adaptée.

1.5.5.2 Caractéristiques structurales

Le gène *Ma* est un TNL constitué de 9 exons séparés par des introns de taille variable (**Figure 8 et Figure 19**). Les quatre premiers exons codent respectivement pour les domaines protéiques TIR, NBS, NLL, LRR, l'exon dit « NLL » (**N**BS-**L**RR **L**inker) codant pour une séquence faisant le lien entre les structures très différentes des domaines NBS et LRR. Les cinq derniers exons codent pour cinq modules, séparés par des introns contenant des microsatellites (Claverie *et al.*, 2011). Les similarités relevées entre ces cinq modules indiquent qu'il s'agit de cinq exons répétés codant pour cinq modules d'environ 200-250 acides aminés et présentant une homologie avec les signatures décrites en position C-terminale chez les TNLs d'*Arabidopsis* (Meyers *et al.*, 2003). Cette particularité, au moins pour les TNLs clonés, fait du gène *Ma* un TNL deux fois plus grand que les TNLs dits classiques. Le clonage du gène avec sa région promotrice (15 kb) a permis de valider ce gène par transformation d'un clone sensible. Il produit un transcrit de 6147 pb (sans les UTRs) et une protéine prédite de 2048 aa (**Figure 19**) (Claverie *et al.*, 2011).

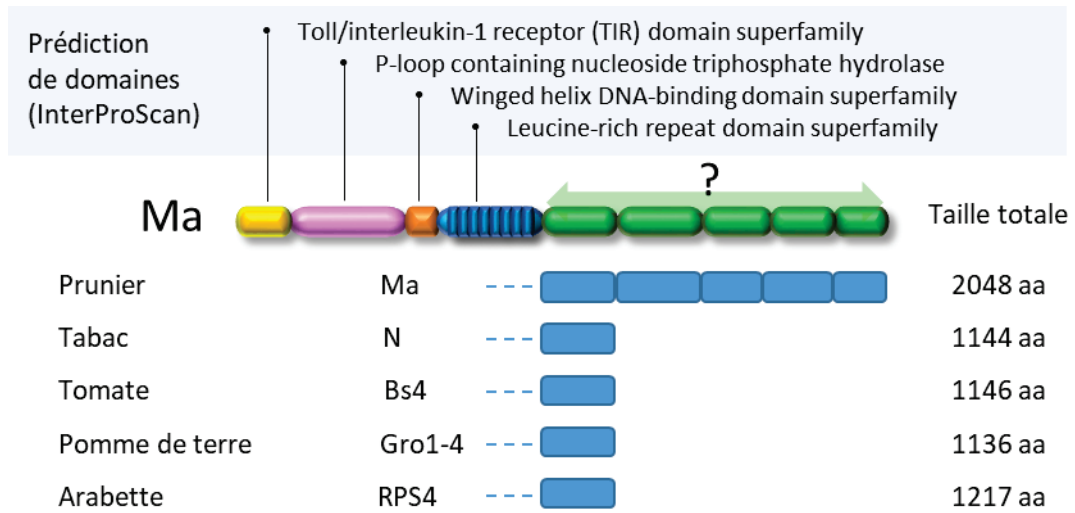


Figure 19 - Annotation fonctionnelle de la protéine prédite Ma et de quelques TNLs de référence.

1.5.6 Ouverture

L'étude du gène *Ma* chez le prunier myrobolan a révélé des propriétés biologiques et des caractéristiques remarquables soulevant ainsi de nombreuses questions. Le gène *Ma* est-il vraiment atypique ? Cette structure est-elle présente chez d'autres *Prunus* et plus largement chez d'autres groupes taxonomiques ? Pourrait-elle être liée à la propriété de résistance aux nématodes à galles et, dans ce cas, quels seraient les déterminants impliqués dans la résistance ?

Plus largement, la nature et les caractéristiques du module nommé domaine Post-LRR (PL) de par sa localisation, restent inconnues. Ce module est-il un composant indispensable au bon fonctionnement des TNLs ? A-t-il été conservé au cours de l'évolution des plantes et quel pourrait être son rôle ? Toutes ces questions appellent des réponses qui permettraient de mieux comprendre le fonctionnement des TNLs dans le processus de résistance.

Afin de fournir des éléments de réponse aux questions soulevées ci-dessus, ce travail de thèse développe les quatre parties suivantes :

- † **Les TNLs dans le génome du pêcher et la caractérisation du domaine PL.** En prenant comme modèle d'étude le génome du pêcher, nous annoterons, au sein de la famille des NBS-LRRs, les récepteurs immuns de la sous-famille des TNLs et caractériserons structurellement le domaine PL. Nous préciserons également la distribution du domaine PL à plus grande échelle en exploitant un large éventail de données génomiques. Cette étape nous permettra de savoir si la particularité structurelle du gène *Ma* liée au domaine PL est originale (i) chez les *Prunus* et (ii) plus largement chez d'autres plantes (Angiospermes). D'un point de vue évolutif, nous chercherons à savoir à quel moment est apparue et comment s'est développée cette structure et proposerons une hypothèse sur son origine.
- † **Les NBS-LRRs chez les conifères et les traits évolutifs de cette famille chez les plantes terrestres.** Afin de valoriser l'expertise acquise dans l'annotation des TNLs et PLs associés aux espèces pérennes du genre *Prunus*, nous élargirons nos travaux aux NBS-LRRs présents chez d'autres plantes pérennes en abordant notamment l'étude de cette famille dans le groupe taxonomique des conifères, représentants majeurs des Gymnospermes. Nous étudierons la distribution des deux sous-familles majeures détectées chez les conifères, à savoir les TNLs et les RNLs, chez les Gymnospermes mais également chez les autres plantes terrestres.
- † **Caractérisation du gène *RMja* chez l'amandier.** Du fait de leur localisation sur le même chromosome, nous préciserons tout d'abord la relation génétique entre le gène *Ma* du prunier et

le gène *RMja* de l'amandier. Cette première étape comprend l'identification de la séquence complète de la région impliquée dans la résistance aux nématodes à galles chez l'amandier. Elle nous permettra de développer une étude comparative entre ces deux gènes dont le spectre de résistance est sensiblement différent. D'un point de vue agronomique, la relation génétique entre ces deux gènes a une influence sur une stratégie impliquant leur pyramidage dans un même porte-greffe afin de conférer une résistance durable aux nématodes à galles.

- † **Données complémentaires relatives aux mécanismes et déterminants de la résistance conférée par le gène *Ma*.** Dans cette dernière partie, nous présenterons nos premiers résultats sur les données d'expression du gène *Ma* et sur les données de diversité de ses allèles et orthologues. Ces éléments contribueront, d'une part, à une meilleure compréhension du mode de fonctionnement de *Ma* vis-à-vis des nématodes à galles et, d'autre part, à la localisation des régions impliquées dans la résistance et *in fine* dans sa spécificité. Ils sont aussi susceptibles d'apporter des pistes de réflexion sur les partenaires potentiels impliqués dans la réponse déclenchée par les TNLs et sur le rôle du domaine PL dans cette réponse.

CHAPITRE 2

MATERIEL ET METHODES

Chapitre 2 : MATERIEL ET METHODES

2.1 Matériel végétal

Le matériel végétal de cette étude recouvre essentiellement des variétés de diverses espèces de *Prunus* dont certaines sont utilisées comme porte-greffes. Il comprend notamment diverses accessions de prunier myrobolan et prunier japonais conservées dans la collection de l'INRA GAFL d'Avignon et qui avaient été phénotypées vis-à-vis des *Meloidogyne* puis conservées à l'UMR ISA (J. L. Poëssel et D. Esmenjaud, com. pers.) (Claverie *et al.*, 2004a,b). La variété *P. persica* cv. 'Lovell' utilisée pour le séquençage du génome, sera considérée comme témoin sensible pêcher. Les accessions de pruniers myrobolans (*P. cerasifera*), 'P. 2175', 'P. 1079' et 'P. 2032' (= variété 'Myrabi'), seront utilisées respectivement comme porteur hétérozygote du gène *Ma*, porteur homozygote du gène *Ma* et comme clone sensible dépourvu du gène *Ma*. Le prunier japonais *P. salicina* cv. 'J222' sera utilisé comme porteur du gène *Rjap*. Enfin, les amandiers *P. dulcis* cv 'Alnem 1' (= 'R1107') et 'Lauranne' ont été utilisés respectivement comme porteur homozygote du gène *RMja* et comme clone sensible hétérozygote dépourvu du gène *RMja*. Pour le marquage fin du gène *RMja*, des individus hétérozygotes F1 (*RMja/rmja*₁ et *RMja/rmja*₂) issus du croisement [Alnem 1 x Lauranne] ont été autofécondés en serre et les amandes F2 ainsi générées ont été collectés et placés sur milieu de culture pour germination. Environ 1500 individus ont été obtenus pour être phénotypés et génotypés. L'ensemble de ces clones sont cultivés en serre en condition hors-gel et les clones *P. cerasifera* P. 2175, P. 2032 et *P. dulcis* Alnem 1 sont également cultivés *in vitro*.

L'étude concernant les répertoires de NBS-LRRs exprimés chez les conifères a été réalisée sur différents tissus issus de sept espèces nord-américaines : l'épinette blanche (*Picea glauca*), l'épinette noire (*Picea mariana*), le pin blanc (*Pinus strobus*), le pin gris (*Pinus banksiana*), le sapin baumier (*Abies balsamea*), le mélèze laricin (*Larix laricina*) et le thuya occidental (*Thuja occidentalis*). Les plants cultivés en serre âgés de 4 ans ont subi un traitement de stress du type sécheresse. Durant ce traitement de 14 jours, la moitié des plants a été arrosée normalement et l'autre moitié a été privée d'arrosage.

2.2 Matériel animal

Le phénotypage des individus du croisement F2 d'amandier en ségrégation pour le gène *RMja* a été effectué en utilisant un isolat de l'espèce *M. javanica*. Cet isolat a été obtenu par élevage d'une masse d'œufs extraite de la population *M. javanica* "Higuera" initialement prélevée à Cabrils (Catalogne, Espagne). Cet isolat a été maintenu en collection sur la tomate *Lycopersicon esculentum* cv. 'St. Pierre'. L'identité de l'espèce de nématode à galles a été vérifiée par électrophorèse estérasiqse selon la méthode décrite par (Janati *et al.*, 1982).

2.3 Extractions d'acides nucléiques

Les protocoles d'extraction d'ADN et d'ARN sur le matériel végétal de type *Prunus* ont été optimisés pour s'affranchir des contaminants fortement présents chez ces modèles et diminuant la qualité et la quantité des acides nucléiques obtenus. Les tissus foliaires sont préférentiellement prélevés le matin afin d'éviter l'accumulation de sucres présents dans les tissus qui affectent la qualité des extractions. Environ 100 - 130 mg de jeunes feuilles (d'aspect encore légèrement translucide) ou d'apex racinaires blancs (le moins lignifiés possible) constituent un échantillon optimisé contenant un minimum de contaminants (lignine, polyphénols, ...). Chaque échantillon prélevé est congelé dans l'azote liquide, et soit immédiatement utilisé pour en extraire l'ADN ou l'ARN, soit stocké à -80°C.

2.3.1 Extraction d'ADN de Prunus

Les échantillons sont broyés à l'aide d'un broyeur à bille (2 minutes – 30 Hz) dans 400 µl de tampon d'extraction (fraîchement préparé ou en rajoutant le métabisulfite au dernier moment) (Sorbitol 0.35M ; Tris 0.1M ; EDTA 5mM ; métabisulfite de sodium 4g/L ; pH 7.5) par tube. Une fois broyés, 400 µl de tampon de lyse (Tris 0.2M ; EDTA 0.05M ; NaCl 2M ; CTAB 2%) et 160 µl de Sarkosyl (5%) ont été ajoutés à chaque tube et les tubes ont été mélangés par inversion puis au vortex. Les échantillons sont incubés 20 à 30 min à 65°C sous agitation. 660 µl de chloroforme/isoamyl alcool (24 :1) froid (4°C) sont ajoutés par échantillon et agités jusqu'à l'obtention d'une émulsion. Les échantillons sont ensuite centrifugés 5 min à 15 000 rpm à 4°C. Le surnageant est homogénéisé avec 2 µl de RNase A (10 mg/ml) et incubé 30 min à 37°C. 480 µl d'isopropanol froid sont ajoutés et les échantillons sont laissés durant la nuit à -20°C. Les échantillons sont centrifugés à 4°C pendant 10 min à 15 000 rpm. Les culots sont lavés au moins deux fois à l'éthanol 70% froid (4°C) préparé avec de l'eau ultra-pure. Les culots sont décollés par tapotement pour les rincer. Après séchage, ils sont repris dans 100 µl d'eau ultra-pure stérile chauffée à 37°C.

2.3.2 *Extraction d'ARN totaux de Prunus*

100 à 150 mg de tissus (racines ou feuilles) sont broyés dans l'azote liquide puis repris dans 800 µl de tampon d'extraction (CTAB 2.5% ;PVPP 2% ; Tris-HCl pH8 100mM ;EDTA 25mM ; NaCl 2M ; β-mercaptoethanol 3%) et incubés à 65°C pendant 30 minutes. 800 µl de chloroforme/isoamyl alcool (24 :1) sont ajoutés et centrifugés (8 min – 13000 rpm). Le surnageant est repris avec le même volume de phénol saturé en tris HCl (pH acide)/chloroforme/isoamyl alcool (25 :24 :1) et laissé sur glace 10 minutes puis les échantillons sont centrifugés (8 min – 13000 rpm). Le surnageant est repris avec le même volume de chloroforme/isoamyl alcool (24 :1) et centrifugé (8 min – 13000 rpm). Le surnageant est repris avec le même volume de NaCl 5M et 1 volume d'isopropanol pour être laissé la nuit à -20°C. Les échantillons sont centrifugés 20 minutes à 13000 rpm puis les culots sont lavés deux fois à l'éthanol 70%. Les culots sont séchés puis repris dans 44 µl d'eau ultra-pure stérile RNase free.

Les 44 µl d'ARN sont ajoutés à 5 µl de tampon 10X TURBO DNase Buffer et 1 µl de TURBO DNase puis incubés à 37°C pendant 30 minutes (kit TURBO DNA-free). 5 µl de tampon d'inactivation de la DNase sont rajoutés et incubés pendant 5 minutes à température ambiante. Après centrifugation les ARNs purifiés sont dosés au nanodrop et déposés sur gels. Seuls les ARNs purs sans contamination d'ADN génomique et présentant des ratios A 260/280 et A 260/230 compris entre 1.9 et 2.2 seront conservés pour la synthèse des ADN complémentaires.

2.3.3 *Synthèse des ADN complémentaires (ADNc) et clonage pour séquençage*

La synthèse des ADNc est effectuée avec le kit 'Maxima H Minus First', en ajoutant 1 µl d'amorce spécifique ou d'oligo dT', 1 µl de 'dNTP mix' et 13 µl d'ARN total DNA free. 4 µl de tampon '5X RT Buffer' et 1 µl d'enzyme 'Maxima H Minus' sont ensuite rajoutés pour un volume total de 20 µl ; le tout est ensuite mis à incuber 30 minutes à 50°C puis 5 minutes à 85°C pour inactiver l'enzyme. Les ADNc sont stockés au congélateur -80°C pour une utilisation en PCR ou clonage.

Afin d'obtenir les séquences de fragments amplifiés par PCR avec les amorces SCAR à partir d'ADN génomique ou d'ADNc, les produits PCR sont soit directement séquencés en Sanger soit clonés avant séquençage en Sanger. Dans ce dernier cas les produits PCR sont analysés sur gel d'agarose (0,8%-2,5% TAE) et extraits en utilisant le kit 'Qiagen QIAquick Gel Extraction Kit' en suivant les instructions du fournisseur sans modifications. Les bandes purifiées sont clonées en vecteur plasmidique avec le kit 'pGEM-T vector systems' en suivant le protocole du fournisseur sans modification. Les fragments fraîchement clonés sont utilisés pour transformer des bactéries thermo-compétentes DH5α en utilisant les consignes et les réactifs du kit 'Invitrogen™ Library Efficiency™ DH5α Competent Cells'. Les colonies bactériennes blanches récupérées après une nuit à 37°C sont analysées par PCR avec les

amorces spécifiques ou amorces des bras du vecteur (amorces universelles T7/SP6) pour vérification et remises en culture. La culture bactérienne est récupérée pour réaliser une extraction des plasmides avec le kit ‘Qiagen, QIAprep Spin Miniprep Kit’ en suivant les instructions du fournisseur. Les plasmides purifiés sont ensuite séquencés en Sanger avec les amorces T7 et SP6 pour une lecture de la séquence dans les deux sens (fragments >0,5 kb).

2.4 Expression du gène *Ma*

A partir de jeunes plants de P. 2175 cultivés en condition *in vitro*, une vingtaine de clones ont été transférés pour moitié en boîte de Pétri sur milieu gélosé (condition *in vitro*) et l’autre moitié a été acclimatée en conteneur avec du sol préalablement stérilisé (condition *in vivo*). Pour chacune des conditions, 5 plants ont été utilisés comme témoins non inoculés et 5 autres ont été inoculés avec environ 200 individus de *M. incognita* au stade infestant J2. Après 3 jours, les racines et les feuilles ont été prélevées (entre 100 et 200 mg en poids frais) pour en extraire l’ARN qui est dosé au nanodrop et dilué pour obtenir des échantillons d’ARN homogènes à 100 ng.µl⁻¹. L’ADN complémentaire synthétisé est utilisé en PCR quantitative avec le kit 5x HOT Pol EvaGreen HRM Mix 9ROX en suivant les spécifications du fournisseur, sur Agilent AriaMx Real-Time PCR System.

2.5 Génotypage

Le génotypage de la descendance d’amandier en ségrégation F2 [Alnem 1 x Lauranne] a été réalisé essentiellement grâce à des marqueurs de type microsatellites (SSRs) et SCAR. La liste des marqueurs utilisés est disponible en **Annexe 1**. Ces marqueurs se sont avérés utilisables entre les différentes espèces de *Prunus* : dans ce genre, les premières cartes génétiques construites pour le croisement amandier x pêcher [Texas x Earlygold] ont servi de cartes de référence pour les différentes espèces des sous-genres *Amygdalus* et *Prunophora* (Joobeur *et al.*, 1999) (Aranzana *et al.*, 2003). Cette propriété a permis, en plus des marqueurs déjà existants, d’utiliser le génome du pêcher ‘*Prunus persica* genome assembly v2.0’ pour définir de nouveaux marqueurs afin de génotyper la descendance F2 d’amandier. Les microsatellites ont été marqués par des fluorochromes en 5’ et, après amplification PCR (kit MyTaq DNA polymerase), ont été analysés sur un séquenceur 3130xl Genetic analyzer. Les marqueurs SCAR ont été amplifiés en utilisant le kit MyTaq DNA polymerase selon les indications du fournisseur avec 35 cycles (94°C-30 sec ; Tm-30 sec ; 72°C -1kb/min), analysés sur gel d’agarose 1.5% et les produits PCR ont été vérifiés en séquençage Sanger.

2.6 Phénotypage

Le phénotypage de 1448 individus d'amandier issus de semis de la descendance F2 en ségrégation [Alnem 1 x Lauranne] a été réalisé en serre pendant la période végétative en deux lots. Un premier lot de 541 plants a été testé entre mars et octobre 2016 et un deuxième lot de 907 individus a été testé entre mars et octobre 2017. Les individus ont été plantés par deux en pots de 5 litres. L'inoculum de *M. javanica* a été produit en parallèle dans des godets de tomates (*S. lycopersicum* cv. 'Saint Pierre') inoculées au stade 5 feuilles et cultivées pendant 2 mois. Après ces deux mois la partie aérienne des tomates a été supprimée. Dans chaque pot contenant deux plants d'amandier, le contenu d'un godet (sol contaminé et système racinaire infesté) a été incorporé. Cette méthodologie permet d'appliquer aux plants une pression d'inoculum élevée et continue grâce à la libération séquentielle de milliers de J2s de *M. javanica*. Après 6 mois, les plants ont été extraits de leur conteneur et leur système racinaire lavé et examiné. Chaque plant est noté résistant (R) en cas d'absence totale de galles ou sensible (S) en cas de présence de galles. Globalement, la notation s'effectue aisément du fait de l'effet fort du gène *RMja*, les plants sensibles portant des symptômes très marqués alors que les porteurs du gène en sont totalement indemnes.

2.7 Analyse de séquences

De nombreuses interfaces ont été utilisées pour analyser les séquences d'acides nucléiques et d'acides aminés générées ou extraites à partir de diverses bases de données (voir paragraphe suivant).

La prédiction et l'annotation structurelle des gènes ont été réalisées avec FGENESH en utilisant les paramètres spécifiques de recherche de gènes dans les espèces *Populus trichocarpa* et *Prunus persica*. Les fichiers de sortie génèrent ainsi le découpage intron/exon, les sites de début et de fin de transcription ainsi que les séquences prédites des transcrits et des protéines associées.

Les alignements de séquences (acides nucléiques et aminés) ont été réalisés avec CLUSTAL OMEGA ou MUSCLE pour les alignements comprenant plus de 20 séquences. Les alignements plus fins ont été réalisés avec T-Coffee. Les alignements sont réalisés avec les paramètres par défaut (gap penalty & extension) et sont examinés manuellement afin de valider les alignements proposés.

Les recherches de séquences par similarité ont été effectuées par BLAST. BLASTP a été principalement utilisé avec comme 'Query' divers NBS-LRRs de référence : les domaines TIR des TNLs N (Q40392.1), Ma (CAR94514.1) et RPS4 (CAB50708.1) ; les domaines NBS des TNLs N, Ma et RPS4, du CNL Gpa2 (AAF04603.1) ; les RNLs ADR1 (BAF00531.1) et NRG1 (AAAY54606.1) ; les domaines RPW8 des RNLs (ADR1, NRG1, RPW8.1 (ACJ05907.1) et RPW8.2 (ACJ72031.1)). Les séquences avec une e-value $<10^{-5}$ ont été analysées en annotation fonctionnelle pour vérifier la présence des domaines considérés. La recherche des domaines PL s'est effectuée par BLASTP avec comme 'Query' les motifs

définis dans notre étude et avec des e-value <30 . La stringence de la recherche a été volontairement diminuée en raison de la taille et de la conservation restreinte des motifs du domaine PL. Les séquences obtenues ont été alignées pour validation manuelle. Alternativement les outils FIMO et MAST (MEME suite) ont été utilisés pour effectuer la recherche de motifs PL à travers des jeux de données protéiques. Les motifs protéiques ont été identifiés en utilisant MEME avec les paramètres par défaut à l'exception (i) de la distribution des sites paramétrée sur 'any number of repetition', (ii) du nombre de motifs paramétré à 10, (iii) de la taille des motifs définie entre 5 et 30 acides aminés (min-max). Seuls les motifs non chevauchants et présents dans une majorité des séquences ont été retenus.

L'annotation fonctionnelle (identification des domaines protéiques) des nombreuses séquences retenues est une étape cruciale de l'étude. Elle s'est donc appuyée sur plusieurs plateformes : dans un premier temps une recherche multi-séquences a été effectuée avec les interfaces Pfam (e-value <1), HMMscan (e-values 0.01 (model) et 0.03 (Hit)) et SUPERFAMILY. Les entrées Pfam sélectionnées sont : NB-ARC (PF00931), TIR (PF01582), RPW8 (PF05659) et LRR 1 à 6, 8, 9, et LRRNT (PF00560, PF07723, PF07725, PF12799, PF13306, PF13516, PF13855, PF14580 et PF01462, respectivement).

Les arbres phylogénétiques ont été construits avec MEGA7 ou l'interface 'phylogeny.fr' en utilisant la méthode de maximum de vraisemblance (ML) PhyML program (v3.1/3.0 aLRT). Les modèles de substitution WAG (Whelan and Goldman) et JTT (Jones- Taylor-Thornton) ont été sélectionnés et la fiabilité des branches a été supportée par valeur de bootstrap ou en utilisant le test aLRT (SH-Like). Les arbres phylogénétiques obtenus ont été dessinés en utilisant Dendroscope3.

La liste détaillée des différents outils informatiques utilisés est présentée en **Annexe 2**.

2.8 Données additionnelles et bases de données

L'étude a fait appel à de nombreux protéomes, transcriptomes et génomes de plantes regroupés sur diverses plateformes. Lorsque l'outil BLAST était mis à disposition sur ces plateformes, celui-ci a été utilisé pour rechercher les différentes séquences d'intérêt. Alternativement, les données de transcriptomes ou de peptides prédits ont été téléchargées afin d'utiliser le BLAST en local. La liste détaillée des bases de données utilisées et les versions des jeux de données testés sont présentées en **Annexe 3**.

CHAPITRE 3

LES TNLs DANS LE GENOME DU PECHER ET LA CARACTERISATION DU DOMAINE POST-LRR

Chapitre 3 : LES TNLs DANS LE GENOME DU PECHER ET LA CARACTERISATION DU DOMAINE POST-LRR

3.1 Introduction

Chez le prunier myrobolan (*P. cerasifera*), le gène *Ma* possède une caractéristique biologique remarquable en raison de la résistance complète qu'il confère vis-à-vis de toutes les espèces de nématodes à galles testées. Il est annoté comme un gène de type TNL possédant une extension atypique en position C-terminale constituée de 5 exons répétés de tailles comprises entre 537 bp et 789 bp (Claverie *et al.*, 2011). Ces cinq exons codent pour cinq domaines présentant des similarités avec certaines signatures précédemment identifiées chez des TNLs connus (Gassmann *et al.*, 1999; Dodds *et al.*, 2001) et au sein des TNLs d'*A. thaliana* (Meyers *et al.*, 2003).

Afin de caractériser l'originalité de la structure du gène *Ma* et plus largement de ce domaine, nommé domaine Post-LRR ou PL du fait de sa position, nous avons dans un premier temps annoté structurellement et fonctionnellement l'ensemble du répertoire des TNLs dans le génome du pêcher (génomme modèle pour les *Prunus*) dans sa version 1.0 puis dans sa version la plus récente (v2.0). Nous avons ainsi identifié, positionné et caractérisé 195 séquences de type TNL et confirmé que ces gènes possèdent une distribution majoritairement organisée en clusters. En examinant la région C-terminale des TNLs, nous avons mis en évidence en aval des domaines LRR, la présence d'un domaine codé le plus souvent par un exon indépendant de celui codant le domaine LRR et caractérisé par trois motifs conservés ainsi que par une structure secondaire prédite exclusivement constituée de feuilletts β . Le premier motif contient notamment un résidu proline extrêmement conservé probablement important dans la configuration qu'il donne au motif. Ce motif contient également deux acides aminés aromatiques consécutifs tryptophane-phénylalanine (WF) (ou majoritairement YF chez *Arabidopsis*) qui, de par leur chaîne latérale pourraient être impliqués dans des liaisons intra ou extra-moléculaires. La deuxième signature identifiée séparée du premier motif par un brin β , est elle-même incluse dans un brin β et contient principalement des résidus hydrophobes. En dehors de ces résidus, une glycine et une cystéine se retrouvent très conservées. Enfin le dernier motif situé à l'extrémité C-terminale du domaine PL contient une cystéine et une glycine (CG) conservées (**Annexe 6**).

A travers la caractérisation fine du domaine PL, dont on pensait avant cette étude qu'il n'était qu'une extension du domaine LRR, nous avons montré que le domaine PL est retrouvé dans les deux tiers environ des TNLs chez le pêcher.

Dans un deuxième temps, l'utilisation plus large des signatures identifiées a montré qu'une proportion assez proche était observée dans les autres génomes de Dicotylédones testés. La recherche du domaine PL à travers divers protéomes indique également que ce domaine n'est présent que chez les TNLs. Enfin, nos résultats indiquent que les TNLs présentant des domaines PL dupliqués restent marginaux (6% des TNLs chez le pêcher). En outre, nous avons montré que la présence de ce domaine en cinq copies chez le gène *Ma* représente une caractéristique très probablement unique chez les Angiospermes.

3.2 Article publié dans BMC Genomics

Van Ghelder, C. and Esmenjaud, D. (2016) TNL genes in peach: insights into the post-LRR domain. *BMC Genomics* 17:317. Doi: 10.1186/s12864-016-2635-0.

NB: les données additionnelles liées à cet article sont disponibles aux adresses suivantes:

<https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-2635-0#Sec24>

<https://datadryad.org/resource/doi:10.5061/dryad.dp20h>

<http://www.ebi.ac.uk/ena/data/view/LT555556-LT555568>

RESEARCH ARTICLE

Open Access

TNL genes in peach: insights into the post-LRR domain

Cyril Van Ghelder^{1,2,3*} and Daniel Esmenjaud^{1,2,3}**Abstract**

Background: Plants develop sustainable defence responses to pathogen attacks through resistance (*R*) genes contributing to effector-triggered immunity (ETI). TIR-NB-LRR genes (TNL genes) constitute a major family of ETI *R* genes in dicots. The putative functions or roles of the TIR, NB and LRR domains of the proteins they encode (TNLs) are well documented, but TNLs also have a poorly characterised C-terminal region, the function of which is unknown in most cases. We characterised this prevalent stress-response protein family in a perennial plant, using the genome of peach (*Prunus persica*), the model *Prunus* species. The first TNL gene from this genus to be cloned, the *Ma* gene, confers complete-spectrum resistance to root-knot nematodes (RKNs) and encodes a protein with a huge C-terminal region with five duplicated post-LRR (PL) domains. This gene was the cornerstone of this study.

Results: We investigated the role of this C-terminal region, by first describing the frequency, distribution and structural characteristics of i) TNL genes and ii) their PL domains in the peach genome, using the v1.0 Sanger sequence together with the v2.0 sequence, which has better genome annotation due to the incorporation of transcriptomic data. We detected 195 predicted TNL genes from the eight peach chromosomes: 85 % of these genes mapped to chromosomes 1, 2, 7 and 8. We reconstructed the putative structure of the predicted exons of all the TNL genes identified, and it was possible to retrieve the PL domains among two thirds of the TNL genes. We used our predicted TNL gene sequences to develop an annotation file for use with the Gbrowse tool in the v2.0 genome. The use of these annotation data made it possible to detect transcribed PL sequences in two *Prunus* species. We then used consensus sequences defined on the basis of 124 PL domains to design specific motifs, and we found that the use of these motifs significantly increased the numbers of PL domains and correlative TNL genes detected in diverse dicot genomes. Based on PL signatures, we showed that TNL genes with multiple PL domains were rare in peach and the other plants screened. The five-PL domain pattern is probably unique to *Ma* and its orthologues within *Prunus* and closely related genera from the Rosaceae and was probably inherited from the common ancestor of these plants in the subfamily Spiraeoideae.

Conclusions: The first physical TNL gene map for *Prunus* species can be used for the further investigation of *R* genes in this genus. The PL signature motifs are a complementary tool for the detection of TNL *R* genes in dicots. The low degree of similarity between PL domains and the neighbouring LRR exons and the specificity of PL signature motifs suggest that PL and LRR domains have different origins, with PL domains being specific to TNL genes, and possibly essential to the functioning of these genes in some cases. Investigations of the role of the oversized *Ma* PL region, in ligand binding or intramolecular interactions for example, may help to enrich our understanding of NB-LRR-mediated plant immunity to RKNs.

Keywords: TIR-NB-LRR, Peach, Genome, *Prunus*, Resistance, Post-LRR domain, NB-LRR superfamily

* Correspondence: Cyril.Van-Ghelder@paca.inra.fr

¹INRA, UMR 1355 Institut Sophia Agrobiotech, 06900 Sophia Antipolis, France²University Nice Sophia Antipolis, UMR 7254 Institut Sophia Agrobiotech, 06900 Sophia Antipolis, France

Full list of author information is available at the end of the article



© 2016 Van Ghelder and Esmenjaud. **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided 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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Throughout their lives, plants have to deal with pressures exerted by diverse pathogens, including viruses, bacteria, oomycetes, fungi, and nematodes. Their survival requires the development and maintenance of effective, sustainable defence responses to these biotic stresses. The first line of defence to pathogen attacks involves the early detection of pathogen-associated molecular patterns (PAMPs) through PAMP triggered immunity (PTI) [1]. Pathogens secrete avirulence factors or effectors that manipulate plant immunity and suppress PTI. These factors, also known as *Avr* gene products [2], are then detected directly or indirectly, by the plants, through a second line of defence known as effector-triggered immunity (ETI). ETI involves specific resistance (*R*) genes [1] and genes encoding nucleotide binding–leucine rich repeat (NB-LRR) proteins are the principal class of *R* genes. A wide range of NB-LRR genes have been identified: about 150 in *Arabidopsis* [3], 400 in rice [4] and in poplar [5] and more than 500 in grapevine [6].

NB-LRR genes can be further classified on the basis of their N-terminal domains, into the Toll/interleukin-1 receptor (TIR) NB-LRR and non-TIR NB-LRR (mostly coiled-coil (CC) NB-LRR (CNL)) families [2]. The TIR-NB-LRR family seems to be older than the non-TIR NB-LRR family [7]. TIR-NB-LRR genes (TNL genes) are rare in monocots [8] in comparison with dicots, in which they seem to have emerged earlier in perennials than in annuals [6, 9, 10]. Most of the well-characterized cloned TNL genes [11] belong to *Arabidopsis* [12], but a few originate from plants of agronomic interest, such as potato [13], plum [14] or flax [15]. TNL genes may control plant pathogens as diverse as viruses (*N/TMV*) [16], bacteria (*RPS4/Pseudomonas syringae*) [17] and eukaryotes, such as fungi (*L6/Melampsora lini*) [15] and nematodes (*Gro1-4/Globodera rostochiensis*, *Ma/Meloidogyne* spp.) [13, 14].

TNLs (the proteins encoded by TNL genes) have a conserved organisation into three major domains: the TIR, NB, and LRR domains (in order, in an N-terminal to C-terminal direction). The N-terminal TIR domain, identified by homology with the *Drosophila* cytoplasmic Toll domain, is involved in downstream protein signalling and pathogen recognition, as shown for the flax *L* gene [18]. TNLs, like the product of the *N* gene in tobacco, form oligomers by direct TIR–TIR interaction [19], as an early event in pathogen detection [20]. The NB domain, a central component of TNLs, is involved in an intramolecular interaction with the TIR and LRR domains and in an extramolecular interaction with ATP/ADP [20]. ATP binding activates TNLs, triggering signalling pathways via conformational changes. Following ATP hydrolysis, the TIR domain activates a

downstream signal cascade leading to the plant hypersensitive response (HR). A smaller NLL domain of conserved size, located between the NB and LRR domains, links these two domains through a probable role in TNL protein folding.

The leucine-rich repeat (LRR) domain is located closer to the C-terminus. It is more polymorphic than the N-terminal domains and is subject to diversifying selection [21]. Pathogen effectors are detected by the LRR domain [22–24], and, possibly, by the TIR domain [18]. These effectors may interact directly with the corresponding *R* gene, as for *Avr-Pita* from the rice blast fungus, which recognises the *Pita R* gene of rice [25]. However, the interaction may also be indirect, mediated by plant co-factors or another plant gene, as hypothesized in the ‘guard’ [2], ‘decoy’ [26], ‘bait and switch’ [27, 28] and ‘integrated decoy’ [29] models. A high frequency of genes involved in indirect recognition would account for the limited numbers of *R* genes relative to the high diversity of pathogens and corresponding effectors. For instance, the *Arabidopsis* RPM1 protein recognizes two avirulent proteins, *AvrRpm1* and *AvrB*, via interaction with the RIN4 protein [21]. Conversely, a single *Avr* product may interact with several *R* proteins, as for *AvrRps4* and the TNL *R* proteins RPS4 and RRS1 [30, 31].

The proteins encoded by NB-LRR genes often end in a C-terminal (CT) region comprising other domains of variable structure generally referred as “CT domains”. TNLs have a longer CT region than CNLs [3]. RRS1-R has a large C-terminus including a DNA binding WRKY domain [32, 33] and functions in tandem with RPS4.

In *Prunus*, the TNL gene *Ma* contains a huge CT region with five duplicated exons (each corresponding to a domain in the protein) designated PL (post-LRR) 1 to PL5 [14]. The deduced amino-acid sequence of this PL region of the corresponding protein, the function of which is completely unknown, is longer than the rest of the gene (i.e. TIR, NB, NLL and LRR domains). Moreover, nothing is known about the frequency, distribution and putative structure of this region in its *Prunus* background, represented by the peach (*Prunus persica*) genome ($2n = 2x = 16$). The *Ma* gene not only has a very unusual structure, it also has a unique biological feature in that it confers complete-spectrum resistance to root-knot nematodes (RKN) of the genus *Meloidogyne* [14].

In peach, two genome versions (v1.0 and v2.0) were obtained from a double-haploid genotype of the cultivar ‘Lovell’. In the v2.0 version, the eight chromosomes cover 225.7 Mb and contain 26,873 protein-coding gene loci [34]. The first objective of our study was to determine the *R* gene repertoire of the *Prunus* genus and the abundance of TNLs with atypical architectures and, in particular, architecture similar to that of the *Ma* gene. To this end, we first identified the complete set of

predicted TNL-related sequences in the peach genome, using both the v1.0 [34] and v2.0 genomes, and established the first physical map of these genes for *Prunus* species. The second objective was to study the CT region of TNLs, investigating the frequency, distribution and structural characteristics of PL domains in peach. We took advantage of the *Ma* gene and its unusual PL region, in our analysis of the structure of PL domains in the peach genome and in plant genomes more generally. We identified specific motifs and structural characteristics of PL sequences of use for the further detection and identification of these sequences. Based on our findings, we suggest some possible origins and functions for PL domains.

Results

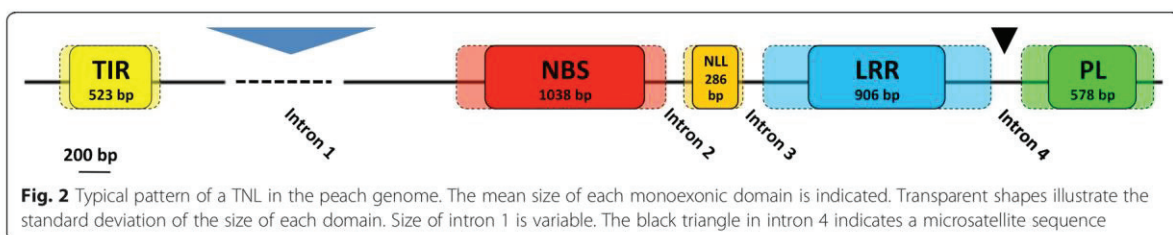
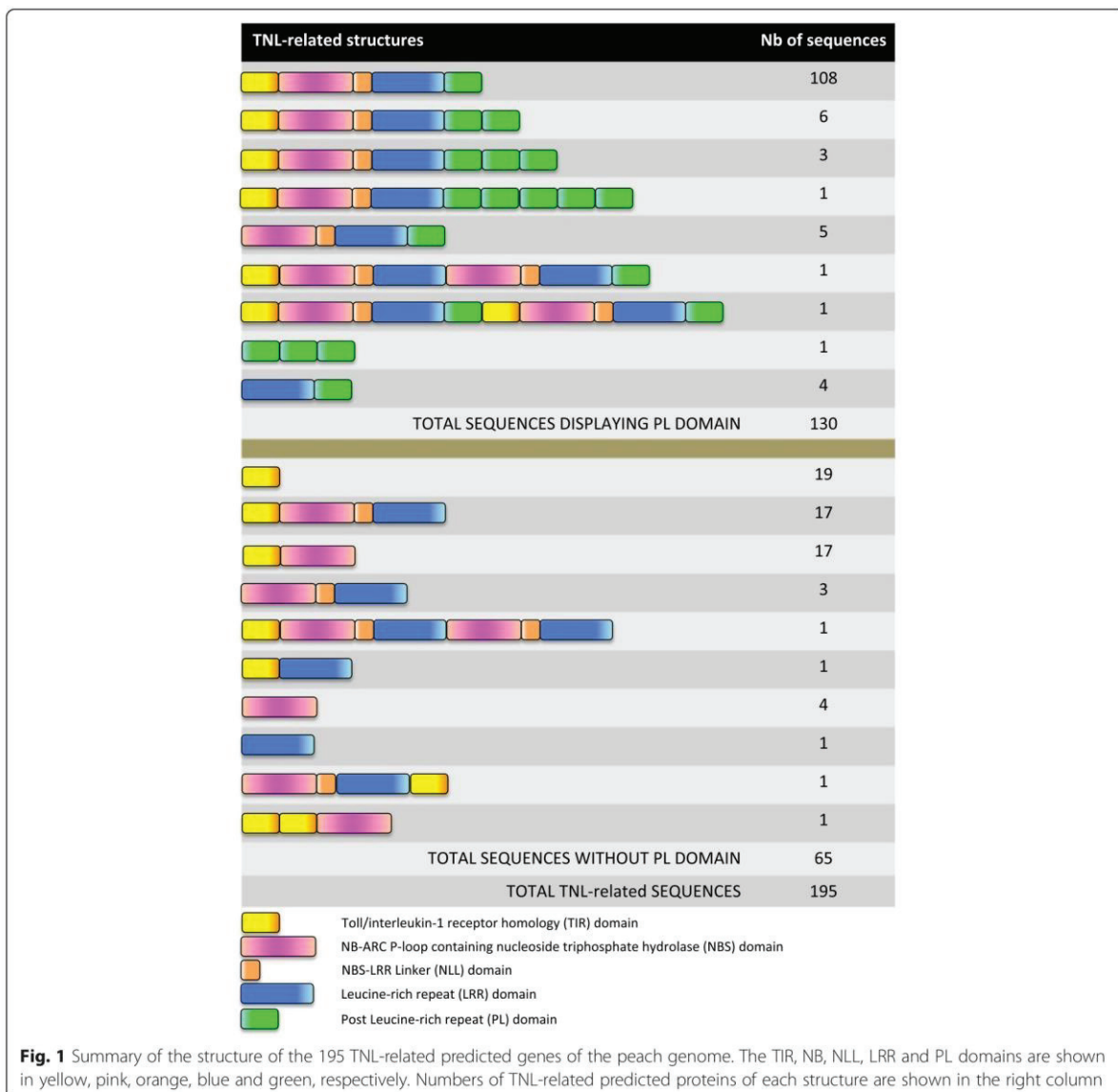
Identification, characterization and distribution of TNLs

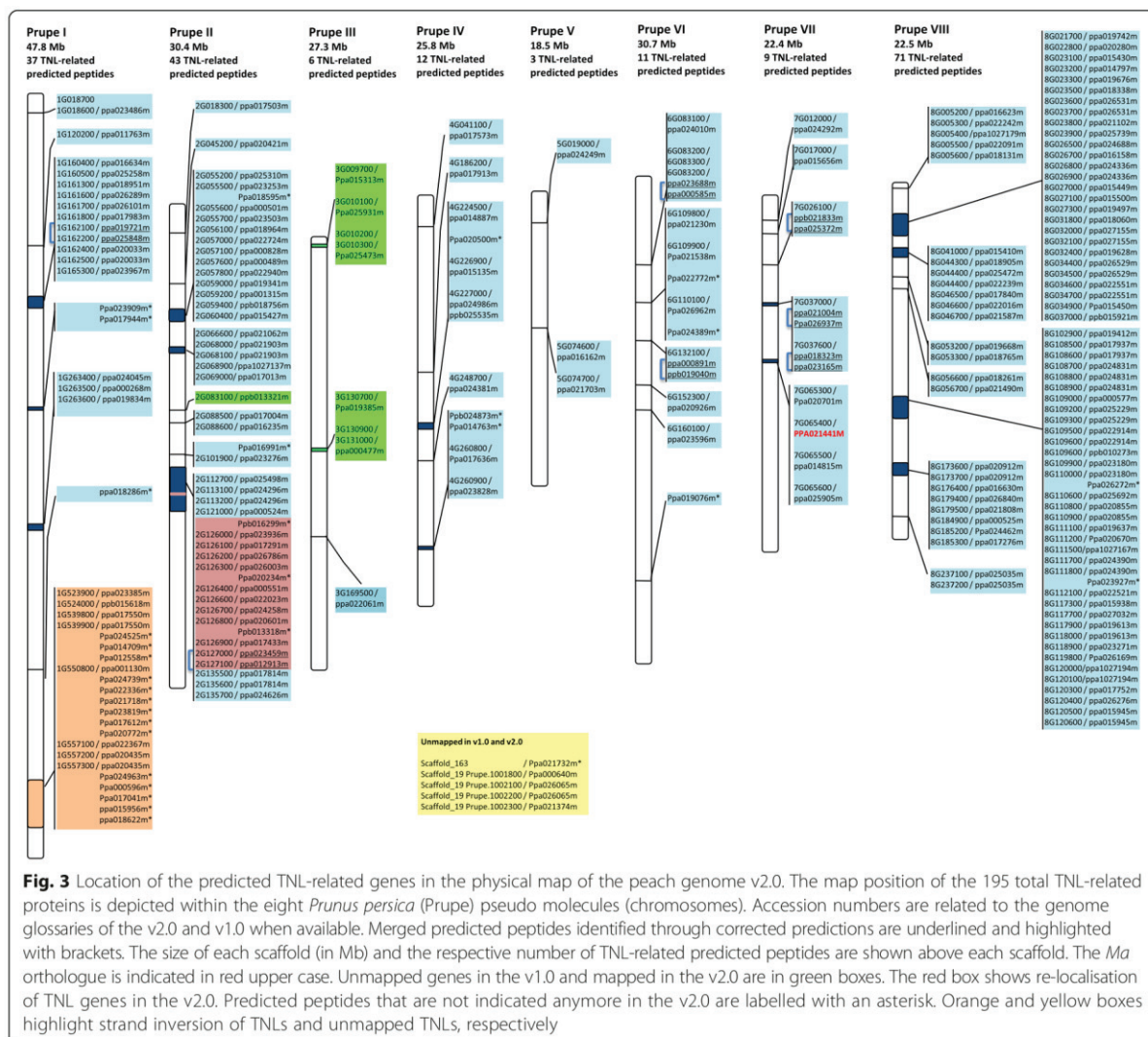
We carried out BLAST P analysis on the predicted peptides database for the peach genome v1.0, with various relevant sequences as queries, to extract the putative peach TNL proteins. With this screening method, we retrieved 205 predicted TNL proteins from the predicted peptide dataset. The complete structure of all predicted TNLs was characterised and refined by retrieving the associated nucleotide sequences together with a 5 kb 5' extension and a 5 kb 3' extension. Analyses of these extended genomic sequences with relevant gene-predicting software suites revealed the presence of unidentified or mispredicted exons, thereby improving v1.0 gene sequence predictions. Fgenesh [35] was the only program to predict the correct intron-exon structure of the *Ma* gene (Additional file 1). The translated DNA coding sequences were then analysed with the InterProScan tool [36], to delineate the various domains in each TNL genes. These procedures made it possible to decrease the total number of TNL genes to 195. This set of TNL genes was also mapped onto the v2.0 genome sequence. The full list of TNL-related genes, with their exon/domain structure, positions and accession numbers (used in both versions of the genome) are indicated in Additional file 2. In parallel, we developed a TNL-specific generic format file (gff3 file, see Additional file 3) to facilitate the comparison between our predictions and the v2.0 genome annotation. We retained the accession numbers from v1.0 of the genome, making it possible to relate the two accession glossaries directly. A comparison of the different types of structures among these TNL genes in the peach genome (Fig. 1) showed that 55 % (108/195) of these genes consisted of five exons corresponding to the following domains assembled in this specific order: TIR, NB, NLL, LRR and PL. The PL domain was present in 67 % (130/195) of the TNL-related sequences. More precisely, 92 % (120/130) of the TNL genes displaying a complete set of classical domains (TIR, NB, NLL and LRR) also had at least one PL

domain. Only 12 of the TNLs had repeats of two or more PL domains and only one peptide (ppa21441m), with the highest degree of similarity to the *Ma* protein, displayed a five-PL domain pattern. Incomplete TNL sequences, often due to recombination processes and frameshift mutations, have also been found and transcriptomic data for the v2.0 sequence have ruled out the possibility of these sequences being pseudogenes. There seems to be a conserved general pattern with a few atypical structures resulting from exon duplication or deletion.

Based on the complete data for the peach genome, we were able to define a typical TNL pattern (Fig. 2). Each single exon of a TNL gene corresponds to a single domain in the corresponding protein and the mean sizes of the TIR, NB and NLL exons were 523, 1038 and 286 bp, respectively. The exons corresponding to the LRR and PL domains had mean sizes of 906 and 578 bp, respectively. These domains displayed greater size variation than the N-terminal domains, consistent with their greater polymorphism and the difficulties encountered in their accurate prediction. The N, BS4 [GenBank:AAR21295] [37], and RPS4 [GenBank:CAB50708] [38] genes, used as reference TNLs, have a single exon encoding both the LRR and PL domains. This feature was rare in the peach genome, in which the intron between the LRR and PL exons was absent from only 12 % of TNL-PL-related sequences. This intron includes a microsatellite sequence in about 60 % of the TNL-PL-related sequences.

The TNL genes were very unevenly distributed between the eight peach chromosomes. Most TNL genes (85 %) mapped to four chromosomes (1, 2, 7 and 8), and such genes were particularly abundant on chromosome 8, which harboured more than one third of all the TNL genes (Fig. 3). Analyses of the v2.0 sequence did not predict 31 of the TNL gene sequences identified in the v1.0 genome sequence (Additional files 2 and 3 and Fig. 3), and the use of the former sequence resulted in slight changes in the predicted distribution of the other TNL genes. Indeed, a large cluster of TNL genes originally mapped to chromosome 7 was predicted to be located on chromosome 2 (Additional file 4). Similarly, the use of the v2.0 sequence made it possible to map five TNLs, that had remained unmapped in the v1.0 sequence, to be localised to chromosome 3. In this new distribution of TNL genes, the numbers of TNL genes on chromosomes 3, 4, 5, 6 and 7 are similar, and there are TNL gene hot spots on chromosomes 1, 2 and 8. The number of TNLs is not correlated with chromosome size, as illustrated for chromosome 8, which is one of the smallest chromosomes present. The 12 TNL genes with multiple PL domains were evenly spread between chromosomes 2, 4, 6, 7, and 8. In this context, the *Ma* orthologue (ppa021441m), with its five-PL domain pattern on chromosome 7, appears to be unique within the peach genome.





Structure of TNL domains
TIR, NB, NLL and LRR domains

We extracted each domain individually, to create a set of relevant sequences for use in motif-based analysis. The vast majority of the sequences obtained from the peach genome included sequences encoding TIR and NB domains, and contained five robust motifs each. Most of these motifs have already been described elsewhere, notably for *Arabidopsis* [3] and *Solanum tuberosum* [39]. All the NB domain sequences contained the characteristic P-loop, RNBS-B and GLPL motifs. For the NLL domain, we identified three recurrent motifs present in 95 % of the sequences. As expected, the repeated nature of the sequences of the LRR domain resulted in a higher degree of polymorphism. From the LxxLxL motif, we developed LRR signatures (Table 1) supported by robust e-values

and characterised by amino-acid sequences also highly conserved in other plant species.

PL domains

We initially defined the PL domains as sequences located in C-terminus of the TNL-related peptides and not matching any known protein domain. We extracted all the PL domain amino-acid sequences and added the corresponding PL domains from the reference TNLs, N, Ma, BS4, RPS4 and Gro1-4 [GenBank:AAP44390] [13]. By combining expert annotation, alignment and size selection, we obtained a complete set of 124 PL domains that we considered reliable for use in subsequent motif detection. Through MEME analysis, we developed three specific signatures of the PL domain (Table 1), which were recovered in 95, 89, and 83 % of the PL sequences,

Table 1 TIR, NB, NLL, LRR, and PL motifs detected from TNLs within the peach genome and from reference TNL genes

TNL sequence numbers	Motif	Signature	MEME motif (consensus sequence)	E-value	Sequences displaying the motif	Similar to (Meyers et al.) [3]	Similar to (Jupe et al.) [39]
146	TIR1	VFLSFRGxDTRxxFxxHL	YDVFLSFRGEDTRKGFTHLYHALXXKGI	3.6e-2895	98 %	TIR-1	motif 18
	TIR2	DDxxLxRG	TFRDDDELERG	2.9e-634	95 %		
	TIR3	YAxSxWCLDEL	YSPELLKAIEESRISIVFSKNYASSTWCLDELVKIL	3.0e-3730	99 %	TIR-2	motif 15
	TIR4	PxFYxVDPsVRxQ	VLPIFYDVPDPSDVRKQTGSFA	1.2e-2248	99 %	TIR-3	motif 13
	TIR5	KVxxWRxAL	DKEKVLWRRAALTEVANLSGW	2.9e-1463	92 %		
120	NB1	GIWGMGGIGKTT	RMVGIWGMGGIGKTTIAKAVYNSISHKFE	8.3e-2337	100 %	P-loop	motif 1
	NB2	GSRIIIT	GKDWFGPGSRIITTRDKHLL	9.0e-1589	100 %	RNBS-B	motif 5
	NB3	GLPLAL	YLELSKRVDYAGGLPLALKVLGSSLYGR	8.1e-1978	100 %	GLPL	motif 3
	NB4	FLDIACF	LDDMEKEIFLDIACFFKGGK	2.7e-1129	92 %	RNBS-D	-
	NB5	MHDL	NKLWMHDLLEQEMGREIVREE	1.9e-1115	92 %	MHDV	motif 7
146	NLL1	FxxMxxLxxL	EAFSKMKNLRLILSNV	5.8e-1047	95 %	-	-
	NLL2	LxxLxWxxxPLxxLP	LRWLCWSGYPLKSLP	2.2e-1037	95 %	-	-
	NLL3	FxPxxLxxLxxMxxS	SNFQPEKLVENMPYSKLRQL	2.6e-1017	95 %	-	-
118	LRR1a	PDFxxxPNLxxLxxCxxL	TKTPDFSGIPNLERLNLEGCTSLVEVHPSI	2.8e-1638	91 %	-	-
	LRR1b	LxxLxxLxxGCxxLxxLP	LKSLETILSGCSKLEKLEPEI	1.1e-789	91 %	-	-
	LRR2a	LxxLxLxGTxxLxxLPxSI	MESLKELDLSGTAIRELPSSI	4.0e-767	92 %	-	-
	LRR2b	LxxCxxLxxLPxxI	NLKDCKNLLKSLPSSI	2.4e-461	90 %	-	-
124	PL-1	IPxWF	PGSEIPEWFSHQSVGD	1.0e-643	95 %	-	-
	PL-2	GxAxCxV	KWMGLALCAVFE	3.2e-386	89 %	-	-
	PL-3	VKKCGxxL	VKKCGVRLVYEQDVEELNQT	9.0e-704	83 %	-	-

Numbers of domain sequences used to detect the motifs are indicated in the left column. Each motif is characterised by its specific signature, its MEME consensus sequence, its e-value and the percentage of sequences displaying it. Some motifs from TIR and NB domains correspond to those described in the literature. Motifs are reported in sequence order

respectively. When present, these motifs were always detected in the same order from the N- to the C-terminus of the protein. The three motifs included highly conserved amino-acids (Fig. 4a) also present in the PL sequences of the five TNL reference proteins (Fig. 4b). The first motif (PL-1) is located at the N-terminal end and is characterised by a highly conserved “proline-X-tryptophan-phenylalanine” sequence. Its hydrophobic region includes conserved proline and aromatic residues (W, F) and is surrounded by polar, negatively charged amino-acids (D or E) upstream, and by polar, positively charged amino-acids (K, R or H) downstream. The second motif (PL-2), which begins close to the end of the PL-1 motif, contains hydrophobic amino-acids and is characterised by conserved glycine and cysteine residues. Secondary structure predictions indicated the presence of beta-sheet structures, particularly between motifs 1 and 2 and, more precisely, between two tryptophan residues separated by 15–25 amino-acids. This prediction of secondary structure within PL domains is illustrated in Fig. 4c for the reference TNLs. For the five PL repeats encoded by the *Ma* gene, this structure was predicted to occur once for the PL1 and PL2 domains and twice for the PL3, PL4 and PL5 domains (Additional file 5). A

sequence length polymorphism between motifs 2 and 3 accounts for the variable size of this domain. Motif 3 contains hydrophobic and polar amino-acids located at the C-terminal end of the domain and the protein.

TNL gene expression

We amplified PL sequences from cDNA, by PCR, to assess the expression of TNL genes and their PL domains in *Prunus* spp. We chose 30 candidate TNL genes from throughout the peach genome that met the requirements shown in Fig. 2, focusing particularly on genes with an intron separating the sequences encoding the LRR and PL domains. Using our sequence predictions, we designed, for each candidate, specific primer pairs binding to the “C-terminal” ends of the LRR and PL exons, respectively (Fig. 5a). For representative material from the genus *Prunus*, we used cDNA from the peach cultivar ‘Nemared’ from the subgenus *Amygdalus* and cDNA from the plum cultivar ‘Myrabi’ from another major *Prunus* subgenus (*Prunophora*). We designed primers amplifying specific fragments from both the genomic DNA and cDNA of the two *Prunus* cultivars. We retained eight candidates, distributed over six chromosomes and expressed in peach and/or plum (Fig. 5c).

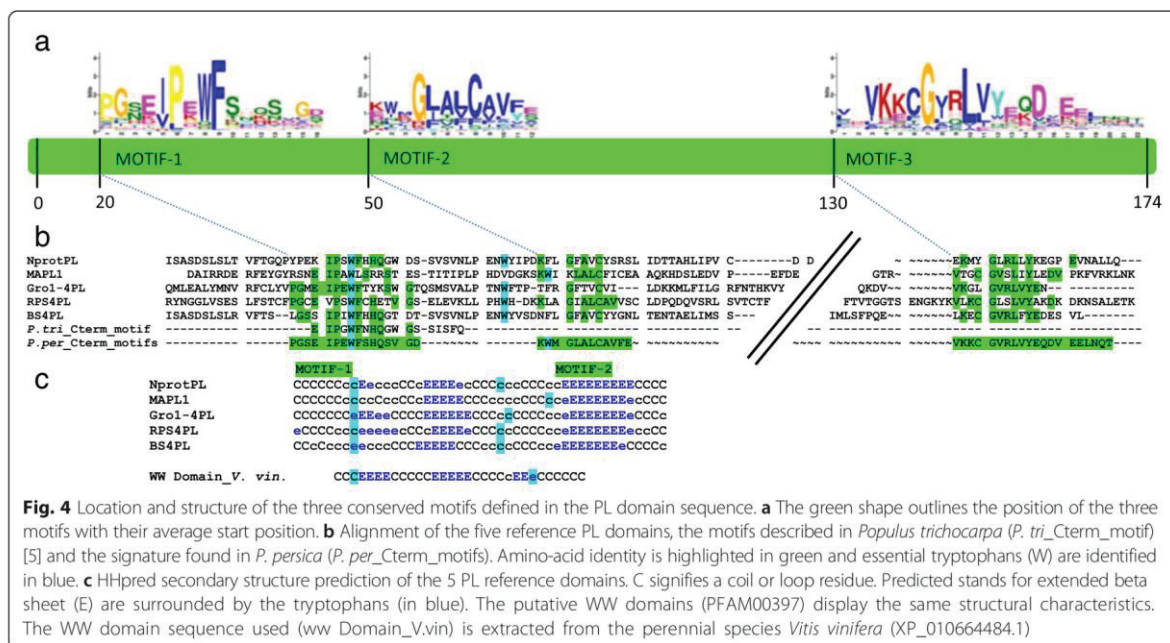


Fig. 4 Location and structure of the three conserved motifs defined in the PL domain sequence. **a** The green shape outlines the position of the three motifs with their average start position. **b** Alignment of the five reference PL domains, the motifs described in *Populus trichocarpa* (*P. tri_Cterm_motif*) [5] and the signature found in *P. persica* (*P. per_Cterm_motifs*). Amino-acid identity is highlighted in green and essential tryptophans (W) are identified in blue. **c** HHpred secondary structure prediction of the 5 PL reference domains. C signifies a coil or loop residue. Predicted stands for extended beta sheet (E) are surrounded by the tryptophans (in blue). The putative WW domains (PFAM00397) display the same structural characteristics. The WW domain sequence used (ww Domain_V.vin) is extracted from the perennial species *Vitis vinifera* (XP_010664484.1)

Three of these candidate genes (ppa022336m, ppa000596m and ppa021732) were amplified despite not having been predicted from the peach genome v2.0 sequence. Otherwise the ppa021732m gene could not be mapped to any of the v2.0 chromosomes and the ppa026786m gene was predicted to be located on chromosome 7 on the basis of the v1.0 sequence, but on chromosome 2 on the basis of the v2.0 sequence. A complete PL sequence was expressed for both genes, as demonstrated by the sequencing of PCR products (Additional file 6) (EMBL accession numbers LT555556 to LT555568). For each candidate, the alignment of the peach ‘Nemared’ and plum ‘Myrabi’ cDNA sequences (when available) with their genomic sequences and their sequences as predicted from mRNA revealed the correct location of the intron between the LRR and PL exons (Fig. 5b). As expected, a lower level of polymorphism has been observed between the genome and the ‘Nemared’ peach sequence than between the genome and the ‘Myrabi’ plum sequence (Fig. 5b).

Phylogenetic analysis

A phylogenetic analysis of the PL domain, conducted with 103 trimmed sequences, showed that PL sequences could be organised into five clades (Fig. 6). As expected from the TNL gene distribution in peach (Fig. 3), most PL sequences mapped to chromosomes 1, 2 and 8, and these sequences could be grouped into identifiable clades. Clade 2 groups together PL sequences from chromosome 1. Clades 1 and 4 were both initially located in chromosome 7 on the basis of the v1.0 genome sequence. Analysis of the v2.0 sequence relocated clade 1 to chromosome 2, whereas the

sequences of clade 4, including the *Ma* gene orthologue in particular, mapped to chromosome 7. The PL sequences from chromosome 8 (which had the highest proportion of PL sequences) were split between clades 3 and 5. These data highlight the correlation between the similarity and physical location of the PL domain sequences. This correlation made it possible to attribute the unmapped TNLs ppa026065m and ppa000640m from clade 3 to chromosome 8 and the unmapped TNL ppa021732m from clade 2 to chromosome 1. Similarly, the four unmapped TNLs from the same subclade unmapped in the v1.0 sequence, ppa015313m, ppa025931m, ppa025473m and ppa000477m, were localised to the same cluster on chromosome 3 of the v2.0 sequence (Additional file 4).

Within the *Ma* gene cluster, the higher degree of conservation of the PL sequences between *Ma* [GenBank: CAR94514] and its peach orthologue (ppa021441m) than between the *Ma* peach orthologue (ppa021441m) and its peach paralogue (ppa014815m), all belonging to clade 4, indicates that the separation of the peach and plum lineages occurred after the TNL duplication. As expected, the three reference TNL genes from the Solanaceae family (*BS4*, *N* and *Gro1-4*) were grouped together into the same subclade, whereas the crucifer *RPS4* TNL gene belonged to another subclade. Nevertheless, these outer groups are not isolated from the peach TNL genes.

PL detection among genomes

MAST and BLASTP analyses were conducted with the PL motifs and the Plant Genome Database, to determine the number of TNL-related protein genes that could be

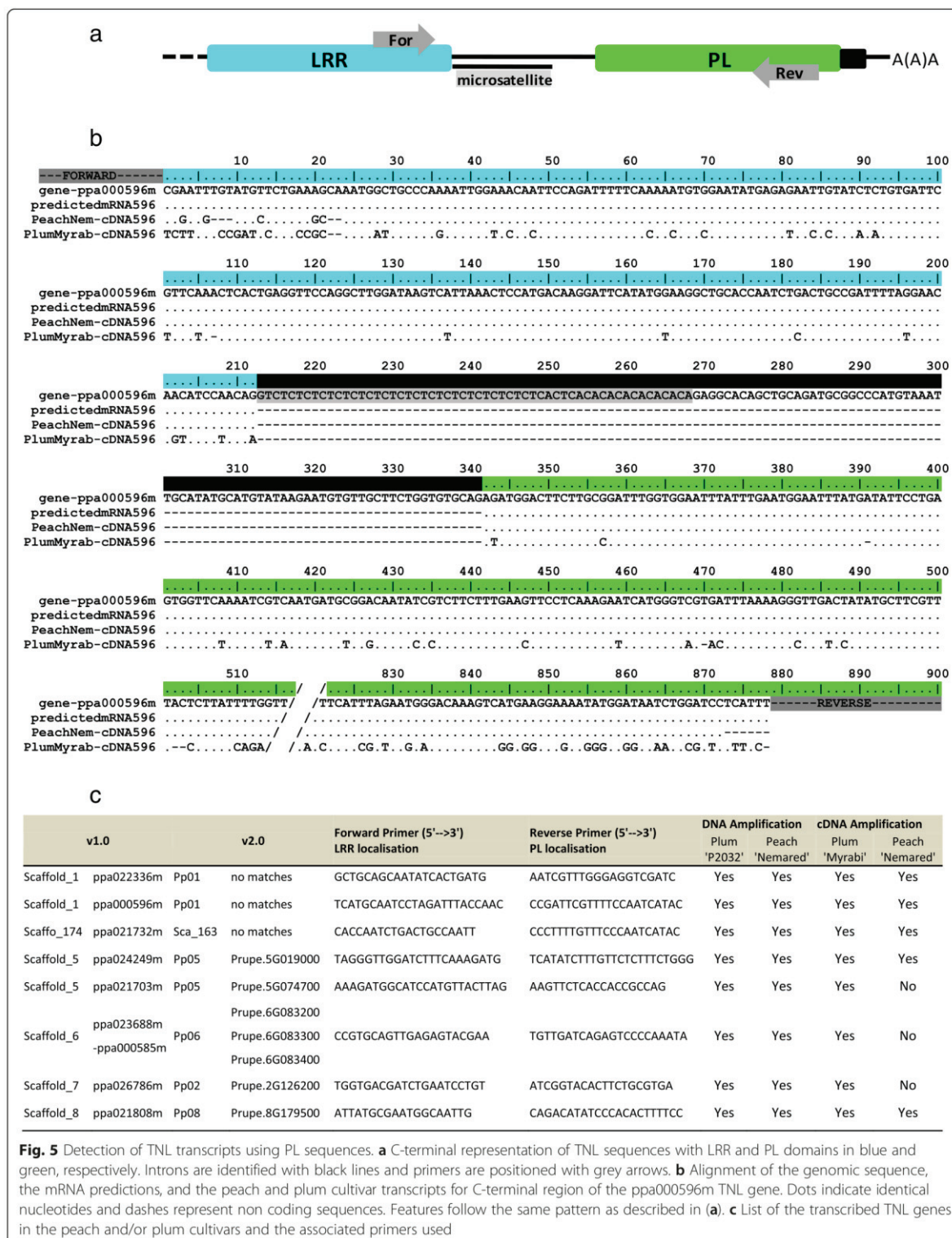


Fig. 5 Detection of TNL transcripts using PL sequences. **a** C-terminal representation of TNL sequences with LRR and PL domains in blue and green, respectively. Introns are identified with black lines and primers are positioned with grey arrows. **b** Alignment of the genomic sequence, the mRNA predictions, and the peach and plum cultivar transcripts for C-terminal region of the ppa000596m TNL gene. Dots indicate identical nucleotides and dashes represent non coding sequences. Features follow the same pattern as described in (a). **c** List of the transcribed TNL genes in the peach and/or plum cultivars and the associated primers used

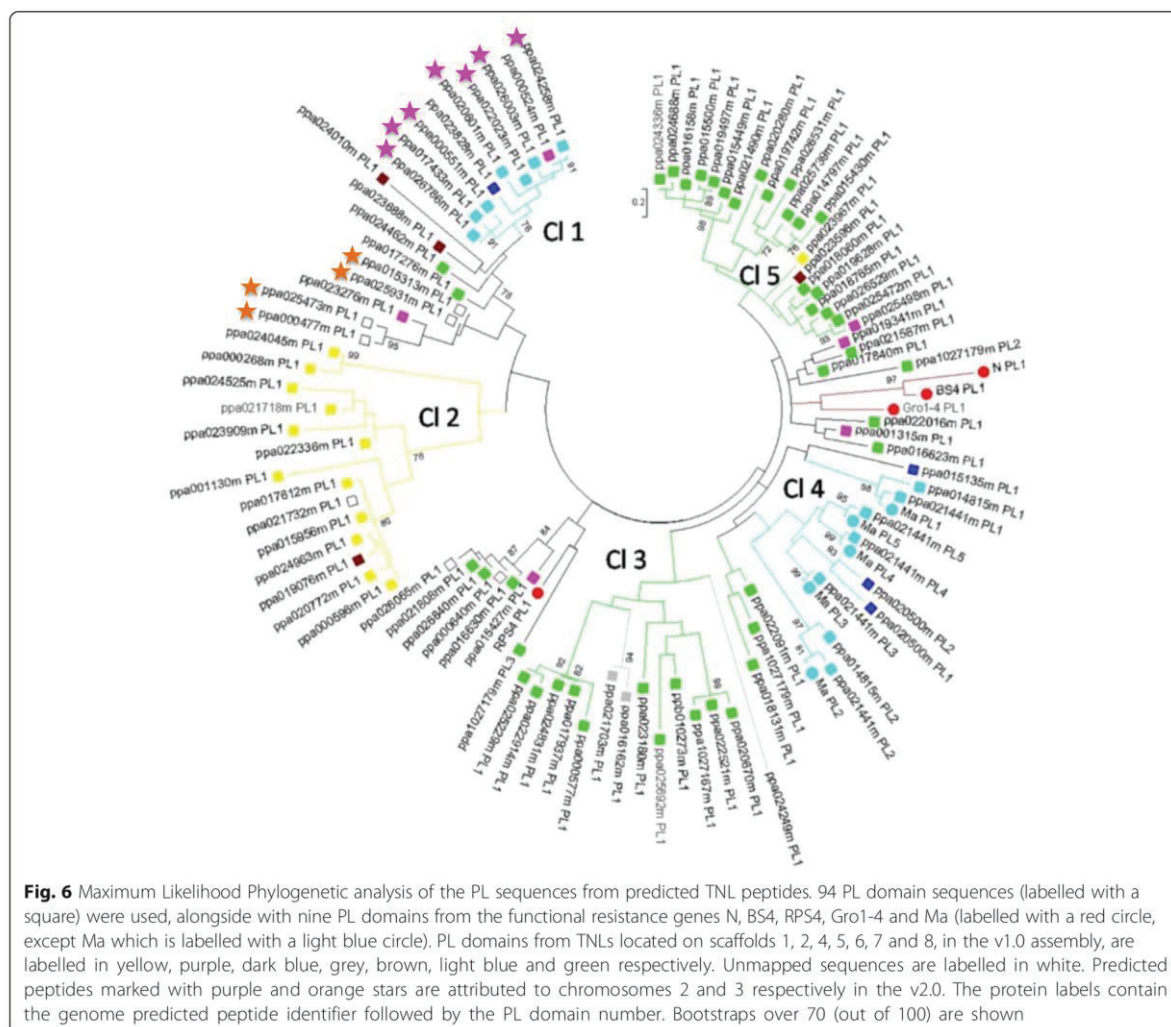


Fig. 6 Maximum Likelihood Phylogenetic analysis of the PL sequences from predicted TNL peptides. 94 PL domain sequences (labelled with a square) were used, alongside with nine PL domains from the functional resistance genes N, BS4, RPS4, Gro1-4 and Ma (labelled with a red circle, except Ma which is labelled with a light blue circle). PL domains from TNLs located on scaffolds 1, 2, 4, 5, 6, 7 and 8, in the v1.0 assembly, are labelled in yellow, purple, dark blue, grey, brown, light blue and green respectively. Unmapped sequences are labelled in white. Predicted peptides marked with purple and orange stars are attributed to chromosomes 2 and 3 respectively in the v2.0. The protein labels contain the genome predicted peptide identifier followed by the PL domain number. Bootstraps over 70 (out of 100) are shown

detected in ten reference genomes of various sizes. The *P. persica* dataset was used to assess the relevance of our approach, Using PL motifs, we were able to detect 72 % (94 of 130) of all peach TNLs with one or several PL domains (Table 2). Indeed, the automated prediction for v1.0 did not detect PL domains for any of the 36 missing sequences that were later annotated as PL-carrying TNL genes. Nevertheless, a combination of the use of transcriptomic data and the reannotation of the v2.0 genome sequence resolved most of the mispredicted PL domains. In dicots genomes, between 50 and 87 % (Table 2) of TNL sequences could be detected using PL signatures. For example, in the *P. trichocarpa* genome [5] 69 of the 91 TNL related-genes identified presented PL signatures. Thus, at least three quarters of the TNLs of this species carry a PL domain. The proportion of TNL genes with a PL domain was highest in the potato genome (87 %). By

contrast, no PL domain sequences were detected in the three Poaceae species (*O. sativa*, *S. bicolor* and *Z. mays*), consistent with the almost total absence of TNLs in monocots. Moreover, no non-TNL sequences were found to have PL motifs with a consistent e-value (<0.01), suggesting that these motifs are specific to TNLs. The detailed results of the BLASTP analysis are reported in Additional file 7. Furthermore, the use of our motifs led to the detection of particular PL domain repeats in peptide databases. We detected a few proteins with atypical structures in *V. vinifera* [Grape Genome Browser:GSVIVT01037222001] and in *P. trichocarpa* [Plant GDB:POPTR_0019s09760.1], with LRR-PL-LRR-PL-LRR-PL and TIR-NBS-LRR-PL-PL domain chains, respectively (Additional file 7). Surprisingly, the structure of the *Prunus* Ma protein, with its five-PL domains, remained unique, as no other predicted TNL proteins with a similar pattern of multiple PL domains

Table 2 Number of TNL-related proteins recovered in seven dicots and three monocots genomes

	Plant species	Genome size	Nb NBS-LRR	Nb TNL	Source	Nb TNLS detected using the PL motifs	% of TNL-related protein containing PL signatures
Dicots	<i>Prunus persica</i>	265	437	185	[34]	94	51
	<i>Populus trichocarpa</i>	485	402	91	[5]	69	76
	<i>Vitis vinifera</i>	487	233	97	[6]	62	64
	<i>Arabidopsis thaliana</i>	135	149	94	[3]	67	71
	<i>Solanum tuberosum</i>	844	438	77	[39]	67	87
	<i>Glycine max</i>	1115	319	116	[58]	50	43
	<i>Medicago truncatula</i>	500	333	156	[59]	78	50
Monocots	<i>Oryza sativa indica</i>	466	653	0	[60]	0	0
	<i>Sorghum bicolor</i>	730	274	0	[61]	0	0
	<i>Zea mays</i>	2400	109	0	[62]	0	0

Numbers of TNLS harbouring PL motifs in the Plant GDB and their ratio are indicated from total putative TNLS reported in the literature

were identified in any of the other plant genomes tested. An orthoMCL analysis on 31 plant genomes conducted with the protein orthologous to Ma in peach (ppa021441m) revealed an orthoMCL group (GRP18645) including the other three members of the ppa021441m cluster (ppa014815m, ppa020701m, ppa025905m) (Fig. 3) and six other members from the apple protein dataset (*Malus domestica*) [GDR:MDP0000267764, MDP0000176003, MDP0000272063, MDP0000158765, MDP0000264373, MDP0000143700]. MDP0000176003 is presumably the Ma apple orthologue (lacking the TIR domain), as it has a similar five-PL domain structure. Finally, we tried to detect these motifs in other *R* genes encoding proteins with LRR domains, such as transmembrane-LRR and CNL proteins. We detected no PL motifs in any of the genes tested (complete list of genes in Additional file 8).

Discussion

Distribution and characterization of TNLS

Following on from the 2003 study by Meyers et al. in *Arabidopsis* [3], many studies have made use of the increasing number of plant genomes available to describe the distribution and diversity of NB-LRR genes in genomes [5, 39–41]. We focused on TNL-related sequences as the major group within the NB-LRR superfamily. We first identified the predicted TNLS within the peach genome, which is considered to be representative of the genus *Prunus*. We found that the 195 predicted TNL sequences retrieved from the peach genome were very unevenly distributed between the eight chromosomes, with 85 % located on four chromosomes (1, 2, 7 and 8), one of which (chromosome 8) carried more than a third of all the predicted TNL sequences. As expected, a large proportion (86 %) of these TNLS were grouped into clusters of at least two duplicated sequences separated by short physical distances and displaying a high degree of sequence similarity. Three of these clusters

included at least 20 TNLS. These data thus confirm the involvement of tandem or segmental gene duplication and recombination as major mechanisms of *R* gene expansion in peach and other plant species [42]. In perennial plants, which have a much longer generation time than annuals, and particularly in those of the Rosaceae, these mechanisms may constitute an adapted response to pathogens, contributing substantially to the great diversity of this group of proteins, as observed in apple [40]. We used the large set of sequences obtained from peach to identify motifs specific to the various TNL domains. Our findings confirm the conserved motifs previously identified for the TIR and NB domains by Meyers et al. [3] and Jupe et al. [39] and identify original specific motifs for NLL and LRR sequences.

By contrast, little is known about the C-terminus of TNLS, other than for the WRKY domain [32, 43]. This study, in which peach was used as the reference genome, increases what is known about the frequency, distribution and structural characteristics of PL domains in TNL genes. With this objective in mind, given that automatic PL domain predictions are currently not very reliable, we first re-annotated the complete set of peach TNL sequences. This re-annotation increased the number of PL domain sequences detected within TNL genes from 94 initial sequences (in the v1.0 genome annotation) to 130 final sequences. The v2.0 genome, with the incorporation of transcriptomic data, thus represents a real improvement. Based on these high-quality PL sequences, we were able to identify three conserved motifs flanking polymorphic regions. Using these conserved motifs, we were able to retrieve PL sequences from the vast majority of sequences encoding TNLS with complete classical domains (TIR, NB, NLL and LRR). Beyond *in silico* prediction, we were able to amplify TNL transcripts generated from genes located on different chromosomes. Our results illustrate that (i) TNL transcripts can be identified through

the detection of PL domain sequences, (ii) our predictions are relevant even for transcripts from the v2.0 genome sequence, and (iii) the peach genome can be used as a reliable matrix for gene studies in the various species (peach, almond, plum, etc.) of the genus *Prunus*.

The cloning and sequencing of the *Ma* gene from *Prunus* revealed the presence of duplicated PL domains [14], providing us with an opportunity to enhance our knowledge of this particular region. Using peach PL sequences and PL sequences from reference TNL genes, we then carried out the first phylogenetic analysis of the PL domains in the peach genome. We found that the PLs present on the same chromosome evolved with their mother TNL genes, through duplication processes [42], resulting in similar sequences and physical positions. Moreover, TNLs with multiple PL domains were found to be rare, with only 12 TNL genes found to encode proteins with more than one PL domain. The *Ma* orthologue was the only protein encoded by the peach genome that was found to include five PL domains. The *Ma* gene, as the only functional TNL gene with multiple domains identified to date, remains a reference sequence for structural analyses of PL domains. The phylogenetic localisation of PL domain sequences from the *Ma* cluster and from genes orthologous to *Ma* demonstrated that TNL diversification occurred before the splitting of the *P. persica* lineage from the plum species acting as the *Ma* donor, *P. cerasifera*. Moreover phylogenetic and orthoMCL analyses reveal that the 5-PL *Ma* structure, presumably present in *Ma* orthologous genes from other *Prunus* species, is also present in its apple (*Malus domestica*) orthologue. The genera *Prunus* and *Malus* belong to the Rosaceae family, which can be subdivided into two prevalent subfamilies, Spiraeoideae (*Prunus/Malus*) and Rosoideae (*Fragaria/Rosa*) [44]. Thus, within the Spiraeoideae, ancestor genes with the multiple-PL domain structure of *Ma* predate the separation of the genera *Malus* and *Prunus* [45]. Finally, our results show that exon duplications have led to the multiple-PL domain structure of *Ma* and generated a relatively high level of polymorphism. This polymorphism may have given rise to the putative original function of the duplicated PL region of *Ma*. We found that 60 % of PL-carrying TNLs had a microsatellite sequence in the intron separating the LRR and PL domains, providing clues to a possible modular mechanism of evolution. This polymorphic site is suitable as a molecular marker for the detection and monitoring of *R* genes, for plant breeding purposes, for example.

Our findings confirm that PL domains are basic components of TNLs not only in peach, but also in other dicots. As expected, no PL sequences were detected in the Poaceae species, consistent with the almost total absence of TNLs in monocots [8]. Using the specific PL motifs defined, we were able to recover PL domains with similar

frequencies from the TNLs of dicots from various botanical families, and with different genome sizes and NB-LRR pools. In addition to detecting TNLs, this approach also made it possible to determine the prevalence of this domain within dicot species. Indeed, the proportion of PL-carrying TNLs ranged from 43 % in soybean to 87 % in potato (including the functional *R* gene *Gro1-4*) [13]. We can assume that these values are underestimates, due to the difficulties in the annotation of this C-terminal domain, which are likely to be resolved by transcriptomic data. Our results highlight the unusual nature of the PL domain, with its specific signature motifs, providing strong support for the use of this domain as a specific marker of TNLs.

Putative origin of the PL domain

The location of the PL domain downstream from the LRR domain, as illustrated in peach, suggests that it may be considered a simple extension of the LRR domain. This hypothesis is supported by the finding that the LRR and PL domains are sometimes encoded by a single exon, as for the reference genes BS4, RPS4 and N. This situation occurred at low frequency (12 %) in the peach genome, suggesting that the two domains may originally have been separated, with fusion into a single exon resulting from intron deletion. It is commonly thought that the PL domain originated from an expansion, by (partial) duplication, of the LRR domain. There are two implications of this hypothesis. Firstly, PL domains should also be present in other *R* gene families containing LRR domains, such as transmembrane-LRR (*Cf* genes) [46], kinase-LRR (*Xa21* gene) [47], and CNL (*Mil-2* gene) [48] genes. However, this was not found to be the case, as none of the genes from these *R* gene families were found to carry PL domains. The second implication is that LRR motifs should also be detected, even residually, in the PL domains and vice versa, whereas this was not the case. As already reported for *Arabidopsis* [3], the C-terminal domain of TNL proteins has no conserved LRR motif. It therefore seems highly unlikely that the PL domain originated from the LRR domain and it seems more probable that the only relationship between these domains is their location within the protein.

Putative role of the PL domain in ligand binding and/or intramolecular interactions

The absence of PL domains from proteins other than TNLs and their systematic location in the C-terminal position after the LRR domain in TNLs suggest a possible role of these domains in recognition via ligand binding. Recent studies on the CNL gene RGA5 have highlighted the importance of a C-terminal domain located after the LRR domain. Indeed, the heavy metal interaction domain concerned is directly involved in the

binding and recognition of rice blast fungus effectors [49]. The PL domain is not related to the C-terminal sequence of RGA5, but its specific amino-acid sequence or structural features may be consistent with a particular role or function. Interestingly, all three PL-specific motifs are present together in the functional TNL genes *Ma* (several PL repeats), *Gro 1–4*, *RPS4*, *BS4* and *N* (single PL domains). These three conserved motifs, two of which (motifs 1 and 2) were identified by Dodds et al. [23] and confirmed by Claverie et al. [14], are separated by a conserved beta-fold structure. The WW domain has several features in common with the PL domain in that it consists of two tryptophan residues separated by 20–23 amino-acids, with one conserved proline and three beta-sheet structures. WW domains are involved in binding to particular proline rich-motifs [AP]-P-P-[AP]-Y in other proteins [50]. Such similarities, together with the characteristics of the PL domain, suggest that the PL domain may be involved in ligand recognition. The PL domain may also be involved in the intramolecular interactions of TNLs, and, in particular, in interactions or joint action with the TIR domain. Like the TIR domain, which is involved in both downstream signalling and ligand recognition [51], PL domains may modulate activity and interact with the TIR domain. Indeed, in the closed three-dimensional structure of TNLs, these two domains are located in close proximity.

TNLs involved in atypical processes, with other specific signatures, such as a signal peptide (L6 from flax) or a WRKY domain (*RRS1* from *Arabidopsis*), have no PL domains. This is consistent with a putative general function of PL domains in TNLs located in the cytoplasm. Recent studies on *RRS1* [52, 53] have highlighted a particular process involving dimerisation between *RPS4* and *RRS1* via the TIR domain, but did not attribute a function to the PL domain of *RPS4* (named CTD). The possibility of a role for the PL domain in downstream signalling cannot, therefore, be excluded. Indeed in paired R proteins, this domain is present in each of the TNL partners, *RPS4* and *RPS4B*, of the WRKY-carrying proteins, *RRS1* and *RRS1-B*, respectively [30]. However, no *RRS1* orthologue and no other TNL with a WRKY domain has been found in the peach genome, suggesting that this particular process may not occur in peach.

Given the small number of sequences with multiple PL domains detected, TNL genes with multiple PL domains should be considered exceptional in plant genomes. The unique five-PL domain structure of *Ma* may be involved in the recognition of effectors or guard proteins and/or in modulation of the plant hypersensitive response. The *Ma* gene confers high, complete-spectrum resistance to all *Meloidogyne* species tested, including the predominant species from the polyphagous root-knot

nematode (RKN) complex, *M. arenaria*, *M. incognita*, and *M. javanica*, and the invasive species *M. enterolobii* [14]. By contrast, other RKN *R* genes from *Prunus*, such as the peach *RMia* gene (acting against *M. incognita* and *M. arenaria*) and the almond gene *RMja* (acting against *M. javanica*), display a more restricted resistance pattern [54, 55]. The complete resistance spectrum of *Ma* may be linked to a broad-spectrum recognition of effectors or guard proteins. We cannot, therefore, exclude the hypothesis that this recognition is mediated by both the LRR domain and one or several repeats in the PL region. The rare event of PL duplication, initiated from *PL1*, may have generated the steric conformation of the *Ma* TNL, ultimately leading to more sustainable and extensive resistance. In *Prunus*, gene editing with C-terminally truncated or mutated regions should shed light on the role of these PL repeats. Just as the Rx response to the potato virus X coat protein was shown to depend on the C-terminal region of the LRR domain [56], modifications of the PL region may influence the degree or spectrum of resistance conferred by the *Ma* gene.

Conclusions

The peach genome contains numbers of TNL genes similar to those in other perennial plants with genomes of similar size. The distribution of the genes of this family and of the variants of its specific domains between peach chromosomes confirmed the hypothetical mechanisms of TNL expansion. PL domains were recovered from about two thirds of the TNLs identified. The PL dataset for the peach genome was used to define specific motifs that will be useful complementary tools for detecting and localising TNLs and their variants in plant genomes. PL domains may be required for the function of some TNL proteins, particularly through specific motifs potentially involved in the recognition of effectors or guard proteins. TNL mapping and the definition of PL domain signature sequences should make it possible to identify functional TNLs from peach or, more widely, in other perennial plants and members of the Rosaceae. TNLs with multiple PL domains are rare in peach and, presumably, in other plants and the five-PL domain structure of *Ma* appears to be unique in peach and, more generally, among perennial and annual plants. Investigations of the putative role or function of the oversized PL region of the *Ma* gene in the resistance properties (high, broad-spectrum resistance) should increase our understanding of the mechanisms of resistance involving NB-LRR-based plant immune receptors.

Methods

Detection and mapping of the TNL genes

TNLs in the *Prunus persica* v1.0 genome were detected by Blast P analysis with the NCBI BLAST tool available

from the GDR website (https://www.rosaceae.org/tools/ncbi_blast). The blast analyses were carried out with the default parameters, except for the description and alignment parameters, which were increased to 500 to visualize as many results as possible in the graphical overview. We used the sequences of the TMV resistance protein N (Q40392.1) [57], the RKN resistance protein Ma (CAR94514.1) [14], the TIR-1, -2, -3, and -4 motifs developed by Meyers et al. [3] and the TIR motifs 13, 15, and 18 developed by Jupe et al. [39], as queries, and the peach genome v1.0 predicted peptides as the database. The N and Ma sequences were chosen because they appeared to be the most relevant for TNL detection. Indeed the N gene is the reference TNL gene for plants in general, and the Ma gene is the only *Prunus* TNL R gene to have been functionally validated. The TIR domain is characteristic of the TNL family. We therefore retained the TIR motifs developed from genome-scale data for *Arabidopsis* and potato. As a complementary tool for the detection of TNL proteins, we used the PL motifs defined in this study to query the database. We retained the predicted peptides with an e-value below 1e-04 for detection with N and Ma, below 0.068 for detection with the TIR motifs and below eight for detection with the PL motifs. The protein sequences retrieved were subjected to InterProScan analysis to confirm that they corresponded to TNLs. We also cross-referenced the data with the complete list of proteins from the predicted proteins database available from the GDR website (ftp://ftp.bioinfo.wsu.edu/species/Prunus_persica/Prunus_persicagenome.v1.0/homology/Prunus_persica_v1.0_vs_swissprot.xls). Finally, all the predicted TNLs recovered were visualised individually with the GDR Gbrowse tool (http://www.rosaceae.org/gb/gbrowse/prunus_persica/). The link between the v1.0 and v2.0 genomes was obtained with the Blast tool. The mRNA sequences predicted from the v1.0 genome were blasted against the v2.0 genome pseudomolecules and scaffolds, to determine intron/exon positions, and ORFs were defined for each exon of each candidate for annotation purposes (Additional file 3). This file must be uploaded onto a local Gbrowse tool, such as IGV, or obtained via the Gbrowse tool of the GDR website (https://www.rosaceae.org/gb/gbrowse/prunus_persica_v2.0.a1/), together with the *Prunus persica* v2.0 genome.

Identification and structure of TNL-related proteins

The nucleotide sequences of predicted TNL genes were extracted with 5 kb of upstream sequence and 5 kb of downstream sequence, with the GDR Sequence Retrieval tool available from <http://www.rosaceae.org/retrieve/> sequences. Fgenesh, an HMM-based *ab initio* gene predictor (gene-finding parameters: *Solanum lycopersicum*,

Vitis vinifera, *Hevea brasiliensis*, and *Populus trichocarpa*), was used to predict TNL gene structures. The individual analysis and expert annotation of these predicted genes made it possible to assemble their intron-exon structures. The predicted amino-acid sequences were then analysed with InterProScan 5, which was used to identify the various domains with default parameters and with the whole set of available databases. This approach made it possible to associate exon structures with families of domains.

PL transcript sequences

Two members of the genus *Prunus*, the peach (*P. persica*) cultivar 'Nemared' and the Myrobalan plum (*P. cerasifera*) cultivar 'Myrabi' (P.2032), were grown in a growth chamber. Young leaves from each cultivar were collected so as to obtain samples of 100 mg of biological material. Half of each sample was used for DNA extraction [55], the rest being used for total RNA extraction with the TQ RNA Cells Tissues Kit (Talent). An additional DNase digestion was carried out with Turbo DNA-free (Ambion), and the quality of the final RNA preparation was checked by electrophoresis in a 1.5 % agarose gel. First-strand cDNA synthesis was performed with oligo(dT) primers and the Maxima H Minus first-strand cDNA synthesis kit (Thermo Fisher). PCR was carried out with the MyTaq polymerase kit (Bioline), specific primer pairs (Fig. 5c) and the following steps: initial denaturation for two minutes at 94 °C, followed by 35 cycles of 30 seconds at 94 °C, 30 s at the annealing temperature and one minute at 72 °C, and a final extension step of three minutes at 72 °C. PCR products were run on 1.5 % agarose gels and the fragments observed were extracted from the gel with the Minelute gel extraction kit (Qiagen). Purified PCR products were sequenced with Sanger technology. All experiments were conducted according to the kit manufacturer's instructions, unless otherwise specified.

Sequence alignment and generation of a phylogenetic tree for PL domains

We trimmed the 124 sequences to obtain a final alignment composed of 103 PL domains (including 9 PL domains from N (1), Ma (5), RPS4 (1), BS4 (1) and Gro1-4 (1)). The full sequence alignment (246 positions, gaps included) included the three PL motifs. Domain sequences from the five cloned functional reference TNLs, RPS4 [GenBank:CAB50708], N [SwissProt:Q40392], BS4 [GenBank:AAR21295], Gro1-4 [GenBank:AAP44390], and Ma [GenBank:CAR94514], were extracted as previously described. The amino-acid sequences of the domains were manually curated and aligned, using MEGA6 and Muscle with default settings. The phylogenetic tree for PL domains was

constructed with MEGA6, using a Maximum Likelihood method with 100 Bootstrap replications and the Jones-Taylor-Thornton model.

Definition and analysis of specific PL-domain signatures

Using the complete dataset for TNLs, we confirmed the position of each domain by InterProScan analysis, and we extracted each domain individually. For each of the TIR, NB, NLL, and LRR domains, we removed those (i) encoded by more than two exons and (ii) with domain lengths of less than 450, 950, 260, and 750 bp, respectively. We then performed a rough alignment for each domain, to remove potential aberrant sequences. The final set of sequences obtained was analysed with MEME software. As for other TNL domains, we considered a single exon to correspond to a single domain and we arbitrarily decided to consider exons consisting of more than 400 nucleotides as valid PL domain exons. After this initial selection on a size basis, a rough alignment was used to exclude aberrant sequences. MEME software was used to identify specific signatures from sequence alignments. The settings used were as follows: 0 or 1 occurrence of the motif per sequence, with a maximum of five different motifs and a minimum-maximum motif width of 5–30 amino-acids. MAST and BLASTP analyses conducted with default parameters identified PL domains in predicted peptides. BLASTP analyses with the Plant Genome DataBase website (<http://www.plantgdb.org/cgi-bin/blast/PlantGDBblast>) led to the identification of PL domain-related sequences and, therefore, TNL sequences. The genomes used for this study were those of perennial (peach, poplar, grape) and annual (barrel clover, potato) dicots, monocots (rice, sorghum), a model plant (*Arabidopsis*) and a monocot (maize) and a dicot (soybean) with large genomes. We analysed the secondary structure of the PL domains with HHPred and the Chou Fasman Secondary Structure Prediction Server available on the ExPASy resource portal, for the 124 sequences subjected to MEME analysis and the associated consensus sequence.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The datasets supporting the conclusions of this article are either available in the DRYAD repository, doi:10.5061/dryad.dp20h, <http://dx.doi.org/10.5061/dryad.dp20h>, or in <https://www.ebi.ac.uk/ena>, <http://www.ebi.ac.uk/ena/data/view/LT555556-LT555568> for biological sequences [EMBL

accession numbers LT555556 to LT555568] or included within the article and its additional files.

Additional files

Additional file 1: Structure of the *Prunus* TNL gene *Ma* predicted with the predictors Fgenesh and GENESCAN using various specific gene-finding parameters. The first row shows the exact size of each exon (and associated domain in bp) of the *Ma* gene. The other rows show the exon predictions. NA not appropriate. (DOCX 48 kb)

Additional file 2: Exon and domain structures of all predicted TNL-related peptides. Exon sizes are shown above the colour shapes reflecting the different domains. (XLSX 115 kb)

Additional file 3: TNL prediction mapped in the *Prunus persica* genome v2.0. This Generic Feature Format Version 3 (GFF3) can be uploaded as a custom track directly in a local Genome browser or the online GDR Gbrowse, together with the genome v2.0. (GFF3 88 kb)

Additional file 4: Physical map of the TNL-related genes in the peach genome v1.0 and revision observed in v2.0. The map position of the 195 total TNL-related proteins is depicted within the eight scaffolds (chromosomes). Accession numbers are related to the genome glossary of the v1.0. Merged predicted peptides identified through corrected predictions are underlined and highlighted with brackets. The size of each scaffold (in Mb) and the respective number of TNL-related predicted peptides are shown under each scaffold. The *Ma* orthologue is indicated in red upper case. Unmapped genes in the v1.0 and mapped in the v2.0 are in green boxes. The red box shows re-localisation of TNL genes in the v2.0. Predicted peptides that are not indicated in the v2.0 are labelled with an asterisk. (PDF 167 kb)

Additional file 5: Secondary-structure prediction of the PL motif-1. The table shows the consensus sequence obtained from the alignment of the peach PL sequences, the 5 PLs of the *Ma* protein and the WW-domain consensus sequence obtained from *Vitis vinifera* sequences. H signifies a helix residue. Predicted stands for extended beta sheet (E) are surrounded by the tryptophans (in blue). (DOCX 17 kb)

Additional file 6: Transcribed PL sequences. Sequences of the transcribed genes ppa022336m, ppa000596m, ppa021732m, ppa024249m, ppa021703m, ppa023688m-ppa000585m, ppa026786m and ppa021808m for the peach 'Nemared' and plum 'Myrabi' (= 'P.2032') cultivars. (FAS 7 kb)

Additional file 7: TNL-related peptides obtained with BLASTp analysis using MEME motifs of the post-LRR domain in the Plant Genome Database website. Results are shown for *Populus trichocarpa*, *Vitis vinifera*, *Medicago truncatula*, *Arabidopsis thaliana*, *Solanum tuberosum*, *Glycine max*, *Oryza sativa*, *Prunus persica*, *Sorghum bicolor*, and *Zea mays*. Identified TNL-related sequences, unrelated TNL sequences and unknown sequences are highlighted in green, blue and yellow, respectively. (DOC 138 kb)

Additional file 8: Main functional resistance genes described in the literature. The table includes names, host plants, pathogens controlled and structural characteristics of the resistance genes. (XLSX 15 kb)

Abbreviations

CC: coiled-coil; ETI: effector-triggered immunity; GDR: genome database for rosaceae; NB-LRR: nucleotide-binding site leucine-rich repeat; PAMP: pathogen-associated molecular pattern; PGDB: plant genome database; PL: post-leucine-rich repeat; PTI: pamp-triggered immunity; TIR: toll/Interleukin1 receptor; TMV: tobacco mosaic virus.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CVG carried out the sequence analyses, experiments, designed and performed the manual re-annotation and analysed the physical and phylogenetic relationships. DE and CVG designed the study and wrote the manuscript. Both authors read and approved the final manuscript.

Authors' information

The authors belong to the Plant-Nematode interactions team, Institut Sophia Agrobiotech (ISA), INRA PACA, France, and have an expertise in plant resistance genes and plant nematology.

Acknowledgements

We would particularly wish to thank Emeline Deleury for help with sequence extraction and feedback. We thank Julia Truch and Simon Saucet for feedback on the manuscript and Ariane Fazari for sequencing. We thank the ISA bioinformatics group for orthoMCL analyses.

Funding

This work has been funded by the Plant-Nematode interactions team, Institut Sophia Agrobiotech (ISA), INRA PACA, France.

Author details

¹INRA, UMR 1355 Institut Sophia Agrobiotech, 06900 Sophia Antipolis, France. ²University Nice Sophia Antipolis, UMR 7254 Institut Sophia Agrobiotech, 06900 Sophia Antipolis, France. ³CNRS, UMR 7254 Institut Sophia Agrobiotech, 06900 Sophia Antipolis, France.

Received: 29 October 2015 Accepted: 22 April 2016

Published online: 30 April 2016

References

- Jones JDG, Dangl JL. The plant immune system. *Nature*. 2006;444(7117):323–9.
- Dangl JL, Jones JDG. Plant pathogens and integrated defence responses to infection. *Nature*. 2001;411(6839):826–33.
- Meyers BC, Kozik A, Griego A, Kuang HH, Michelmore RW. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell*. 2003;15(4):809–34.
- Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol Gen Genomics*. 2004;271(4):402–15.
- Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W, Martin F. Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Mol Biol*. 2008;66(6):619–36.
- Yang SH, Zhang XH, Yue JX, Tian DC, Chen JQ. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol Gen Genomics*. 2008;280(3):187–98.
- Yue JX, Meyers BC, Chen JQ, Tian DC, Yang SH. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytol*. 2012;193(4):1049–63.
- Tarr DE, Alexander HM. TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Res Notes*. 2009;2:197.
- Terefe-Ayana D, Kaufmann H, Linde M, Debener T. Evolution of the Rdr1 TNL-cluster in roses and other Rosaceae species. *BMC Genomics*. 2012;13.
- Wan HJ, Yuan W, Ye QJ, Wang RQ, Ruan MY, Li ZM, Zhou GZ, Yao ZP, Zhao J, Liu SJ, et al. Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogues in pepper: characterization, genetic variation, functional divergence and expression patterns. *BMC Genomics*. 2012;13.
- Hulbert SH, Webb CA, Smith SM, Sun Q. Resistance gene complexes: evolution and utilization. *Annu Rev Phytopathol*. 2001;39:285–312.
- Noel L, Moores TL, van der Biezen EA, Parniske M, Daniels MJ, Parker JE, Jones JDG. Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus of Arabidopsis. *Plant Cell*. 1999;11(11):2099–111.
- Paal J, Henselewski H, Muth J, Meksem K, Menendez CM, Salamini F, Ballvora A, Gebhardt C. Molecular cloning of the potato Gro1-4 gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. *Plant J*. 2004;38(2):285–97.
- Claverie M, Dirlewanger E, Bosselut N, Van Ghelder C, Voisin R, Kleinhenz M, Lafargue B, Abad P, Rosso MN, Chalhou B, et al. The *Ma* gene for complete-spectrum resistance to *Meloidogyne* species in *Prunus* is a TNL with a huge repeated C-terminal post-LRR region. *Plant Physiol*. 2011;156(2):779–92.
- Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG. The L6 gene for flax rust resistance is related to the Arabidopsis bacterial-resistance gene RPS2 and the tobacco viral resistance gene-N. *Plant Cell*. 1995;7(8):1195–206.
- Dinesh-Kumar SP, Tham WH, Baker BJ. Structure-function analysis of the tobacco mosaic virus resistance gene N. *Proc Natl Acad Sci U S A*. 2000;97(26):14789–94.
- Hinsch M, Staskawicz B. Identification of a new Arabidopsis disease resistance locus, RPS4, and cloning of the corresponding avirulence gene, avrRps4, from *Pseudomonas syringae* pv. pisi. *Mol Plant-Microbe Interact*. 1996;9(1):55–61.
- Ellis JG, Lawrence GJ, Luck JE, Dodds PN. Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell*. 1999;11(3):495–506.
- Mestre P, Baulcombe DC. Elicitor-mediated oligomerization of the tobacco N disease resistance protein. *Plant Cell*. 2006;18(2):491–501.
- Takken FLW, Govere A. How to build a pathogen detector: structural basis of NB-LRR function. *Curr Opin Plant Biol*. 2012;15(4):375–84.
- McHale L, Tan XP, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. *Genome Biol*. 2006;7(4):212.
- Belkadir Y, Subramaniam R, Dangl JL. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. *Curr Opin Plant Biol*. 2004;7(4):391–9.
- Dodds PN, Lawrence GJ, Ellis JG. Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. *Plant Cell*. 2001;13(1):163–78.
- Ellis JG, Dodds PN, Lawrence GJ. Flax rust resistance gene specificity is based on direct resistance-avirulence protein interactions. *Annu Rev Phytopathol*. 2007;45:289–306.
- Bryan GT, Wu KS, Farrall L, Jia YL, Hershey HP, McAdams SA, Faulk KN, Donaldson GK, Tarchini R, Valent B. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene Pi-ta. *Plant Cell*. 2000;12(11):2033–45.
- van der Hoorn RAL, Kamoun S. From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell*. 2008;20(8):2009–17.
- Collier SM, Moffett P. NB-LRRs work a "bait and switch" on pathogens. *Trends Plant Sci*. 2009;14(10):521–9.
- Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet*. 2010;11(8):539–48.
- Cesari S, Bernoux M, Moncuquet P, Kroj T, Dodds PN. A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. *Front Plant Sci*. 2014;5:606.
- Saucet SB, Ma Y, Sarris PF, Furzer OJ, Sohn KH, Jones JDG. Two linked pairs of Arabidopsis TNL resistance genes independently confer recognition of bacterial effector AvrRps4. *Nat Commun*. 2015, 6 doi:10.1038/ncomms7338
- Wirthmueller L, Zhang Y, Jones JDG, Parker JE. Nuclear accumulation of the Arabidopsis immune receptor RPS4 is necessary for triggering EDS1-dependent defense. *Curr Biol*. 2007;17(23):2023–9.
- Deslandes L, Olivier J, Theulieres F, Hirsch J, Feng DX, Bittner-Eddy P, Beynon J, Marco Y. Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive *RRS1-R* gene, a member of a novel family of resistance genes. *Proc Natl Acad Sci U S A*. 2002;99(4):2404–9.
- Noutoshi Y, Ito T, Seki M, Nakashita H, Yoshida S, Marco Y, Shirasu K, Shinozaki K. A single amino acid insertion in the WRKY domain of the Arabidopsis TIR-NBS-LRR-WRKY-type disease resistance protein SLH1 (sensitive to low humidity 1) causes activation of defense responses and hypersensitive cell death. *Plant J*. 2005;43(6):873–88.
- Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, et al. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet*. 2013;45(5):487–U447.
- Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol*. 2006;7 Suppl 1:S10.1–12.
- Zdobnov EM, Apweiler R. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 2001;17(9):847–8.
- Schorneck S, Ballvora A, Gurlebeck D, Peart J, Ganai M, Baker B, Bonas U, Lahaye T. The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. *Plant J*. 2004;37(1):46–60.
- Gassmann W, Hinsch ME, Staskawicz BJ. The Arabidopsis RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *Plant J*. 1999;20(3):265–77.

39. Jupe F, Pritchard L, Etherington GJ, MacKenzie K, Cock PJA, Wright F, Sharma SK, Bolser D, Bryan GJ, Jones JDG, et al. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics*. 2012;13.
40. Jia Y, Yuan Y, Zhang Y, Yang S, Zhang X. Extreme expansion of NBS-encoding genes in Rosaceae. *BMC Genet*. 2015;16.
41. Velasco R, Zharkikh A, Troglio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, FitzGerald LM, Vezzulli S, Reid J, et al. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE*. 2007;2(12):e1326.
42. Marone D, Russo MA, Laido G, De Leonardi AM, Mastrangelo AM. Plant Nucleotide Binding Site-Leucine-Rich Repeat (NBS-LRR) genes: active guardians in host defense responses. *Int J Mol Sci*. 2013;14(4):7302–26.
43. Deslandes L, Olivier J, Peeters N, Feng DX, Khounloham M, Boucher C, Somssich I, Genin S, Marco Y. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc Natl Acad Sci U S A*. 2003;100(13):8024–9.
44. Potter D, Eriksson T, Evans RC, Oh S, Smedmark JEE, Morgan DR, Kerr M, Robertson KR, Arsenault M, Dickinson TA, et al. Phylogeny and classification of Rosaceae. *Plant Syst Evol*. 2007;266(1–2):5–43.
45. Illa E, Sargent DJ, Girona EL, Bushakra J, Cestaro A, Crowhurst R, Pindo M, Cabrera A, van der Knaap E, Iezzoni A, et al. Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. *BMC Evol Biol*. 2011;11.
46. Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JDG. The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell*. 1996;84(3):451–9.
47. Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, et al. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science*. 1995;270(5243):1804–6.
48. Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM. The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *Plant Cell*. 1998;10(8):1307–19.
49. Cesari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, Rivas S, Alaux L, Kanzaki H, Okuyama Y, et al. The rice resistance protein pair RGA4/RGA5 recognizes the magnaporthe oryzae effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell*. 2013;25(4):1463–81.
50. Sudol M, Chen HI, Bougeret C, Einbond A, Bork P. Characterization of a novel protein-binding module - the WW domain. *Febs Lett*. 1995;369(1):67–71.
51. Luck JE, Lawrence GJ, Dodds PN, Shepherd KW, Ellis JG. Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. *Plant Cell*. 2000;12(8):1367–77.
52. Le Roux C, Huet G, Jauneau A, Camborde L, Tremousaygue D, Kraut A, Zhou B, Levaillant M, Adachi H, Yoshioka H, et al. A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell*. 2015;161(5):1074–88.
53. Sarris PF, Duxbury Z, Huh SU, Ma Y, Segonzac C, Sklenar J, Derbyshire P, Cevik V, Rallapalli G, Saucet SB, et al. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell*. 2015;161(5):1089–100.
54. Esmenjaud D, Voisin R, Van Ghelder C, Bosselut N, Lafargue B, Di Vito M, Dirlwanger E, Poessel JL, Kleinhentz M. Genetic dissection of resistance to root-knot nematodes *Meloidogyne* spp. in plum, peach, almond, and apricot from various segregating interspecific *Prunus* progenies. *Tree Genet Genomes*. 2009;5(2):279–89.
55. Van Ghelder C, Lafargue B, Dirlwanger E, Ouassa A, Voisin R, Polidori J, Kleinhentz M, Esmenjaud D. Characterization of the *RMja* gene for resistance to root-knot nematodes in almond: spectrum, location, and interest for *Prunus* breeding. *Tree Genet Genomes*. 2010;6(3):503–11.
56. Bendahmane A, Farnham G, Moffett P, Baulcombe DC. Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the Rx locus of potato. *Plant J*. 2002;32(2):195–204.
57. Whitham S, Dinesh Kumar SP, Choi D, Hehl R, Corr C, Baker B. The product of the tobacco mosaic virus resistance gene-N: similarity to toll and the interleukin-1 receptor. *Cell*. 1994;78(6):1101–15.
58. Kang YJ, Kim KH, Shim S, Yoon MY, Sun S, Kim MY, et al. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC Plant Biol*. 2012;12.
59. Ameline-Torregrosa C, Wang B-B, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB. Identification and characterization of nucleotide-binding site-Leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol*. 2008;146(1):5–21.
60. Shang J, Tao Y, Chen X, Zou Y, Lei C, Wang J, Li X, Zhao X, Zhang M, Lu Z. Identification of a New Rice Blast Resistance Gene, *Pid3*, by Genomewide Comparison of Paired Nucleotide-Binding Site-Leucine-Rich Repeat Genes and Their Pseudogene Alleles Between the Two Sequenced Rice Genomes. *Genetics*. 2009;182(4):1303–11.
61. Cheng X, Hang H, Zhao Y, Qian Y, Zhu S, Cheng B. A genomic analysis of disease-resistance genes encoding nucleotide binding sites in *Sorghum bicolor*. *Genet Mol Biol*. 2010;33(2):292–7.
62. Cheng Y, Li X, Jiang H, Ma W, Miao W, Yamada T, Zhang M. Systematic analysis and comparison of nucleotide-binding site disease resistance genes in maize. *Febs J*. 2012;279(13):2431–43.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



CHAPITRE 4

LES NBS-LRRs CHEZ LES CONIFERES ET TRAITs EVOLUTIFS DE CETTE FAMILLE MULTIGENIQUE CHEZ LES PLANTES TERRESTRES

Chapitre 4 : LES NBS-LRRs CHEZ LES CONIFERES ET TRAITS EVOLUTIFS DE CETTE FAMILLE MULTIGENIQUE CHEZ LES PLANTES TERRESTRES

4.1 Introduction

Au cours de ces vingt dernières années, de nombreuses études ont caractérisé la famille des NBS-LRRs chez divers génomes de plantes, essentiellement des Angiospermes, et plus récemment chez quelques organismes archaïques (algues vertes, Bryophytes, ...) (Shao *et al.*, 2016; Gao *et al.*, 2018). L'expertise acquise concernant l'annotation des TNLs à l'échelle du génome du pêcher a ouvert la voie à un élargissement de nos travaux à d'autres plantes pérennes. Ainsi, après avoir fait le point sur le domaine PL et ses TNLs associés chez les Angiospermes, nous avons choisi d'étudier les NBS-LRRs chez un groupe taxonomique très différent, prépondérant dans les écosystèmes forestiers et pour lequel très peu de données relatives à cette famille multigénique sont disponibles : les conifères (Gymnospermes). En effet, il s'agit d'un groupe important écologiquement et économiquement mais dont les données génomiques disponibles sont parcellaires du fait de la taille et de la complexité des génomes qu'il contient.

Pour ce faire, nous avons analysé les transcriptomes de référence de sept espèces de conifères. L'accès à ces données de qualité nous a permis d'identifier et d'annoter plus de 3800 séquences de NBS-LRRs. Cette nouvelle ressource revêt une grande importance pour les chercheurs travaillant sur la résistance des plantes de par la rareté des séquences disponibles au sein de ce groupe taxonomique.

Nous avons ainsi pu établir que le répertoire de NBS-LRRs chez toutes les espèces de conifères testées était large et varié, avec une prépondérance des TNLs et un nombre important de RNLs. Cette caractéristique également mentionnée chez les Rosaceae, nous a conduit à évaluer les effectifs de ces deux sous-familles chez divers génomes de plantes. Nous avons montré qu'une relation numérique existait entre RNLs et TNLs selon un ratio approximatif de 1 : 10. Ce résultat, s'ajoutant à de récentes études fonctionnelles (Qi *et al.*, 2018; Castel *et al.*, 2019), indique que les RNLs sont des partenaires privilégiés dans le déclenchement de la réponse immune associée aux TNLs. Par ailleurs, notre étude a révélé une grande diversité chez les RNLs constitués de quatre groupes dont un est surexprimé précocement en réponse au déficit en eau. Nous avons également noté qu'un déficit en eau prononcé chez les conifères avait pour effet de sous-réguler un grand nombre de NBS-LRRs et notamment de TNLs. Enfin, notre annotation fonctionnelle des TNLs a révélé une absence des motifs que nous avons

précédemment identifiés dans le domaine PL des Angiospermes. Par ailleurs, aucune autre signature C-terminale claire n'a pu être identifiée chez les TNLs de conifères. L'absence de domaines PL chez les conifères ainsi que chez des plantes archaïques indique que ce domaine est probablement apparu tôt chez les plantes à graines mais n'a connu une expansion adaptative que chez les Angiospermes et plus précisément chez les Dicotylédones (**Figure 20**).

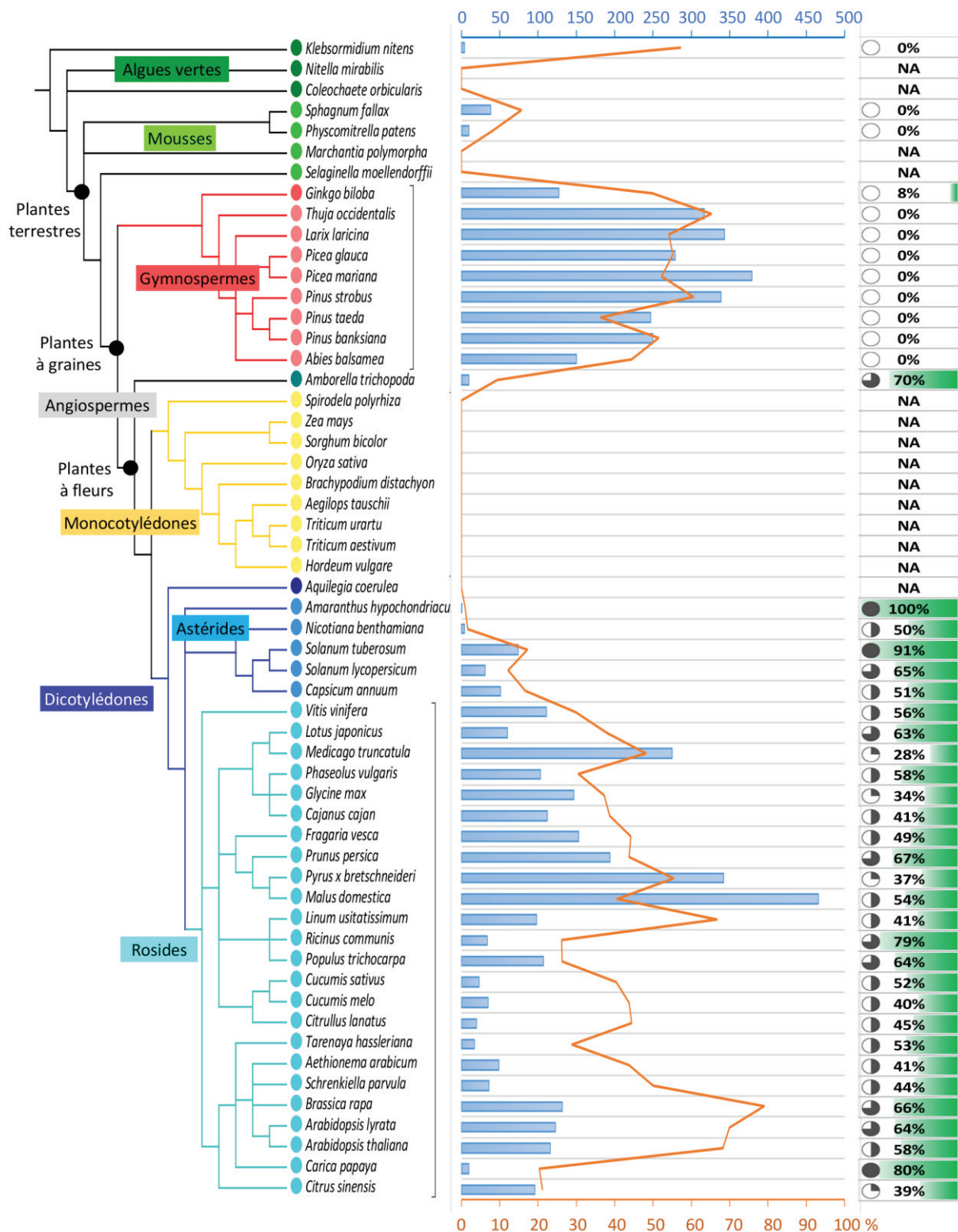


Figure 20 - Distribution des TNLs chez 56 génomes et transcriptomes du règne végétal.

L'histogramme en bleu représente les effectifs de TNLs. En orange, la part (%) que représentent les TNLs par rapport au nombre total de NBS-LRRs par espèce. Le panneau latéral droit illustre le pourcentage de TNLs disposant d'un domaine PL. Les valeurs et références associées sont indiquées en Annexe 4. La relation phylogénétique entre les différentes espèces a été établie avec l'outil 'Taxonomy Common Tree' du NCBI, disponible à l'adresse : <https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>.

4.2 Article publié dans Scientific Reports

Van Ghelder, C. Parent G.J., Rigault, P., Prunier, J., Giguère, I., Caron, S., Stival Sena, J., Deslauriers, A., Bousquet, J., Esmenjaud, D. and MacKay, J. (2019) **The large repertoire of conifer NLR resistance genes includes drought responsive and highly diversified RNLs.** *Scientific Reports* **9** :11614. doi: 10.1038/s41598-019-47950-7.

NB : Dans cet article, pour répondre à la demande d'un des reviewers, les termes « NBS-LRR » et « domaine NBS » ont été remplacés, respectivement, par « NLR » et « domaine NB-ARC ».

Les données additionnelles liées à cet article sont disponibles à l'adresse suivante :

<https://www.nature.com/articles/s41598-019-47950-7#Sec22>

OPEN **The large repertoire of conifer NLR resistance genes includes drought responsive and highly diversified RNLs**

Received: 15 January 2019

Accepted: 26 July 2019

Published online: 12 August 2019

 Cyril Van Ghelder^{1,2}, Geneviève J. Parent^{1,2,3}, Philippe Rigault^{4,5}, Julien Prunier³, Isabelle Giguère³, Sébastien Caron⁴, Juliana Stival Sena⁶, Annie Deslauriers⁷, Jean Bousquet^{3,8}, Daniel Esmenjaud¹ & John MacKay^{2,3}

The NLRs or NBS-LRRs (nucleotide-binding, leucine-rich-repeat) form the largest resistance gene family in plants, with lineage-specific contingents of TNL, CNL and RNL subfamilies and a central role in resilience to stress. The origin, evolution and distribution of NLR sequences has been unclear owing in part to the variable size and diversity of the RNL subfamily and a lack of data in Gymnosperms. We developed, searched and annotated transcriptomes assemblies of seven conifers and identified a resource of 3816 expressed NLR sequences. Our analyses encompassed sequences data spanning the major groups of land plants and determinations of NLR transcripts levels in response to drought in white spruce. We showed that conifers have among the most diverse and numerous RNLs in tested land plants. We report an evolutionary swap in the formation of RNLs, which emerged from the fusion of an RPW8 domain to a NB-ARC domain of CNL. We uncovered a quantitative relationship between RNLs and TNLs across all land plants investigated, with an average ratio of 1:10. The conifer RNL repertoire harbours four distinct groups, with two that differ from Angiosperms, one of which contained several upregulated sequences in response to drought while the majority of responsive NLRs are downregulated.

The NLR (nucleotide-binding, leucine-rich repeat) gene family is the main group of cytoplasmic receptors involved in the recognition of specific pathogens as part of effector-triggered immunity in plants¹. NLRs may confer resistance to biotic aggressors including viruses, bacteria, fungi, nematodes and insects^{2–6}, and abiotic stresses such as drought in some cases⁷, and present a wide range of variation across flowering plants⁸. This fundamental role of NLR proteins in resilience to biotic and abiotic stresses has encouraged genome-scale studies of their evolution, primarily in Angiosperms^{9–14}.

The NLR protein family may have originated in green algae¹⁵ and was well-defined early in the land plant lineage¹⁶. It has evolved through gene duplication, domain fusions, and acquisition of amino-acid sequences that enable the detection of pathogens, protein dimerization, the activation of immune responses and possibly the triggering of a type of programmed cell death known as hypersensitive-like response¹⁷. The central nucleotide binding NB-ARC domain binds ADP or ATP molecules, which induces a protein conformation change and the switch from an 'OFF' to an 'ON' state. The NB-ARC domain, which is often used to perform phylogenetic studies, contains highly conserved motifs (Ploop, kinase, RNBS, GLPL, MHD) involved in intra- and extra-molecular

¹INRA, Université Côte d'Azur, CNRS, ISA, 400 route des Chappes, BP167, 06903, Sophia Antipolis, France.

²Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK. ³Forest Research Centre and Institute for Systems and Integrative Biology, Université Laval, 1030 rue de la Médecine, Québec, QC, G1V 0A6, Canada. ⁴Gydlé Inc., 1135 Grande Allée Ouest Suite 220, Québec, QC, G1S 1E7, Canada. ⁵Center for Organismal Studies (COS), University of Heidelberg, Im Neuenheimer Feld 345, 69120, Heidelberg, Germany.

⁶Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Québec, G1V 4C7, Canada.

⁷Département des Sciences Fondamentales, Université du Québec à Chicoutimi, 555 Boulevard de l'Université, Chicoutimi, QC, G7H2B1, Canada. ⁸Canada Research Chair in Forest Genomics, Université Laval, 1030 rue de la Médecine, Québec, QC, G1V 0A6, Canada. Correspondence and requests for materials should be addressed to C.V.G. (email: cyril.van-ghelder@inra.fr)

interactions^{18,19}. Among these motifs, the RNBS-A, RNBS-D and the MHD motifs show some variations that are correlated to the N-terminal domain identity and can be used to classify the sequences into subfamilies. The C-terminal LRR domain mainly supervises the recognition function through interactions with effectors or molecules modified by effectors through residues in its horseshoe 3D-structure²⁰.

The conserved Toll interleukin-1 receptor (TIR) domain at the N-terminus defines the large TIR-NB-ARC-LRR (TNLs) subfamily²¹, many members of which display a C-terminal post-LRR domain (PL) of unknown function²². The two other subfamilies are made up of non-TIR containing NLRs with a diversity of N-terminal sequences and structures: (1) the CC-NB-ARC-LRR (CNL) protein subfamily, which displays a coiled-coil structure in their N-terminus⁹ and (2) the RPW8-NB-ARC-LRR (RNL), which shares homology with RPW8-only proteins (RESISTANCE TO POWDERY MILDEW8) from *Arabidopsis thaliana*²³. In addition, partial NLRs such as RBA1 (TIR-only protein), TN2 (TIR-NB-ARC-only), RPW8.1 and RPW8.2 (RPW8-only) were recently shown to confer resistance to pathogens^{23–25}; therefore, we considered truncated versions in our study of the NLR gene family.

The occurrence of the different NLR subfamilies is variable among the major plant taxonomic groups. For instance, Monocots have many CNL encoding genes and lack the TNL subfamily²¹. The RNL subfamily is found in primitive plants such as *Physcomitrella patens* but has few or no members in other plant lineages, except in the Rosaceae, where it underwent a major expansion²⁶. The RNL subfamily have been classified as a sister clade to the CNLs and TNLs¹¹, but its evolution remains poorly resolved owing to the variability in both CNL and RNL subfamilies and its sequence motifs remain poorly defined. In Angiosperms, RNLs are subdivided into two identified subclades²⁷ based on the homology of their RPW8 domain to the one of (1) the N-required gene 1 (NRG1), co-responsible of the Tobacco Mosaic Virus resistance in *Nicotiana benthamiana*²⁸, or (2) the activated disease resistance gene 1 (ADR1), involved in both pathogen resistance and drought tolerance in *Arabidopsis thaliana*⁷. Functionally, RNL genes function as helper proteins acting downstream of sensor NLRs^{27,29–31}.

The recent analysis of *Pinus taeda* L. and *Picea abies* (L.) H. Karst. genome drafts recently suggested the occurrence of numerous RNL sequences²⁶ but a comprehensive analysis of conifer NLRs is lacking. A few conifer NLR sequences have been described³² and linked to resistance against pathogens such as *Cronartium ribicola*, which is responsible for the white pine blister rust^{33,34}. Conifers possess very large genomes sizes and abundant pseudogenes, which hinder genome assembly and analysis of large gene families, but transcriptome sequencing provides an effective alternative to rapidly acquire large datasets of expressed genes^{35–37}.

The overall goal of this study was to investigate the makeup of NLRs in conifers with the aim of improving our understanding of the RNL subfamily in plants. We introduced 3816 expressed NLR sequences in seven conifers in the *Pinaceae* and the *Cupressaceae* families. Our specific objectives were to: (1) identify the repertoire of expressed NLR sequences through the analysis of seven conifer transcriptomes; (2) classify the poorly known conifer NLRs based on amino acid domains and motif analyses; (3) investigate the diversity and the drought responsiveness of conifer RNLs, testing whether or not conifers form a single group external to the Angiosperm groups as recently proposed²⁷. We found that conifer RNLs presented a level of sequence diversity that is unparalleled in tested Angiosperms and possibly in other plant taxa. Our results also suggest that they play a key role in drought response.

Results

NLRs are abundant and diverse in the *Pinales*. We developed a dataset of 3816 expressed NLR sequences (Supplementary Data S1) by analysing the transcriptomes of seven conifer species in five genera of the *Pinaceae* and *Cupressaceae*. The sequences were identified using the Pfam database (31.0) in the predicted amino acid sequence datasets obtained from the assembled mRNA sequence. The expressed sequences contained at least one of the canonical NB-ARC, RPW8, or TIR domains while CC-only and LRR-only were excluded as they may be mispredicted or not specific to NLRs. We filtered the full dataset to eliminate potentially redundant NLR peptides of a same encoding-gene. The number of expressed NLR genes thus identified ranged from 338 in *Abies balsamea* to 725 in *Picea mariana* (Table 1), in line with the total of 679 NLR genes identified in the genome of *P. taeda*¹⁵. The total number of distinct NLR transcripts represented 0.73% to 1.35% of the assembled transcriptome depending on the species, which are substantial proportions of such genes when comparing with other Angiosperm transcriptomes (Supplementary Table S1).

Domain and motif specificities. The predicted NLR proteins followed the same domain organisation in conifers as described in Angiosperms, which included the differential N-terminal domains TIR, CC or RPW8, a conserved central NB-ARC domain, and a polymorphic C-terminal domain (LRR) (Supplementary Data S2). The TIR domain of conifer and Angiosperm TNLs was generally conserved, unlike the non-TIR N-terminus of the other NLRs, which was highly polymorphic and mainly made up of conformational domains (coiled-coil). The conifer TIR domain sequences were similar to those observed in Angiosperms and were characterised by a site displaying hydroxyl groups (poly-serine or threonine) close to the initial methionine, followed by the TIR-1 to TIR-5 motifs found in Angiosperms (Fig. 1). The TIR-1, 2 and 5 corresponded to the human TLR boxes 1 to 3, respectively. Within the NB-ARC domain, highly conserved sequence signatures and variations were observed in conifers as found in Dicots. The conserved elements included the P-loop, Kinase 2, Kinase 3, RNBS-C and GLPL motifs, which are similar in all NLR subfamilies, and the RNBS-A, RNBS-D and MHD, which were specific to each of the NLR subfamilies (Fig. 1).

We report two specific signatures in the RNBS-D (CFLDLGxFP) and MHD (QHD) motifs that enabled an accurate detection of RNL sequences in conifer sequences. Among these motifs that have been previously identified in Angiosperms³⁸, the RNBS-D motif was particularly useful to discriminate between different NLR sequences. It harboured conserved aspartic acid, glycine and cysteine residues that characterise each of the NLR forms (Fig. 1). We found that the QHD motif exclusively contained a glutamine (Q1) residue before the conserved

	<i>Abies balsamea</i>	<i>Larix laricina</i>	<i>Picea glauca</i>	<i>Picea mariana</i>	<i>Pinus banksiana</i>	<i>Pinus strobus</i>	<i>Thuja occidentalis</i>	Dist [†] (%)
Number of transcripts	46178	50712	37491	58751	47473	45447	38767	
TOTAL NLR	338	633	506	725	486	560	486	
Ratios	0.73%	1.25%	1.35%	1.23%	1.02%	1.23%	1.25%	
Non-TIR-NLR								
CNL								
NB _{CNL} -(LRR)	65	146	115	166	82	103	51	22%
Ratios	0.14%	0.29%	0.31%	0.28%	0.17%	0.23%	0.13%	
CNL2								
NB _{CNL2} -(LRR)	47	18	17	17	36	21	34	6%
Ratios	0.10%	0.04%	0.05%	0.03%	0.08%	0.05%	0.09%	
RNL								
RPW8	5	5	6	13	8	12	3	9%
NB _{RNL} -(LRR)	11	8	6	10	21	12	13	
RPW8-NB _{RNL} -(LRR)	16	33	19	16	26	16	15	
Total RNL-related	32	46	31	43	55	40	31	
Ratios	0.07%	0.09%	0.08%	0.07%	0.12%	0.09%	0.08%	
TIR-NLR								
TNL								
TIR	43	100	121	144	94	89	72	63%
NB _{TNL} -(LRR)	52	97	99	128	104	122	91	
TIR-NB _{TNL} -(LRR)	55	143	58	105	52	126	154	
TIR-LRR	0	3	1	2	0	2	0	
Total TNL-related	150	343	279	379	250	339	317	
Ratios	0.32%	0.68%	0.74%	0.65%	0.53%	0.75%	0.82%	
Atypical NLR	15	26	8	14	7	14	14	
Undetermined NLR	29	54	56	106	56	43	39	

Table 1. Distribution of NLRs and their subfamily assignment in seven conifer transcriptomes. [†]The distribution is calculated using the total sequence number of each subfamilies in proportion of the total of assigned sequences (CNL, CNL2, RNL and TNL).

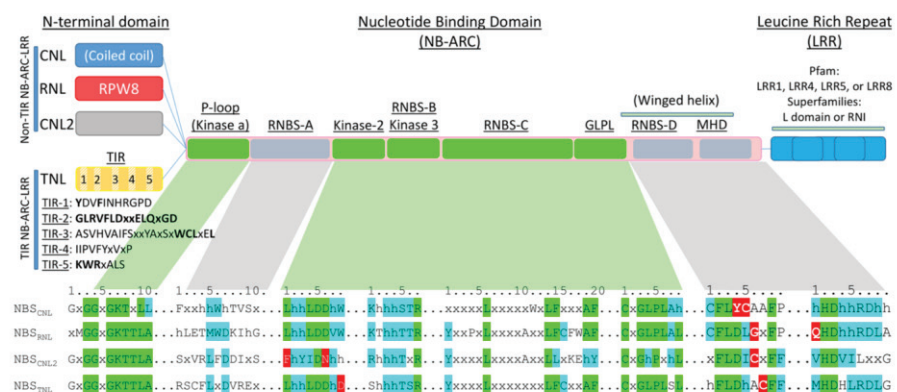


Figure 1. Schematic representation of the four subfamilies of NLRs in conifers. The detailed motifs of the NB-ARC domain are split into conserved motifs through NLR subfamilies (green) and discriminative motifs (grey). Pfam and Superfamilies entry identifiers are indicated above the green bar. Discriminative amino acids are highlighted in red, 'h' indicates various hydrophobic residue.

histidine (H2) and aspartic acid (D3) in the conifer RNL sequences, in contrast to the conventional methionine (M1) or other hydrophobic residues found in CNL and TNL proteins. We also identified an unconventional non-TIR NLR subfamily (CNL2), which had a unique combination of features in the NB-ARC domain, including a conserved cysteine in the sixth position (C6) after the conserved aspartic acid (D4) in the RNBS-D motif, a VHD motif and a polymorphic RNBS-A motif (Table 1, Fig. 1). A blastp analysis based on the RNBS-D motif only identified sporadic NB-ARC_{CNL2} sequences in recent Angiosperms.

We scanned the conifer sequences for conserved C-terminal signatures downstream of the LRR domain but none were clearly identified, in contrast to Dicot TNLs where post-LRR (PL) domains are ubiquitous. None of the known Dicot PL motifs were detected in the entire dataset of predicted peptides by using blastp or FIMO analysis. We also searched in datasets from ancient plants among Embryophytes (*P. patens*, *Sphagnum fallax*), *Ginkgo biloba*, another conifer (*P. abies*), and an early-diverged Angiosperm (*Amborella trichopoda*), and only found complete PL motifs in *A. trichopoda* and a few partial ones in *G. biloba*. These observations suggest a partial acquisition early in seed plant evolution, and a potential adaptive expansion in early flowering plants (Supplementary Fig. S1).

Conifer sequences are broadly represented in all NLR subfamilies. We assigned each of the transcripts to specific subgroups CNL, CNL2, RNL or TNL based on N-terminal sequences and specific motifs in the NB-ARC domain when possible, considering only the sequences containing *bona fide* NB-ARC domains. The TNL sequences were most abundant in all of the species and represented 63% of the total (Table 1), except in *A. balsamea*, which harboured as many CNLs as TNLs. The respective proportions of the different subfamilies were similar in the two *Picea* species and *Larix laricina* (χ^2 , $P > 0.999$), the proportion of TNL sequences was slightly higher in *Thuja occidentalis* and *Pinus strobus* and lower in *A. balsamea*, while RNL sequences were most abundant in *Pinus banksiana* (0.12% of all the transcripts) (Table 1).

Retracing the origin and distribution of RNL identifies a potential link to TNL evolution. The analysis of NB-ARC motifs is crucial to assign sequences to particular subfamilies. We reliably identified motifs in the NB-ARC_{RNL} domain including specific signatures that were absent from all of the NB-ARC_{TNL} or NB-ARC_{CNL} sequences, owing to the large and relatively uniform numbers of RNL sequences in conifers (Table 1). We used these results on RNL motifs to compare conifers and ancient plants to gain insights into the formation of RNL, and to investigate RNL sequence contingent sizes across land plants, aiming to explain differences observed in the seed plant phylogeny.

In early land plants, we identified two, four and ten putative RNL sequences in *Marchantia polymorpha*, *P. patens* and *S. fallax*, respectively, by using RPW8 and NB-ARC signatures. Although we had detected RPW8 domains in these sequences, their adjacent NB-ARC domains did not possess the NB-ARC_{RNL} motifs and clearly displayed CNL signatures in the RNBS-A, RNBS-D and MHD motifs (Supplementary Fig. S2). Among all the plant proteomes tested, this peculiarity was only found in the early land plants. These same species also harboured an RPW8-containing sequence associated with other domains such as Ubox/Arm, PLAC8 and PKinase (Mapoly0115s0019.1.p, Pp3c12_21520V3.1.p, Sphfalx0341s0009.2.p). The data showed that the RPW8 domain fused to a core NB-ARC_{CNL} or connected to other domains already existing in early land plants.

The abundance of RNLs was then investigated in 42 plant proteomes distributed among diverse taxonomic groups of land plants in addition to those of the seven conifers (Fig. 2 and Supplementary Table S2). Zhong and Cheng (2016) identified the expansive RNL subfamily in several of the *Rosaceae* using blastp and also suggested that the *Pinus* and *Picea* had numerous RNLs²⁶. Using the same blastp procedure, we also found large RNL contingents in both *Pinus* and *Picea* as well as in the two other groups of the Pinaceae family, namely Abietoideae and Laricoideae, and in the Cupressaceae family (Fig. 2a). We also detected thirteen RNLs in the early-diverged Gymnosperm *G. biloba* (Fig. 2a and Supplementary Table S2), suggesting that significant RNL contingents may be a common feature of Gymnosperms.

The data showed large variations across the land plants ranging from a single RNL sequence in Monocots to more than 30 in some of the Angiosperms without a clear pattern related to phylogeny. We retrieved TNL sequences from the same 42 plant proteomes either from prior surveys of NLR sequences or by using blastp searches (Fig. 2a and Supplementary Table S2) and performed a regression analysis between the numbers of RNLs and TNLs among the major groups of land plants. A significant relationship was uncovered, showing a 1:10 ratio (RNL:TNL; Fig. 2b). The overall regression model was significant, $F(1, 47) = 130.36$, $P = 3.75e-15$, $R^2 = 0.73$ (Fig. 2b). A similar analysis between RNLs and CNLs failed to detect any quantitative relationship (Fig. 2c).

Conifers present a high diversity in RNL gene sequences. We tested whether the large number of conifer RNLs was associated with increased sequence diversity, which could represent an ancient sub-group expansion. We extracted and trimmed 185 full-length RPW8 domains from the seven conifer sequences to identify potential subgroups. We only retrieved RPW8 domains from RPW8-only and RNL peptides. Following the approach described by Zhong & Cheng (2016)²⁶, we produced an unrooted tree including all of the full-length conifer RPW8 domains and four Angiosperm reference RPW8 domains to delineate the diversity of conifer RPW8 and assign sequences to specific groups. The resulting tree contained four distinct groups of unequal size (Fig. 3a). Two of the groups clustered with the Angiosperm sequences ADR1 (Group 2) or with the weakly supported NRG1 group (Group 4). The tree contained two other large groups (1 and 3) entirely made up of conifer sequences (Fig. 3a). Each of the seven species investigated here appeared to be homogenous in diversity among the four conifer groups. Considering the group 1, the *Thuja* sequences were entirely contained in subgroup 1b, suggesting an ancient divergence followed by recent duplications specific to this group of sequences (1) in the Cupressaceae (Fig. 3a). The group 2 represented by ADR1 had the most uniform representation, with three sequences on average per conifer species (ranging from two to five in *T. occidentalis* and *P. banksiana*, respectively).

Each of the groups was submitted to a clustering filter, which showed strong sequence clustering in groups 1 and 3 using a cut-off value of 80% identity, further suggesting multiple duplications for some of the sequences (Supplementary Fig. S3). The tree topology obtained with the representative sequences by using an identity-based clustering (cut-off of 60%) did not differ from the groups delimited with the full set of sequences (Supplementary Fig. S4). We carried out a motif analysis of the RPW8 domain in the four groups separately and found increasing

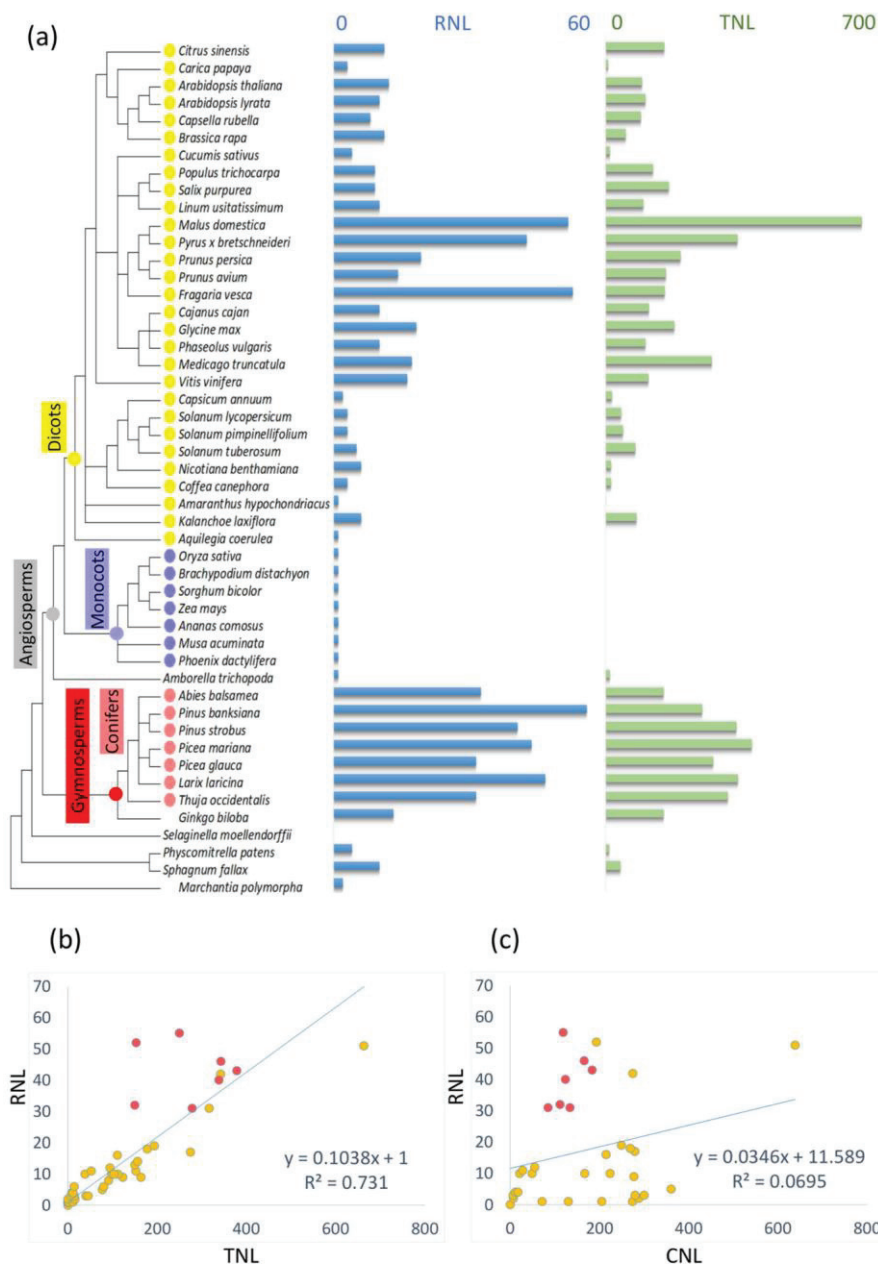


Figure 2. Distribution of the RNL and TNL genes in 49 land plants. (a) RNL- and TNL-related sequence numbers are mapped to the species tree created from the NCBI Common Taxonomy Tree (<https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wcmt.cgi>). Abscissa axis represents the protein numbers situated above the diagram. (b) RNL versus TNL numbers. (c) RNL versus CNL numbers. Light red dots represent the conifer species used in this study. The linear regression equation and the R^2 coefficient are indicated.

polymorphism in the N and C-termini from groups 1 to 3 and a conserved central section in the RPW8 domain (Fig. 3b and supplementary Table S3). Some motifs show signs of recombination (f), loss (a, e) or *de novo* formation (b, h, i). The group 4 appeared to be more diverged compared to groups 1 to 3 (Fig. 3b).

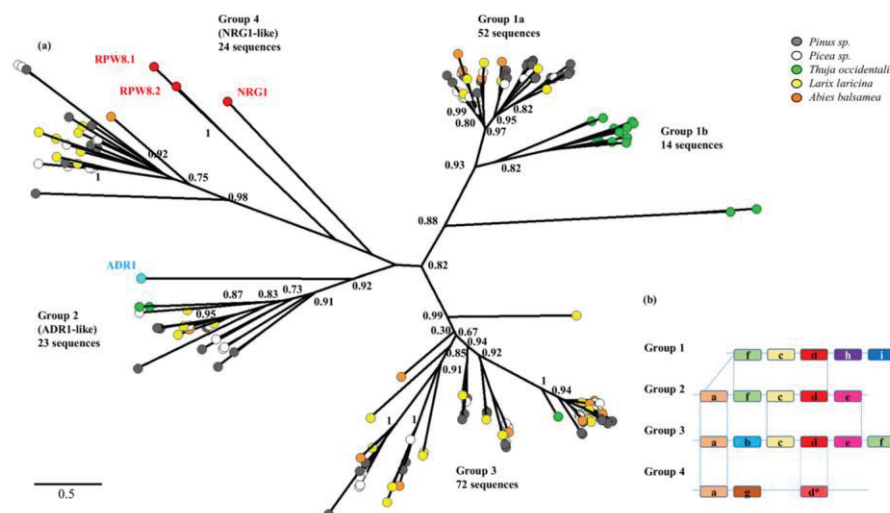


Figure 3. Unrooted ML phylogenetic tree and motif analysis for the RPW8 domain. (a) 185 RPW8 domains identified in conifers in this study together with RPW8 domains of ADR1 (blue dot), RPW8.1, RPW8.2 and NRG1 (red dots) were selected to build the ML phylogenetic tree. The number of conifer sequences contained in each of the groups are indicated below group numbers. For clarity, only main support values >0.70 are displayed (full data available in Supplementary Fig. S6). (b) Motif pattern diversification within the RPW8 domain associated with each group. The details of the motifs are available in Supplementary Table S3.

Next, we examined the conifer diversity of RNLs in view of the RNL land plant sequences. We retrieved the RPW8 domain from other plants (counted sequences in Fig. 2), together with *Picea abies* peptide sequences obtained from an independent dataset³⁹, and applied an identity-based clustering (cut-off of 60%) on all of the sequences to examine the extent of RPW8 domain diversity. We thus obtained a final set of 197 RPW8 sequences to build an expanded unrooted tree. The RNL sequences generally clustered according to their taxonomic clades within each of the groups (Fig. 4). The Gymnosperm sequences *P. abies* and *G. biloba* clustered together with the conifer sequences reported in this study (groups 1–4). All of the Angiosperm sequences clustered into groups 2 and 4 (Fig. 4). All the tested Monocot sequences only clustered in the ADR1 group (2). Most of the sequences of the early land plants, presumably the ancestors of RNLs, were located in NRG1 group (4). A few sequences of *S. fallax* also clustered in the Gymnosperm group 3, indicating a probable onset of RNL diversification. The phylogenetic tree also highlighted some particular sequences that underwent large expansions in conifers (group 3) or Rosaceae (group 4) (Fig. 4).

NLRs are differentially expressed under drought conditions. We investigated the drought responsiveness of NLR genes by RNA-seq analysis of samples from an independent study in *P. glauca* by Stival Sena *et al.*⁴⁰. The study monitored plants of three unrelated genotypes that were well-watered (controls) or not watered at all (drought treatment) over a 22-day period, with sampling time points at 0, 7, 14, 18, 22 days. A principal component analysis of the entire dataset of foliage transcript levels showed that the drought treatment was the main factor that explained 53% of the observed variance (Supplementary Fig. S5) and supported samples groupings by treatment and genotype. The data showed major effects of the drought treatment after 18 and 22 days, independently of the clone used, in addition to smaller and more variable effects between clones at 14 days (Supplementary Fig. S5).

We identified 119 differentially expressed (DE) NLR genes (adjusted p -value <0.05) in drought-treated plants compared to well-watered controls across the NLR repertoire, independently of the genotype. The major changes in NLR expression were observed after 14 days of treatment and increased at 18 and 22 days, thus following the physiological trends in water potential (Fig. 5). The data showed that a large majority ($>87\%$) of the DE NLR genes were downregulated, from slightly (the TNL, PG_019785_T.1; $\text{Log}_2\text{FC} = -0.25$) to strongly downregulated (the CNL, PG_022220_T.1; $\text{Log}_2\text{FC} < -5$) (Fig. 5b and Supplementary Table S4). We found that all NLR subfamilies were represented in the pool of DE sequences. A clustering analysis based on expression patterns identified 14 groups including six expression groups that only contained members of one subfamily (Fig. 5 and Supplementary Table S4).

While the DE CNL and TNL sequences were almost exclusively downregulated, 62% of the RNLs were upregulated. In total, 8 out of all the 15 upregulated NLRs belonged to the RNL subfamily (Fig. 5).

Gymnosperm-specific RNLs have drought responsive expression. We identified 13 drought responsive RNL genes, including 8 upregulated (up to $\text{Log}_2\text{FC} > 8$) and 5 downregulated sequences (up to

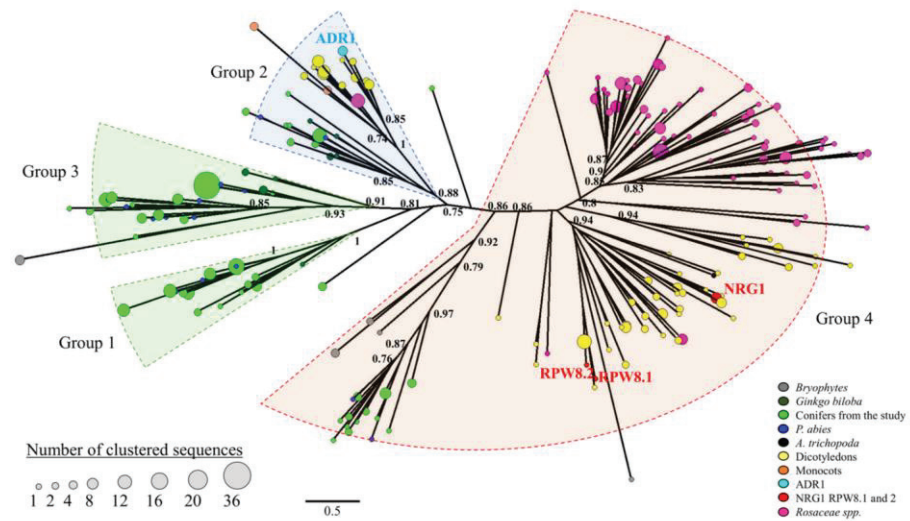


Figure 4. Unrooted ML phylogenetic tree of the RPW8 domain in land plants. 197 representative RPW8 domains retrieved from 50 land plants (49 species used in the Fig. 3 and *Picea abies*) were used for the phylogenetic analysis. The size of the circles reflects the number of sequences that are clustered (sequence identity cut-off of 0.60) with the representative sequence. For clarity, only main support values >0.70 are displayed (full data available in Supplementary Fig. S7). Colours are used to reveal taxonomic groups and reference sequences.

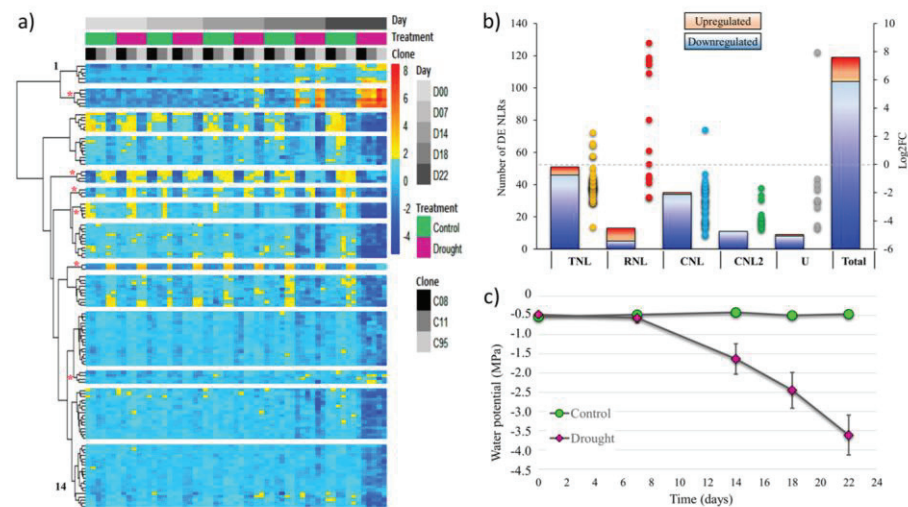


Figure 5. Expression patterns and distribution of drought responsive NLR genes in *Picea glauca*. (a) Heatmap illustrating the 14 expression groups formed of the 119 NLR sequences with drought responsive transcript levels for each genotype, treatment and sampling day. Red asterisks indicate expression groups with a single subfamily of NLR. (b) Number of drought responsive NLR genes and Log₂ Fold Change observed in the four subfamilies and undetermined (U). (c) Measurement of water potential in needles for control and water-stressed plants at five sampling times. Each point is the average of 12 biological replicates (3 *P. glauca* genotypes with 4 replicates each) with standard deviation (adapted from Stival Sena *et al.*⁴⁰).

Log₂FC < -2) (Fig. 6 and Supplementary Figure S9). Examining the set of five most upregulated RNL sequences, we observed a steady increase in transcript levels starting from the day 7, which was before any noticeable change was observed in the leaf water potential (Fig. 5). All of the 13 drought responsive RNLS belonged to the group 1

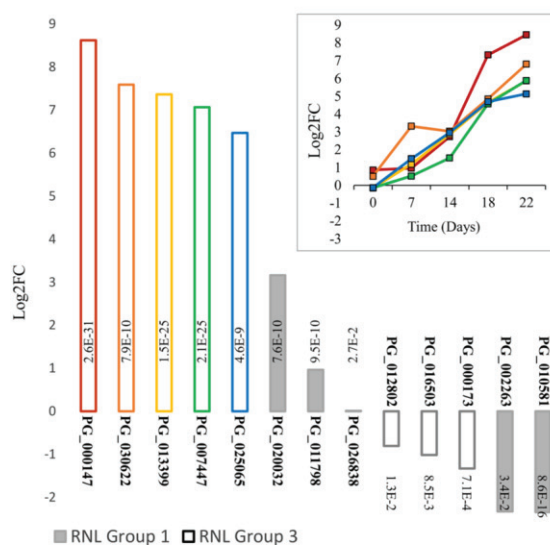


Figure 6. Expression of the 13 drought responsive RNLs and their RNL group assignment. The main figure shows transcript levels (Log2 fold change) after 22 days without water. Adjusted p-values are the result of the statistical tests for differential expression between drought-treated and control groups. Time course of expression profiles (Log2 fold change) are shown for the five most upregulated RNLs (top).

or 3 of sequences that are specific to Gymnosperms. None of them clustered with the ADR1 or the NRG1 group of sequences (group 2 and 4) that are in common with Angiosperms. This observation suggests that group 1 and 3 likely harbour gene sequences specifically involved in the response to drought in conifers (Fig. 6).

A genomic cluster of group 1 RNL sequences was identified based on the *P. glauca* linkage map⁴¹. The sequences were identified based on similarity (99% identity) between the full set of *P. glauca* RNL sequences described here and the sequences that were positioned on the linkage map⁴¹. Many of the RNL sequences did not produce any matches but four RNL sequences with matches were localised within 0.4 centi Morgans (cM) of each other (positions 3.86 to 4.18) on linkage group 6. Their patterns of expression ranged from downregulated (PG_010581_T.1), upregulated (PG_011798_T.1) to not differentially expressed genes (PG_012538_T.1; PG_008655_T.1) (Supplementary Table S5).

Discussion

Most NLRs are constitutively expressed in healthy tissues and some are stress responsive in plants²⁰. We produced a transcriptome-wide repertoire of NLR sequences from seven species in the Pinaceae and Cupressaceae families by RNA sequencing of several tissues and species in normal and stressed conditions. We found a large and varied NLR gene arsenal in these conifers compared to Angiosperms that have reduced or even eliminated either TNL or RNL subfamilies. We discuss interspecific variations and evolution in non-TIR NLRs including their ratio to TNLs, which may depend on their involvement in successful stress responses, resulting in their selection followed by expansion through duplications in some cases.

Motif usage in conifers and implications in NLR evolution. Extensive studies on NB-ARC domains have highlighted their motif flexibility involving the reorganisation of pre-existing building blocks that can be observed at a domain or motif level¹⁶. In our dataset, we observed conserved motifs (P-loop, Kinase 2/3, and GLPL) involved in crucial ADP/ATP binding and shifting⁴². These motifs are conserved in all NLR subfamilies and in NB-ARC of prokaryotes, which do not display upstream N-terminal domains¹⁶. Thus, these motifs are not influenced by the N-terminal domain identity. The over-representation of RNL sequences in several conifers allowed us to define reliable NB-ARC_{RNL} signatures. We showed that RNL proteins strictly required a glutamine, first residue of the MHD motif in Fig. 1 (Q1) in place of the methionine (M1) (i.e. QHD). This glutamine residue was first identified in *Arabidopsis* ADR1 homologs⁴³ and probably possesses a functional role through its ability to create hydrogen bonds possibly involved in ATP processing. Indeed, mutations on H2 or D3 residues in the MHD motif (Fig. 1) favour ATP binding leading to the auto-activation of NLR and cell death^{44,45}. The shift in sensitivity between the ON/OFF states is crucial to confer a finite lifetime to active forms and reliably unlock OFF states initiating safe cell death⁴⁶. The RNBS-D motif is shaped according to the N-terminal domain identity but its exact function remains unclear¹⁹. By contrast to other NLR subfamilies, RNLs did not harbour a conserved cysteine (C5, C6 or C7) but a conserved glycine in the sixth position (G6) of the RNBS-D motif (Fig. 1). The ability of cysteine residues to form disulphide bonds may play a role in intramolecular interactions whereas the glycine in RNLs could reduce the steric hindrance. Our motif analysis showed that CNL2 sequences display polymorphic

motifs and an intermediate RNBS-D form that resembles that of the archaic kinase-NB-ARC-LRR in *P. patens*¹⁰. These findings and the near absence of CNL2 sequences in plants other than conifers suggest that CNL2 could represent an intermediate form in the evolutionary swaps between NLR subfamilies.

Diversity and drought responsiveness of RNL genes. The data reported here showed that conifer RNLs include sequences that are distinct from those previously described in addition to members of the well-known ADR1 and NRG1 groups^{26,27}. The presence of ADR1-like sequences among all flowering plants and the absence of NRG1-like representatives in some taxonomic groups reinforced the previously confused state of knowledge^{27,43}. We showed that conifer RNL genes split into four groups, two of which correspond to the only known Angiosperm groups (groups 2 and 4) while the two others (groups 1 and 3) are Gymnosperm-specific (Fig. 4, 5). Our results are at odds with the suggestion of a single independent Gymnosperm group²⁷ and places the NRG1-ADR1 differentiation before the divergence between Angiosperms and Gymnosperms, at more than 300 Myr ago⁴⁷. RNL sequences from *G. biloba* and *P. abies* support this interpretation (Fig. 5).

Although they are usually found in low numbers, we detect that ADR1-like sequence group has a few sequences in all of the seed plants tested to date, which contrast with other RNL groups. Taken together, the absence of plant species lacking ADR1-like sequences, the involvement of ADR1 in *Arabidopsis thaliana* drought tolerance and its identification as a cornerstone in NLR signalling³¹ clearly suggest that ADR1-like sequences are crucial for plant survival and fitness. Considering the phylogenetic distance between species, the compact aspect of the corresponding branch in the phylogenetic tree supports a significant sequence conservation previously documented in flowering plants⁴³. It has been suggested that ADR1-like proteins carry out a general function in disease resistance or play a critical role to control common pathogens²⁷. Its characteristics may also fit with a role in adaptation to abiotic stresses that affect most plants in their natural habitats, or a hub function acting downstream of a sensing network. In contrast, the heterogeneous NRG1-like group is made up of more polymorphic sequences, which is reflected by its scattered distribution in the phylogenetic tree (Fig. 4).

Most importantly, our data identify two Gymnosperm-specific groups (1 and 3) of RNLs that may harbour many recent duplications based on their lower levels of sequence divergence. The RPW8 domain motif organisation in these latter two groups shows relatedness with ADR1-like sequences, suggesting a shared origin with the ADR1 group. The extensive duplication and diversification of RNLs in conifers argues in favour of a role in adaptation to ubiquitous stress such as harsh environmental conditions.

Our RNA-Seq analysis in *P. glauca* suggests that the majority of differentially expressed NLRs are downregulated after a water deprivation of several days, which is in line with increased susceptibility to pathogens in stressed conifers^{48,49}. Interestingly, only RNLs from groups 1 and 3 appeared to be responsive to drought. Specifically, a subset of RNLs from group 3 were strongly overexpressed after drought treatment and few sequences were slightly downregulated, as observed for a majority of stress-responsive NLRs. In *Arabidopsis*, ADR1 was shown to be actively involved in drought tolerance⁷. Here we report for the first time in conifers, RNL sequences from a nearby group of ADR1 that respond significantly and early to drought stress. Their genomic localisation indicates that some of them may be clustered in tandem arrays. The comparative study of drought-responsive RNL clusters among clones or species that show differential resilience to drought should be of interest. Analyses of copy number variation, gene organisation, sequence diversity or expression among these clusters may reveal molecular mechanisms acting in adaptation to drought in conifers⁵⁰.

RNL origin and distribution in plant lineage. The recent sequencing of ancient genomes has shed light on the origin of NLR proteins¹⁵. The RNL subfamily arose before the Gymnosperm - Angiosperm split²⁷ putatively *de novo* from a non-coding region of an NLR-encoding gene or following domain fusion²⁶. No RNLs are mentioned in green algae¹⁵; nevertheless, we identified a few RNL sequences in the tested early land plants, including the liverwort *M. polymorpha*, and in the mosses *P. patens* and *S. fallax*, which harboured typical NB-ARC_{CNL} but no NB-ARC_{RNL} motifs. In these three species we also identified other RPW8-containing proteins. NB-ARC_{RNL} signatures were detected in all tested seed plants, including ancient plants such as *G. biloba* and *A. trichopoda*. It is possible that a pre-RNL emerged from a pool of CNL sequences in which the fusion of the RPW8-encoding exon occurred early in land plant followed by a transition toward RNL forms that arise before the divergence between Gymnosperms and Angiosperms. However, a study aiming to confirm this hypothesis and to decipher the RNL evolutionary history should include more sequences from ancient plant lineages (e.g. green algae, mosses, lycophytes, and ferns) but also other Gymnosperm sequences from Cycads and Gnetophytes. Expansion of the RNL sequence family is observed in groups 1 and 3 in the conifers and in group 4 in the Rosids. By contrast, taxonomic groups such as the Monocots have only a single ADR1-like RNL and have lost any other group representatives. A striking co-absence of NRG1-like and TNL sequences has been highlighted in Lamiales (a subgroup of Asterids), early-diverged Dicots (*Aquilegia coerulea*), and Monocots²⁷.

We identified a high abundance of both RNL and TNL genes in conifers. We also observed that the number of TNL genes correlates well with that of RNLs but not with that of CNLs in diverse seed plants. The quantitative link between TNLs and RNLs (approximate ratio of 1:10, RNLs:TNLs) is based on 49 plant species, despite potential variations in determinations of gene numbers among species. This asymmetrical ratio linking the abundances of RNLs and TNLs is surprising especially given that the two subfamilies are probably not physically linked, e.g. the chromosome 7 of *Prunus persica* contains 80% of the RNL pool but only 5% of the TNLs (Supplementary Table S6), and the dynamic NLR family has undergone several gene expansions and losses during evolution¹¹. In Angiosperms, the functionally characterised RNLs NRG1 and ADR1 act as downstream helpers of conventional sensor NLRs^{29,31}. NRG1 acts downstream of the activation of the TNLs N, Roq1 and RPP1, but not the CNLs Bs2 and Rps2^{28,30}. ADR1 and its paralogues are also thought to act in conjunction with TNL and CNL sensors^{29,31}. Recently, an extensive study revealed a larger NLR network involving numerous TNLs, including pair TNL pairs, and a few CNLs that signal via NRG1 and/or ADR1³¹. In this context, the proportionality between TNLs and

RNLs may be explained by the structure of their functional relationship (*i.e.* multiple sensors that signal via few helpers). RPW8 domains recruit TNL gene-signalling components such as EDS1, PAD4 and EDS5, but do not recruit CNL gene-signalling components⁵². New findings showed that signalling machineries, including EDS1, PAD4 and SAG101, co-evolved within plant species and clades for regulating downstream pathways in TNL immunity⁵³. A recent study that highlighted the loss of numerous NLRs and their gene-signalling components in aquatic Angiosperms paves the way to decipher the complex interactions between the plant immune system and drought tolerance⁵⁴. All together, these results support the hypothesis that RNLs may function redundantly in network, together with TNLs to activate NLR-triggered immunity. By contrast to the complex network recently identified in Solanaceae that is exclusively composed of sensor-helper CNLs⁵⁵, the complex TNLs/EDS1/RNLs might have evolved independently to provide a parallel machinery that triggers immunity especially in Rosids and in Gymnosperms.

Methods

Plant material and RNA extraction for transcriptomes. The study relied on four year-old saplings and germinated seeds of seven conifer species, including two spruces (*Picea glauca* and *Picea mariana*), two pines (*Pinus strobus* and *Pinus banksiana*), a fir (*Abies balsamea*), a larch (*Larix laricina*) of the Pinaceae family and a representative of the Cupressaceae family (*Thuja occidentalis*). All of the individuals were obtained from open-pollinated seed lots. The saplings were cultivated under standard greenhouse conditions⁵⁶ and included four individuals per species. Before the sampling, we deprived half of the saplings of each species of water for 14 days and watered the other half normally. We sampled six vegetative tissues from these greenhouse grown plants. Total RNAs were extracted from the different tissues by using the protocol described in⁵⁷ with modifications⁵⁸, and the MasterPure™ Plant RNA Purification kit. The detailed protocol is available in Supplementary Methods.

RNA-Seq library synthesis, sequencing and assembly. We sampled different conifer tissues and applied a drought stress to generate a maximum number of RNAs. We pooled all of the RNAs from the different tissues for each species separately (Supplementary Methods). We used 500 ng of total RNA from each of the species pools to synthesize mRNA libraries using TruSeq® Stranded mRNA kit following the supplier's instructions. Second strand synthesis and post PCR clean-up were performed using Axygen® AxyPrep™ Mag PCR Clean-Up Kit. Each library was quantified using a Nanodrop ND-1000 and characterised with an Agilent Bioanalyzer 2100 using High Sensitivity DNA chips. The Genome Quebec Innovation Centre at McGill University (Montreal, Quebec, Canada) carried out the sequencing of each pool using the "rapid run" procedure (paired-end 250 bp) with an Illumina HiSeq 2500 sequencing system.

The transcriptome assemblies were performed with Gyde software NUCLEAR version 3.2.16, RESOLVE version 2.6.12 and VISION version 2.6.12, and guided using the white spruce gene catalogue obtained by the sequencing of full-length insert cDNAs (FLICs) from white spruce cDNA libraries as described in⁵⁹. The detailed protocol of assembly is available in Supplementary Methods. The accessions numbers for all of the transcriptome assemblies and the sequence reads are mentioned in the additional data and a detailed list is available in Table 2 of the Supplementary Methods.

Functional annotation and motif discovery. For each of the seven conifer species, we applied a first filter by analysing the entire predicted protein dataset using the Pfam database with a cut-off value of 1.0. Sequences displaying the Pfam entries NB-ARC (PF00931), TIR (PF01582), RPW8 (PF05659) and LRR 1 to 6, 8, 9, and LRRNT (PF00560, PF07723, PF07725, PF12799, PF13306, PF13516, PF13855, PF14580 and PF01462, respectively) were extracted. Then, we characterised the domain organisation using HMMER web server version 2.19.0 with the selection of Pfam, TIGRFAM and superfamily HMM databases with significance e-values set to 0.01 (model) and 0.03 (Hit)⁶⁰. Atypical NLR sequences were also verified with InterProScan by using the entire database available. Motifs analysis and scanning were carried out using the MEME suite⁶¹ with maximum number of motifs set to 10, minimum-maximum motif width set to 5–30 and 5–50 and minimum sites per motif equal to 2.

Additional data sampling and extraction data. The full list of additional plant protein datasets and the versions that were screened is available in Supplementary Methods. RNL sequences were extracted using BlastP searches with (i) RPW8 domains of ADR1 (BAF00531.1), NRG1 (AAY54606.1), RPW8.1 (ACJ05907.1), RPW8.2 (ACJ72031.1) and a representative of each of the four conifer RNL clades detected in this study and (ii) NB-ARC_{RNL} laying between the RNBS-D and the QHD motifs as queries with broad e-value < 1. When TNL numbers were not available in the literature, we produced our own estimates by using BlastP (e-values < 10⁻⁸) searches with N (Q40392.1), RPS4 (CAB50708.1) and Ma (CAR94514.1) sequences deprived of LRR, as queries. BlastP (e-values < 30) with PL motifs²² were used to retrieve PL domains. All extracted sequence identities were verified using HMMER with default parameters and NB-ARC and PL motifs were aligned to be evaluated.

Sequence alignment, clustering, and phylogenetic reconstruction. Sequence alignments were performed with Muscle⁶² using default parameters and trimmed using MEGA⁷⁶³. Clustering analyses were conducted with CD-HIT⁶⁴ using default parameters and variable sequence identity cut-off values: 0.97 for elimination of redundant sequences (allelic sequences or alternative transcripts) leading the total number of sequences to be reduced to 3734, and 0.50–0.90 for the conifer RPW8 domains. To identify RNL groups in conifers, 185 full-length RPW8 sequences (>100 amino acids) were extracted and trimmed from our dataset. These RPW8 domains together with RPW8 domains of ADR1, RPW8.1, RPW8.2 and NRG1 were aligned to undertake the first phylogenetic analyses of conifer sequences. Then, the conifer sequences were clustered using a cut-off value of 0.60 to obtain 45 representative RPW8 sequences. Simultaneously, a second dataset was prepared using 351 Angiosperm, 13 *G. biloba* and 17 *P. abies* RNL sequences. These sequences were used to similarly extract full RPW8 domains (>100 amino acids) and clustered (cut-off values: 0.60) to obtain 152 representative RPW8

sequences for this second dataset. The 197 representative sequences were included into the phylogenetic analysis conducted using the web service <http://www.phylogeny.fr>⁶⁵. The phylogenetic trees were reconstructed using the maximum likelihood (ML) method implemented in the PhyML program (v3.1/3.0 aLRT). The WAG substitution model was selected with an estimated proportion of invariant sites and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma = 3.981 and gamma = 4.611). Reliability for internal branch was assessed using the aLRT test (SH-Like). The phylogenetic trees were then drawn using Dendroscope 3⁶⁶. Similarly, the JTT substitution model was used (with Gamma shape parameter: 3.648) resulting in the same tree topology (Supplementary Fig. S8).

Transcriptomic analysis of drought stress in three *P. glauca* clones. Foliage samples from the drought experiment in *P. glauca* described in Stival Sena *et al.*⁴⁰ were analysed by RNA-sequencing. The samples were collected at 0, 7, 14, 18 and 22 days from the beginning of the watering treatments. We used three unrelated and clonally-propagated genotypes (clones 8, 11 and 95) each with two replicates, in two conditions (well-watered control and drought with no watering) with five different sampling time points, and obtained 59 RNA samples to analyse (lacking one control replicate for the clone 95 at the time 7 days). The detailed protocol for RNA extraction and RNA-seq library synthesis modified from Stival Sena *et al.* (2018)⁴⁰ is available in Supplementary Methods. The pooling of 59 libraries into one lane for sequencing produced variable numbers of sequence reads per library (Supplementary Fig. S5). In order to limit the effect of size library, we added a filtering step keeping only genes with counts >5 in at least two of the libraries. A principal component analysis (PCA) also showed that replicates clustered mostly together, indicating that plants from the same treatment had similar profiles and suggesting that variation in library size had little impact on expression profiles (Supplementary Fig. S5). The DESeq2 v1.20.0 package in R software (Core Team 2018) was used to normalize count data. We calculated the differential expression between the treatment and the control foliage samples with the DESeq2 v1.20.0 package and the LRT approach. A lists of differentially expressed NLR genes with a FDR adjusted p-value < 0.05 are shown (Supplementary Table S4). Cluster v2.0.7., factoextra v1.0.5, and pheatmap v1.0.10 packages were used to generate heatmap with gene clustering analyses using manhattan distances and the ward.D2 clustering method^{67–76}.

Accession numbers. The raw data from the seven conifer transcriptomes are deposited in the European Nucleotide Archive (ENA) as part of the study SRP134160, bioproject PRJNA437248 (SRA accession SRR6816977 to SRR6816983). A list of accession numbers is available in supplementary methods.

References

- Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–9 (2006).
- Whitham, S. *et al.* The product of the tobacco mosaic virus resistance gene N: similarity to toll and the interleukin-1 receptor. *Cell* **78**, 1101–15 (1994).
- Rossi, M. *et al.* The nematode resistance gene Mi of tomato confers resistance against the potato aphid. *Proc. Natl. Acad. Sci. USA* **95**, 9750–4 (1998).
- Claverie, M. *et al.* The Ma Gene for Complete-Spectrum Resistance to Meloidogyne Species in Prunus Is a TNL with a Huge Repeated C-Terminal Post-LRR Region. *Plant Physiol.* **156**, 779–792 (2011).
- Ravensdale, M. *et al.* Intramolecular interaction influences binding of the Flax L5 and L6 resistance proteins to their AvrL567 ligands. *PLoS Pathog.* **8**, e1003004 (2012).
- Le Roux, C. *et al.* A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* **161**, 1074–1088 (2015).
- Chini, A., Grant, J. J., Seki, M., Shinozaki, K. & Loake, G. J. Drought tolerance established by enhanced expression of the CC-NBS-LRR gene, ADR1, requires salicylic acid, EDS1 and ABI1. *Plant J.* **38**, 810–22 (2004).
- van der Vossen, E. A. G. *et al.* The Rpi-blb2 gene from *Solanum bulbocastanum* is an Mi-1 gene homolog conferring broad-spectrum late blight resistance in potato. *Plant J.* **44**, 208–22 (2005).
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H. & Michelmore, R. W. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**, 809–34 (2003).
- Xue, J.-Y. *et al.* A primary survey on bryophyte species reveals two novel classes of nucleotide-binding site (NBS) genes. *PLoS One* **7**, e36700 (2012).
- Shao, Z.-Q. *et al.* Long-term evolution of nucleotide-binding site-leucine-rich repeat genes: understanding gained from and beyond the legume family. *Plant Physiol.* **166**, 217–34 (2014).
- Jia, Y., Yuan, Y., Zhang, Y., Yang, S. & Zhang, X. Extreme expansion of NBS-encoding genes in Rosaceae. *BMC Genet.* **16**, 48 (2015).
- Qian, L.-H. *et al.* Distinct Patterns of Gene Gain and Loss: Diverse Evolutionary Modes of NBS-Encoding Genes in Three Solanaceae Crop Species. *G3 (Bethesda)*. **7**, 1577–1585 (2017).
- Andolfo, G. *et al.* Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. *BMC Plant Biol.* **14**, 120 (2014).
- Gao, Y. *et al.* Out of Water: The Origin and Early Diversification of Plant R-Genes. *Plant Physiol.* **177**, 82–89 (2018).
- Yue, J., Meyers, B. C., Chen, J., Tian, D. & Yang, S. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytol.* **193**, 1049–63 (2012).
- Heath, M. C. Hypersensitive response-related death. *Plant Mol. Biol.* **44**, 321–34 (2000).
- Moffett, P., Farnham, G., Peart, J. & Baulcombe, D. C. Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. *EMBO J.* **21**, 4511–4519 (2002).
- van Ooijen, G. *et al.* Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. *J. Exp. Bot.* **59**, 1383–97 (2008).
- McHale, L., Tan, X., Koehl, P. & Michelmore, R. W. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* **7**, 212 (2006).
- Meyers, B. C. *et al.* Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* **20**, 317–32 (1999).
- Van Ghelder, C. & Esmejaud, D. TNL genes in peach: insights into the post-LRR domain. *BMC Genomics* **17**, 317 (2016).
- Xiao, S. *et al.* Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by RPW8. *Science* **291**, 118–20 (2001).
- Zhao, T. *et al.* A Truncated NLR Protein, TIR-NBS2, Is Required for Activated Defense Responses in the exo70B1 Mutant. *PLoS Genet.* **11**, 1–28 (2015).
- Nishimura, M. T. *et al.* TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in *Arabidopsis*. *Proc. Natl. Acad. Sci.* **114**, E2053–E2062 (2017).

26. Zhong, Y. & Cheng, Z. M. A unique RPW8-encoding class of genes that originated in early land plants and evolved through domain fission, fusion, and duplication. *Sci. Rep.* **6**, 1–13 (2016).
27. Collier, S. M., Hamel, L.-P. & Moffett, P. Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol. Plant-Microbe Interact.* **24**, 918–31 (2011).
28. Peart, J. R., Mestre, P., Lu, R., Malcuit, I. & Baulcombe, D. C. NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Curr. Biol.* **15**, 968–73 (2005).
29. Bonardi, V. *et al.* Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc. Natl. Acad. Sci.* **108**, 16463–16468 (2011).
30. Qi, T. *et al.* NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. *Proc. Natl. Acad. Sci. USA* **115**, E10979–E10987 (2018).
31. Castel, B. *et al.* Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol.* **222**, 966–980 (2019).
32. Meyers, B. C., Morgante, M. & Michelmore, R. W. TIR-X and TIR-NBS proteins: Two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. *Plant J.* **32**, 77–92 (2002).
33. Liu, J.-J. & Ekramoddoullah, A. K. M. Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex. D. Don.). *Mol. Genet. Genomics* **270**, 432–41 (2003).
34. Liu, J.-J. & Ekramoddoullah, A. K. M. The CC-NBS-LRR Subfamily in *Pinus monticola*: Targeted Identification, Gene Expression, and Genetic Linkage with Resistance to *Cronartium ribicola*. *Phytopathology* **97**, 728–36 (2007).
35. De La Torre, A. R. *et al.* Insights into conifer giga-genomes. *Plant Physiol.* **166**, 1724–32 (2014).
36. Raherison, E. S. M., Giguère, I., Caron, S., Lamara, M. & MacKay, J. J. Modular organization of the white spruce (*Picea glauca*) transcriptome reveals functional organization and evolutionary signatures. *New Phytol.* **207**, 172–87 (2015).
37. Warren, R. L. *et al.* Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *Plant J.* **83**, 189–212 (2015).
38. Shao, Z.-Q. *et al.* Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. *Plant Physiol.* **170**, 2095–109 (2016).
39. Nystedt, B. *et al.* The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579–84 (2013).
40. Stival Sena, J., Giguère, I., Rigault, P., Bousquet, J. & MacKay, J. Expansion of the dehydrin gene family in the Pinaceae is associated with considerable structural diversity and drought-responsive expression. *Tree Physiol.* **38**, 442–456 (2018).
41. Pavy, N. *et al.* A high-resolution reference genetic map positioning 8.8 K genes for the conifer white spruce: structural genomics implications and correspondence with physical distance. *Plant J.* **90**, 189–203 (2017).
42. Danot, O., Marquet, E., Vidal-Ingigliardi, D. & Richet, E. Wheel of Life, Wheel of Death: A Mechanistic Insight into Signaling by STAND Proteins. *Structure* **17**, 172–82 (2009).
43. Chini, A. & Loake, G. J. Motifs specific for the ADR1 NBS-LRR protein family in Arabidopsis are conserved among NBS-LRR sequences from both dicotyledonous and monocotyledonous plants. *Planta* **221**, 597–601 (2005).
44. Williams, S. J. *et al.* An Autoactive Mutant of the M Flax Rust Resistance Protein Has a Preference for Binding ATP, Whereas Wild-Type M Protein Binds ADP. *Mol. Plant-Microbe Interact.* **24**, 897–906 (2011).
45. Bernoux, M. *et al.* Comparative Analysis of the Flax Immune Receptors L6 and L7 Suggests an Equilibrium-Based Switch Activation Model. *Plant Cell* **28**, 146–59 (2016).
46. Urbach, J. M. & Ausubel, F. M. The NBS-LRR architectures of plant R-proteins and metazoan NLRs evolved in independent events. *Proc. Natl. Acad. Sci. USA* **114**, 1063–1068 (2017).
47. Savard, L., Li, P., Strausst, S. H., Chase, M. W. & Michaud, M. Chloroplast and nuclear gene sequences indicate Late Pennsylvanian time for the last common ancestor of extant seed plants. **91**, 5163–5167 (1994).
48. Klutsch, J. G., Shamoun, S. F. & Erbilgin, N. Drought stress leads to systemic induced susceptibility to a necrotrophic fungus associated with mountain pine beetle in *Pinus banksiana* seedlings. *PLoS One* **12**, e0189203 (2017).
49. Ryu, M., Mishra, R. C., Jeon, J., Lee, S. K. & Bae, H. Drought-induced susceptibility for *Cenangium ferruginosum* leads to progression of *Cenangium*-dieback disease in *Pinus koraiensis*. *Sci. Rep.* **8**, 1–14 (2018).
50. Trujillo-Moya, C. *et al.* Drought Sensitivity of Norway Spruce at the Species' Warmest Fringe: Quantitative and Molecular Analysis Reveals High Genetic Variation Among and Within Provenances. *G3 (Bethesda)*, **8**, 1225–1245 (2018).
51. Rathjen, J. P. & Dodds, P. N. Dancing with the Stars: An Asterid NLR Family. *Trends Plant Sci.* **22**, 1003–1005 (2017).
52. Xiao, S. *et al.* The atypical resistance gene, RPW8, recruits components of basal defence for powdery mildew resistance in Arabidopsis. *Plant J.* **42**, 95–110 (2005).
53. Lapin, D. *et al.* A coevolved EDS1-SAG101-NRG1 module mediates cell death signaling by TIR-domain immune receptors. *Biorxiv*, <https://doi.org/10.1101/572826> (2019).
54. Baggs, E. *et al.* Convergent gene loss in aquatic plants predicts new components of plant immunity and drought response. *Biorxiv*, <https://doi.org/10.1101/572560> (2019).
55. Wu, C.-H. *et al.* NLR network mediates immunity to diverse plant pathogens. *Proc. Natl. Acad. Sci. USA* **114**, 8113–8118 (2017).
56. Raherison, E. *et al.* Transcriptome profiling in conifers and the *PiceaGenExpress* database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. *BMC Genomics* **13**, 434 (2012).
57. Chang, S., Puryear, J. & Cairney, J. A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Report.* **11**, 113–116 (1993).
58. Pavy, N. *et al.* Identification of conserved core xylem gene sets: Conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytol.* **180**, 766–786 (2008).
59. Rigault, P. *et al.* A White Spruce Gene Catalog for Conifer Genome Analyses. *Plant Physiol.* **157**, 14–28 (2011).
60. Finn, R. D. *et al.* HMMER web server: 2015 update. *Nucleic Acids Res.* **43**, W30–8 (2015).
61. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–8 (2009).
62. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–7 (2004).
63. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–4 (2016).
64. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–9 (2006).
65. Dereeper, A., Audic, S., Claverie, J.-M. & Blanc, G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol. Biol.* **10**, 8 (2010).
66. Huson, D. H. & Scornavacca, C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61**, 1061–7 (2012).
67. Zhang, J. *et al.* Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. *Cell Host Microbe* **7**, 290–301 (2010).
68. Jupe, F. *et al.* Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics* **13**, 75 (2012).
69. Kale, S. M., Pardeshi, V. C., Barvkar, V. T., Gupta, V. S. & Kadoo, N. Y. Genome-wide identification and characterization of nucleotide binding site leucine-rich repeat genes in linseed reveal distinct patterns of gene structure. *Genome* **56**, 91–9 (2013).
70. Kohler, A. *et al.* Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Mol. Biol.* **66**, 619–36 (2008).
71. Wan, H. *et al.* Genome-wide analysis of NBS-encoding disease resistance genes in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. *BMC Genomics* **14**, 109 (2013).

72. Porter, B. W. *et al.* Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. *Mol. Genet. Genomics* **281**, 609–26 (2009).
73. Yang, S., Zhang, X., Yue, J.-X., Tian, D. & Chen, J.-Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics* **280**, 187–98 (2008).
74. Mun, J.-H., Yu, H.-J., Park, S. & Park, B.-S. Genome-wide identification of NBS-encoding resistance genes in *Brassica rapa*. *Mol. Genet. Genomics* **282**, 617–31 (2009).
75. Guo, Y.-L. *et al.* Genome-Wide Comparison of Nucleotide-Binding Site-Leucine-Rich Repeat-Encoding Genes in *Arabidopsis*. *Plant Physiol.* **157**, 757–769 (2011).
76. Kang, Y. J. *et al.* Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC Plant Biol.* **12**, 139 (2012).

Acknowledgements

We thank F. Larochelle and M.-A. Paré for assistance with plant procurement and growth, S. Laviolette and S. Senneville for project management, and K. Guay, K. Beaupré-Boivin, D. Milette and A. Azaiez for sample collections and processing. This study was developed in part as a collaboration with the ProCoGen project (<https://bfw.ac.at/rz/bfwcms2.web?dok=9020>) (funded by the European Union's Seventh Framework Programme). This work was supported by a research grant from the Québec Ministry for the Economy, Science and Innovation, program for the support of international initiatives to AD, JB and PR. CVG was supported by the Plant – Nematode Interaction team, at ISA-INRA in Sophia Antipolis, France. JM is supported by M. and A. Wood and by the P. Wood Trust.

Author Contributions

J.M., P.R., J.P., G.J.P., A.D. and J.B. designed the transcriptome research, I.G. and S.C. carried out and supervised the field and laboratory work, P.R. analysed the RNA-Seq data, developed the reference transcriptomes and did the data releases, J.S.S. planned and carried out the drought responsiveness experiment and G.J.P. performed the related RNA-Seq data analysis, C.V.G., J.M. and D.E. planned and designed the NLR analyses, C.V.G. developed and analysed the NLR data and drafted the manuscript, C.V.G. and J.M. revised the manuscript, all of the authors commented and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-47950-7>.

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) 2019

CHAPITRE 5

CARACTERISATION DU GENE *RMja* CHEZ L'AMANDIER

Chapitre 5 : CARACTERISATION DU GENE *RMja* CHEZ L'AMANDIER

5.1 Introduction

Le gène *Ma* présent chez le prunier myrobolan (*P. cerasifera*) possède un spectre qui contrôle toutes les espèces testées à ce jour. Les deux autres gènes identifiés chez les *Prunus* ont un spectre plus étroit. Ainsi le gène *RMja* de l'amandier (*P. dulcis*) contrôle l'espèce majeure *M. javanica* mais ne confère pas de résistance à une autre espèce majeure, *M. incognita*, alors que le gène *RMia* chez le pêcher (*P. persica*) contrôle *M. incognita* mais ne contrôle pas *M. javanica*. L'accès à cette variabilité de spectre dans le genre *Prunus* permet d'envisager à terme l'étude des déterminants qui sont responsables de cette 'spécificité' de résistance vis-à-vis des nématodes à galles. Si les gènes TNL à spectre variable qui nous intéressent ont une même origine et donc appartiennent à un même cluster (gènes paralogues ou, mieux, orthologues), une approche comparative des déterminants mis en jeu peut être conduite. Du fait que, parmi les gènes des *Prunus*, *Ma* et *RMja* sont sur le même groupe de liaison (LG7) (Van Ghelder *et al.*, 2010), une étude préalable de leur relation génétique a donc été conduite. Nous avons pour cela travaillé à la localisation fine de *RMja* chez l'amandier (*P. dulcis*). Ce travail s'est appuyé sur le fait, d'une part, que l'amandier est très proche du pêcher, espèce de référence dont le génome séquencé est de très bonne qualité, mais est également proche du prunier myrobolan (*P. cerasifera*), porteur de *Ma* et, d'autre part, que la forte synténie entre ces espèces permet une utilisation interspécifique de leurs marqueurs moléculaires.

Afin de nous permettre d'entreprendre cette cartographie à haute résolution du gène *RMja*, nos partenaires de l'INRA d'Avignon ont créé une descendance F2 d'amandier d'environ 1500 individus issus d'un croisement entre *P. dulcis* cv. 'Alnem 1', porteur homozygote du gène *RMja* (*RMja/RMja*), et *P. dulcis* cv. 'Lauranne', sensible (*rmja/rmja*) à *M. javanica*. Une première descendance d'environ 900 plants âgés de 6 mois a été inoculée avec une population de *M. javanica* au printemps 2016. Nous avons réalisé le phénotypage en analysant les symptômes sur l'ensemble de ces plants en fin d'année 2016. Parallèlement nous avons développé des marqueurs moléculaires de type SSR et SNP en utilisant le génome du pêcher et avons génotypé ces individus. Cette première campagne nous a permis de produire un marqueur très proche du gène *RMja* puisqu'un seul individu recombinant a été identifié pour celui-ci. Afin de renforcer cette cartographie, nous avons testé avec une approche similaire les 600 autres individus entre le printemps et l'automne 2017. Cette deuxième phase d'expérimentation nous a permis de définir un nouveau marqueur situé de l'autre côté du gène et présentant également un seul individu recombinant. La stratégie de cartographie haute-résolution du gène *RMja* est illustrée en Figure

21.

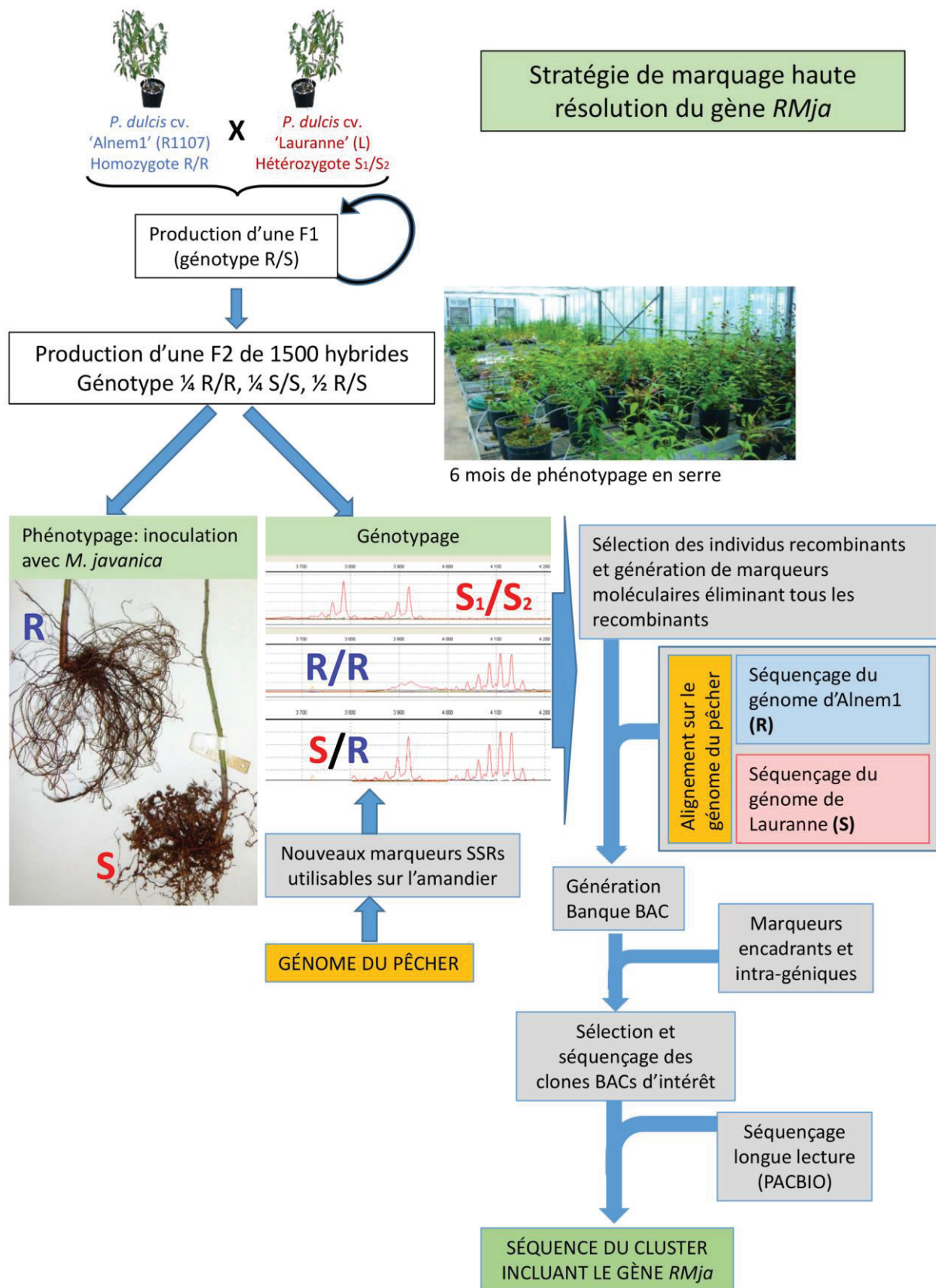


Figure 21 - Logigramme décrivant la stratégie de localisation et identification du gène *RMja* chez l'amandier *P. dulcis* cv. 'Alnem 1'.

Les résultats combinés des deux phases montrent que le gène *RMja*, inclus entre ces deux marqueurs, est localisé dans un intervalle de moins de 100 kb sur le génome du pêcher. Cette fenêtre encadre un cluster de 4 TNLs chez le pêcher parmi lesquels se trouve l'orthologue du gène *Ma*. Le séquençage des génomes des clones parentaux résistant 'Alnem1' et sensible 'Lauranne' et leur alignement sur le génome du pêcher nous ont ensuite permis de révéler la présence de 3 gènes seulement dans le cluster en question chez l'amandier, dont deux ont une structure altérée. Afin d'obtenir une séquence fiable de la zone contenant le gène *RMja* chez l'amandier, nous avons généré, dans un premier temps, une banque BAC du clone 'Alnem 1' avec nos partenaires du CNRGV de Toulouse. Grâce à nos marqueurs, nous avons sélectionné deux clones BAC contenant la zone d'intérêt et réalisé leur séquençage avec une technologie produisant de longs fragments (« long reads »). Leur séquence assemblée, couvrant 336 kb du génome de l'amandier, contient le gène *RMja*. Dans un deuxième temps, une analyse bio-informatique couplée à des données d'expression nous a permis de conclure que le gène *RMja* est très probablement l'orthologue du gène *Ma* chez l'amandier. Dans ce cas, le gène *RMja* présente une architecture très similaire à celle de *Ma* avec cependant des différences marquées localisées principalement dans les domaines PLs qui pourraient expliquer la divergence de spectre de résistance entre les deux gènes. Plus largement, la comparaison des séquences couvrant le cluster de TNLs chez l'amandier avec celles disponibles chez le pêcher (génome) et chez le prunier (séquences BAC issues de son clonage dans notre laboratoire ; (Claverie *et al.*, 2011)) nous ont permis d'aborder l'évolution de ce cluster de TNLs. Nous avons ainsi pu mettre en lumière les mécanismes génétiques ayant conduit au façonnage des structures particulières présentes dans ce cluster.

5.2 Article publié dans *Frontiers in Plant Science*

Van Ghelder, C., Esmenjaud, D., Callot, C., Dubois, E., Mazier, M. and Duval, H. (2018) **Ma Orthologous Genes in Prunus spp. Shed Light on a Noteworthy NBS-LRR Cluster Conferring Differential Resistance to Root-Knot Nematodes.** *Frontiers in Plant Science*. 9:1269. Doi: 10.3389/fpls.2018.01269.

NB : les données additionnelles liées à cet article sont disponibles à l'adresse suivante:

<https://www.frontiersin.org/articles/10.3389/fpls.2018.01269/full#supplementary-material>



Ma Orthologous Genes in *Prunus* spp. Shed Light on a Noteworthy NBS-LRR Cluster Conferring Differential Resistance to Root-Knot Nematodes

Cyril Van Ghelder^{1*}, Daniel Esmenjaud¹, Caroline Callot², Emeric Dubois³, Marianne Mazier⁴ and Henri Duval⁴

¹ Institut Sophia Agrobiotech, INRA, CNRS, Université Côte d'Azur, Sophia Antipolis, France, ² Centre National de Ressources Génomiques Végétales, INRA, CNRS, Université de Toulouse, Castanet-Tolosan, France, ³ MGX-Montpellier GenomiX, CNRS, Montpellier, France, ⁴ Unité de Génétique et Amélioration des Fruits et Légumes (GAFL), INRA, Montfavet, France

OPEN ACCESS

Edited by:

Jens Staal,
Ghent University, Belgium

Reviewed by:

Daguang Cai,
Christian-Albrechts-Universität zu Kiel,
Germany
Volkan Cevik,
University of Bath, United Kingdom

*Correspondence:

Cyril Van Ghelder
cyril.van-ghelder@inra.fr

Specialty section:

This article was submitted to
Plant Microbe Interactions,
a section of the journal
Frontiers in Plant Science

Received: 04 May 2018

Accepted: 10 August 2018

Published: 11 September 2018

Citation:

Van Ghelder C, Esmenjaud D, Callot C, Dubois E, Mazier M and Duval H (2018) *Ma* Orthologous Genes in *Prunus* spp. Shed Light on a Noteworthy NBS-LRR Cluster Conferring Differential Resistance to Root-Knot Nematodes. *Front. Plant Sci.* 9:1269. doi: 10.3389/fpls.2018.01269

Root-knot nematodes (RKNs) are considerable polyphagous pests that severely challenge plants worldwide and especially perennials. The specific genetic resistance of plants mainly relies on the NBS-LRR genes that are pivotal factors for pathogens control. In *Prunus* spp., the *Ma* plum and *RMja* almond genes possess different spectra for resistance to RKNs. While previous works based on the *Ma* gene allowed to clone it and to decipher its peculiar TIR-NBS-LRR (TNL) structure, we only knew that the *RMja* gene mapped on the same chromosome as *Ma*. We carried out a high-resolution mapping using an almond segregating F2 progeny of 1448 seedlings from resistant (R) and susceptible (S) parental accessions, to locate precisely *RMja* on the peach genome, the reference sequence for *Prunus* species. We showed that the *RMja* gene maps in the *Ma* resistance cluster and that the *Ma* ortholog is the best candidate for *RMja*. This co-localization is a crucial step that opens the way to unravel the molecular determinants involved in the resistance to RKNs. Then we sequenced both almond parental NGS genomes and aligned them onto the RKN susceptible reference peach genome. We produced a BAC library of the R parental accession and, from two overlapping BAC clones, we obtained a 336-kb sequence encompassing the *RMja* candidate region. Thus, we could benefit from three *Ma* orthologous regions to investigate their sequence polymorphism, respectively, within plum (complete R spectrum), almond (incomplete R spectrum) and peach (null R spectrum). We showed that the *Ma* TNL cluster has evolved orthologs with a unique conserved structure comprised of five repeated post-LRR (PL) domains, which contain most polymorphism. In addition to support the *Ma* and *RMja* orthologous relationship, our results suggest that the polymorphism contained in the PL sequences might underlie differential resistance interactions with RKNs and an original immune mechanism in woody perennials. Besides, our study illustrates how PL exon duplications and losses shape TNL structure and give rise to atypical PL domain repeats of yet unknown role.

Keywords: almond, *Meloidogyne*, nucleotide-binding site - leucine-rich repeat (NBS-LRR), *Prunus*, resistance gene, root-knot nematodes (RKNs), TIR-NBS-LRR (TNL)

INTRODUCTION

Plant-parasitic nematodes cause huge losses on agriculture worldwide (Nicol et al., 2011). Based on their economic impact, the root-knot nematodes (RKNs) *Meloidogyne* spp. are considered as the number one threat caused by nematodes on crops (Jones et al., 2013). The predominant RKN species are the tropical parthenogenetic mitotic species *Meloidogyne arenaria*, *M. incognita* and *M. javanica* and to a lesser extent the temperate parthenogenetic meiotic *M. hapla*. Nevertheless, minor RKN species such as *M. enterolobii*, *M. chitwoodi*, or *M. fallax* have been recently described as emerging problems for agriculture possibly promoted by climate change or transition in agronomic production (Elling, 2013). The extreme polyphagy of RKNs, characterized by their ability to infect over 5500 plant species, highly increased the interest to identify the actors and generic mechanisms underlying their interaction with plants, especially those involved in plant immunity (Blok et al., 2008).

The plant innate immunity relies on two main processes. The first level involves transmembrane receptors that detect generic molecules usually associated with pathogens. Following this non-specific recognition, this pathway, called pathogen-associated molecular patterns (PAMP) triggered immunity (PTI), activates MAP-kinases cascades leading to plant immunity. This first level of defense can be targeted at different levels by particular pathogen-secreted molecules, known as effectors, leading to inhibition of immunity. In return, effectors can be recognized in a specific manner through a second mechanism so called effector-triggered immunity (ETI) (Jones and Dangl, 2006). The cornerstone of ETI is the NBS-LRR gene family that is involved in plant adaptation to biotic and abiotic stresses. Hundreds of NBS-LRR genes are usually found in genomes of flowering plants (Meyers et al., 1999). They can be subject to deletion through unequal crossing-over or (non)-homologous repair events (Luo et al., 2012) or to expansion involving tandem, ectopic duplications and, to a lesser extent, segmental duplications (Leister, 2004). The dynamic mechanism of gene duplications explained the generic cluster organization of NBS-LRR genes found in plant genomes. This large gene family is split formally into two sub-groups, the TIR-NBS-LRR (TNL) and non-TIR NBS-LRR according to the presence or not of a Toll-interleukin-1 receptor domain in the N-terminal part of the protein (Meyers et al., 1999). Given their modular organization and distinctive domain functions, TNLs act as a multi-function tool for plant immunity. Indeed, the leucine-rich repeat (LRR) domain is polymorphic and often described by its ability to detect pathogens through direct or indirect effector recognition (Jones and Dangl, 2006). The latter process gives rise to some conceptual examples illustrated by the guard protein and decoy models (van der Hoorn and Kamoun, 2008). The central nucleotide-binding site (NBS) domain binds alternatively ADP or ATP leading to conformational modification and by consequence to switch from “OFF” to “ON” states of the protein (Bernoux et al., 2016). The transition between the LRR and the NBS domains is carried by a conserved exon coding for the NBS-LRR Linker (NLL). Once the protein is activated, the TIR domain is released triggering a downstream signaling pathway leading to plant immunity

that is often associated with a programmed cell death called Hypersensitive-like Response (Heath, 2000). However, each domain presumably covers a larger field of competence such as intra or extra-molecular interactions involved in the maintenance of the 3D protein folding, dimerization or alternative effector recognition (Rairdan and Moffett, 2006; Burch-Smith et al., 2007). Besides these three canonical domains, it exists a fourth domain, peculiar to TNLs, named post-LRR (PL) domain after its regular position (Van Ghelder and Esmenjaud, 2016). This domain, with unknown role or structure, is usually present in one copy in most of the cloned TNL genes such as *N*, *RPS4*, or *Gro1-4* (Whitham et al., 1994; Gassmann et al., 1999; Paal et al., 2004). Over the last decades, many efforts have been deployed to identify novel genes conferring strong and sustainable natural resistance in plants with agronomical interest. The *Prunus* species are important fruit crops worldwide and, amongst them, *Prunus persica* (peach), *P. cerasifera* (Myrobalan plum) and *P. dulcis* (almond), *qua* perennials, are challenged continuously by RKNs. The common ancestor of the *Prunus* genus emerged approximately 61 Myr in Eastern. The plums (subgenus *Prunophora*) originating from the Eastern Europe and the Middle East, diverged earlier than the almond and peach (subgenus *Amygdalus*) that are native from Western and Eastern Asia (estimated absolute age of 54.6 Myr for *Prunophora* versus 48.9 Myr for *Amygdalus*) (Chin et al., 2014). Although the almond genome is not available yet, the compact peach genome (265 Mbp; Verde et al., 2013) can be considered as the reference genome for *Prunus* species in particular for the diploid (2n = 16) species, peach, almond, Japanese and Myrobalan plums, and apricot.

The *Ma* gene, from *P. cerasifera*, was the first TNL gene, conferring resistance to RKNs, cloned in a woody perennial plant (Claverie et al., 2011). Conversely to the *Mi-1* tomato CNL R gene (Milligan et al., 1998), the *Ma* gene confers a heat-stable and complete-spectrum resistance to RKNs (Claverie et al., 2011), and no natural or selected virulent isolates have been detected yet (Khallouk et al., 2013). Besides these attractive biological properties, this gene displays a unique C-terminal structure, which is only found in the *Prunus* and *Malus* genera, made of five PL domains completing the TIR-NBS-LRR canonical part (Van Ghelder and Esmenjaud, 2016). Apart from the *Ma* gene, lying on *Prunus* chromosome 7, several other RKNs R genes with more restricted resistance spectra have been identified (Saucet et al., 2016). These are the *RMia* gene from peach, which is most probably a TNL located in a 92-kb interval on the chromosome 2 (Duval et al., 2014) and the *RMja* gene from almond. Interestingly, the latter confers a resistance to *M. javanica* but not to *M. incognita*. A low-resolution mapping strategy carried out on few progenies localized the *RMja* gene on the chromosome 7 (linkage group 7) (Van Ghelder et al., 2010).

The overall goal of this study was to pave the way to identify the molecular determinants involved in the RKN resistance. We have considered three *Prunus* species through accessions that display distinct spectra for resistance to RKNs ranging from complete (plum P.2175/*Ma*) to partial (almond Alnem1/*RMja*) and null (peach Lovell/*Prupe.7G065400*). The specific research objectives were: 1) to localize the *RMja* gene in almond using

high-resolution mapping; 2) to obtain a reliable sequence of the *RMja*-candidate region; 3) to identify the polymorphism that could be involved in the resistance mechanism; 4) and to investigate the evolution of the complete *Ma/RMja* TNL cluster. We showed that the *RMja* gene is located in the *Ma* TNL cluster. Almond NGS genomes, BAC sequences and expression data strongly suggest that the *RMja* gene is the ortholog of the *Ma* plum gene. Additionally, we shed light on the dynamic evolution of the complete *Ma* TNL cluster in plum, peach and almond revealing an atypical focal point for RKN resistance in *Prunus* spp.

MATERIALS AND METHODS

Plant and Nematode Materials

The resistant (R) accession *P. dulcis* cv. 'Alnem1' controls *M. arenaria* and *M. javanica* but not *M. incognita* and a single gene named *RMja* has been identified to confer resistance against *M. javanica* (Kochba and Spiegel-Roy, 1975; Esmenjaud et al., 2009; Van Ghelder et al., 2010). We produced an intraspecific progeny from almond. Accession 'Alnem1,' carrying the homozygous dominant *RMja* gene (*RMja/RMja*), was crossed with the susceptible (S) accession 'Lauranne' (recessive for *RMja*; *rmja1/rmja2*) to create a F1 heterozygous progeny. Several Lauranne x Alnem1 F1 hybrids were self-pollinated in greenhouses to produce a segregating F2 progeny (F2 LxA) with three different resistance genotypes (*RMja/RMja*, *RMja/rmja1or2*, and *rmja1or2/rmja1or2*). Shelled almonds were collected and placed onto a culture medium to induce germination.

The population *M. javanica* "Higuera" from Cabrils (Cataluna, Spain) was reared from a single egg mass and maintained on RKN susceptible tomatoes (*Lycopersicon esculentum* cv. 'St. Pierre'). Its RKN species identity was verified using isoesterase profiling method as defined by Janati et al. (1982).

Experimental Procedure for Phenotyping

A number of 1448 F2 LxA seedlings, 6-month-old, were planted in pairs in 5-liter pots, placed in greenhouse, irrigated individually every 2 days and grown until harvested for rating at a mean temperature of 25°C (extremes 20–30°C). In parallel, 5-leaf tomato plantlets grown in 250-ml containers were inoculated with 500 24–72-h old juveniles of *M. javanica*. After 2 months, aerial parts of tomato plants were removed and the contaminated soil of one container, including the infected root system, was transplanted into each *Prunus* pot to induce a high- and durable-inoculum pressure. After 6 months, the root system of each *Prunus* plant was carefully examined and rated into two classes: resistant (R) (absence of galls) or susceptible (S) (presence of gall(s)).

Nucleic Acid Extraction, cDNA Synthesis and Genotyping

The genomic DNA of all F2 plants was extracted using 100 mg of frozen young leaves. Each sample was ground with a mixer

mill in 330 µl of extraction buffer (sorbitol 0,35M, Tris 0,1M, EDTA 5 mM, 4 mg of sodium metabisulfite), 330 µl of lysis buffer (Tris 0.2 M, EDTA 50 mM, NaCl 2 M, CTAB 2%) and 130 µl of sarkosyl 5%. After a chloroform-isoamylalcohol (24:1) procedure, a precipitation with isopropanol and three ethanol washes, the DNA content was eluted in water and treated with RNase.

RNA was extracted using 100 mg of young roots or leaves according to the procedure of Tong et al. (2012) with modifications. Fresh RNA samples were treated with DNase using the Turbo DNA-free kit and evaluated with a Nanodrop. First-strand cDNA were synthesized using the maxima H minus first strand synthesis kit with either specific and oligo (dT) primers. 3 µl of first-strand cDNA were used in PCR using primers defined in **Supplementary Table S1** to conduct transcript detection. The *P. persica* actin gene Prupe.02G235000 was used as control for cDNA synthesis.

PCR were carried out using the MyTaq DNA polymerase kit or the Expand long range kit for longer fragments following the manufacturer's instructions. SSR primers used (**Supplementary Table S1**) for the genotyping were 5'-fluorescent-labeled with different fluorophores and PCR products were analyzed with a 3130xl Genetic Analyzer.

Genome Sequencing of the Almond Parents

DNA extracted from leaves of accessions 'Alnem1' and 'Lauranne' was used to prepare genomic libraries. Tagmentation and PCR amplifications were carried out using the Illumina Nextera DNA sample preparation kit following the supplier's instructions. The validation of the genomics libraries were performed with a DNA quantification using an Agilent high sensitivity chip and by qPCR. The cluster was carried out inside a flow-cell using the Illumina cluster generation kit. Then, a paired-end 125 sequencing was achieved using an Illumina HiSeq 2500 with a sequence by synthesis (SBS) technique.

Image analysis and base calling were performed using the HiSeq Control Software and Real-Time Analysis component. Demultiplexing was performed using Illumina's Conversion Software (bcl2fastq 2.17). The quality of the data was assessed using FastQC from the Babraham Institute and the Illumina software SAV (Sequence Analysis Viewer). Potential contaminants were investigated with the FastQ Screen software from the Babraham Institute.

BWA-MEM (v 0.7.12) was used to align reads to the *Prunus persica* genome (v 2.0). MGX-Montpellier GenomiX core facility performed the sequencing of both clones. The raw data of both genomes have been deposited with links to the NCBI BioProject accession number PRJNA448729 (BioSamples numbers SAMN08863554 and SAMN08863555).

Genomic Library and Bacterial Artificial Chromosome (BAC) Sequencing

High molecular weight (HMW) genomic DNA was prepared from young frozen leaves from accession 'Alnem1' as described by Peterson et al. (2000) and Gonthier et al. (2010). Agarose embedded HMW DNA was partially digested with

HindIII, subjected to two size selection steps by pulsed-field electrophoresis using a CHEF Mapper system. DNA was eluted, ligated into the pIndigoBAC-5 HindIII-Cloning Ready vector and transformed in *Escherichia coli* electrocompetent cells. Pulsed-field migration programs electrophoresis buffer and ligation desalting conditions were performed based on Chalhoub et al. (2004). The resulting library represents ~7-fold coverage of accession 'Alnem1' (48 plates, 18,432 BAC clones and mean insert size of 134 kb). BAC clones were spotted on nylon membrane, screened with radioactive labeled probes designed in exons PL1, PL2 and PL5 and revealed by High-density filter reader program. The positive clones were verified by real-time PCR using the specific primers designed in exons PL1, PL2 and PL5. 2 µg of each individual BAC clone of interest were pooled for the construction of a SMRT® library using the standard Pacific Biosciences preparation protocol for 10 kb library with PacBio® Barcoded Adapters. The pool was then sequenced in one SMRT cell using the P6 polymerase with C4 chemistry. Sequencing was performed on PacBio RS II sequencer. After a demultiplexing step, the sequence assembly was performed following the HGAP PacBio workflow (Chin et al., 2013), and using the SMRT® Analysis (v2.3) software suite for HGAP implementation¹. BAC ends sequences confirmed the position of selected clones, on the peach genome. BAC library construction and screening, and BAC clone sequencing and assembly were performed by the INRA-CNRGV. The data have been deposited with links to NCBI BioProject accession number PRJNA448736.

¹<https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP>

Sequence Analysis

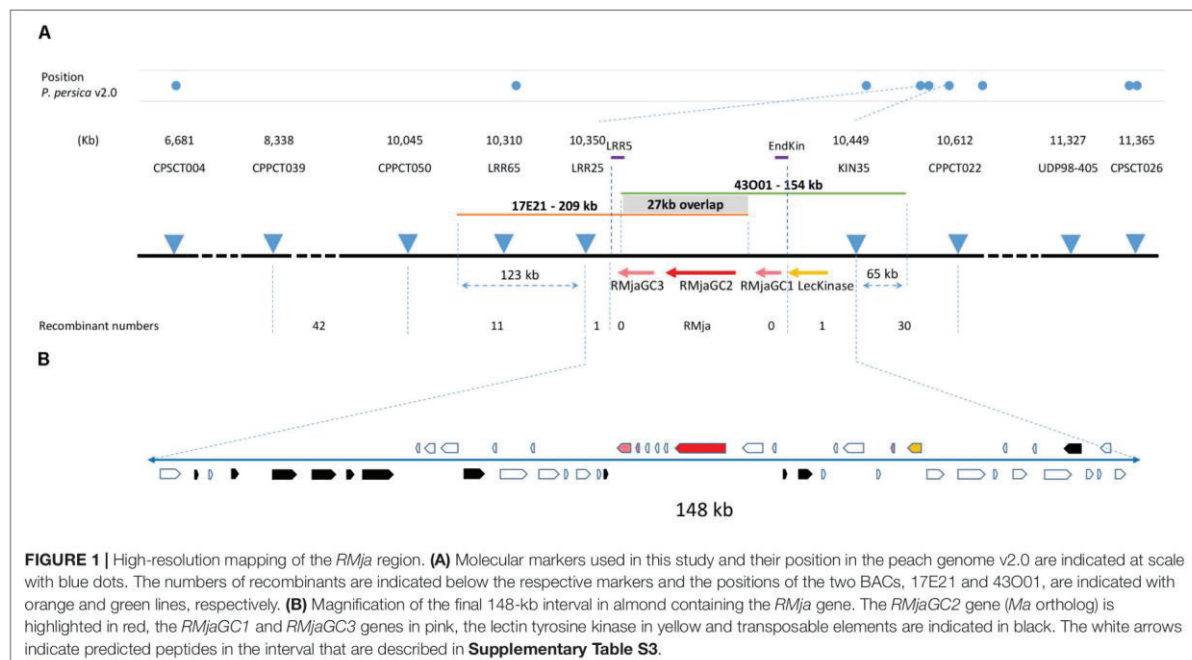
HMM-based gene structure predictions were carried out using FGENESH (Solovyev et al., 2006) and the *P. persica* specific gene-finding parameters. Read alignment and index files (BAM and BAI files), genomic feature files (GFF3) and *P. persica* v2.0.a1 genome² were analyzed with the Integrative Genomics Viewer (IGV) (Robinson et al., 2011). Alignments, sequence trimmings, mutation rate and phylogenetic analyses were evaluated using MEGA7 (Kumar et al., 2016). Conservative, semi- and non-conservative amino acid substitutions were determined using the Clustal Omega assignment (Sievers et al., 2011). Detection of promoter and regulatory elements were carried out using Cister (Frith et al., 2001) with the *Ma*, *RMjaGC2* and *Prupe.7G065400* genes extended with 7101 bp upstream each of them. We screened 5414, 5414, and 2721 bp upstream the *Ma*, *RMjaGC2* and *Prupe.7G065400* genes, respectively, for W-boxes using PLACE 26.0 (Higo et al., 1999) and PlantTFDB 4.0 (Jin et al., 2017).

RESULTS

High-Resolution Mapping of the *RMja* Gene

We localized the *RMja* gene between two SSR markers that cover a window of 0.14 cM. This region corresponds to a physical interval of 99 kb in the peach genome v2.0.a1. The high-resolution mapping strategy conducted on 1448 total F2

²ftp://ftp.bioinfo.wsu.edu/species/Prunus_persica/Prunus_persica-genome.v2.0.a1/assembly/



almond seedlings revealed a single recombination event for each of the flanking markers, LRR25 and KIN35. Internal primers named LRR5 and EndKin designed within this final interval, amplified fragments that eliminated any recombination events allowing us to refine the *RMja* gene localization (Figure 1A). This mapping was actually performed in two steps. In a first step, we used a progeny composed of 907 F2 seedlings that clearly split into 625 R individuals and 282 S individuals. We tested the S and R parental accessions together with twenty hybrids to select reliable and polymorphic 5'-fluorescent-labeled SSR markers spread in this LG7 region from the peach genome (Verde et al., 2013). Genotyping the 907 progenies with the flanking markers CPPCT039 and CPPCT022 (Supplementary Table S1) revealed 27 and 16 recombinant individuals on each side of the gene, respectively. In a second step, the 541 F2 individuals obtained segregated into 411 R: 130 S individuals (3 R: 1S ratio; $\chi^2 = 0.27$; $P = 0.6$). The overall individuals segregated into 1036 R: 412 S individuals ($\chi^2 = 9.2$; $P = 0.002$). In the candidate interval, we developed three new SSR markers, LRR65, LRR25 and Kin35 (Supplementary Table S1) that revealed eleven, one and one recombinant individuals in the total progeny, respectively (Figure 1A).

The *RMja* Gene Is Located in a Region Containing the *Ma* Ortholog

In the peach genome, a further examination of the *RMja* final interval marked out by the LRR25 and KIN35 markers, revealed eleven candidate genes named *Prupe.7G065100* to *Prupe.7G066100* (Supplementary Table S2). Among these candidate genes, a functional annotation identified one tyrosine kinase (*Prupe.7G065700*) and four TNLs (*Prupe.7G065300* to *Prupe.7G065600*) including the *Ma* ortholog (*Prupe.7G065400*). We sequenced the genome of the parental accessions 'Alnem1' (R) and 'Lauranne' (S) to further characterize the candidate region in almond. The phyletic proximity between peach and almond allowed us to map accurately the reads on the *P. persica* genome assembly (v2.0.a1). All over the region, the alignment of both the R and S almond reads onto the *P. persica* genome indicated a similar pattern involving alternation between syntenic regions with peach, and indels of unknown size (Figure 2A). We identified two profiles in the syntenic regions showing either (i) a standard coverage ($20\times - 50\times$) and moderate nucleotide polymorphism or (ii) a high coverage ($>100\times$) and nucleotide polymorphism. The first pattern corresponds to orthologous loci whereas the second, presumably, reflects the misalignment of ectopic transposable elements (TEs). Magnification of the region spanning from the first TNL (*Prupe.7G065300*) to the lectin kinase (*Prupe.7G065700*), arbitrarily named R region, highlights the absence of the ortholog of the TNL *Prupe.7G065500* in almond (Figure 2B). The almond orthologs of *Prupe.7G065700* (lectin kinase) and *Prupe.7G065400* (*Ma* ortholog TNL) display a significant level of homology (illustrated in gray in Figure 2B) unlike the orthologs of *Prupe.7G065300* and *Prupe.7G065600* showing nucleotide polymorphism and numerous indels. More specifically, the *Prupe.7G065400* gene shows conserved structural features and low polymorphism in predicted coding regions in

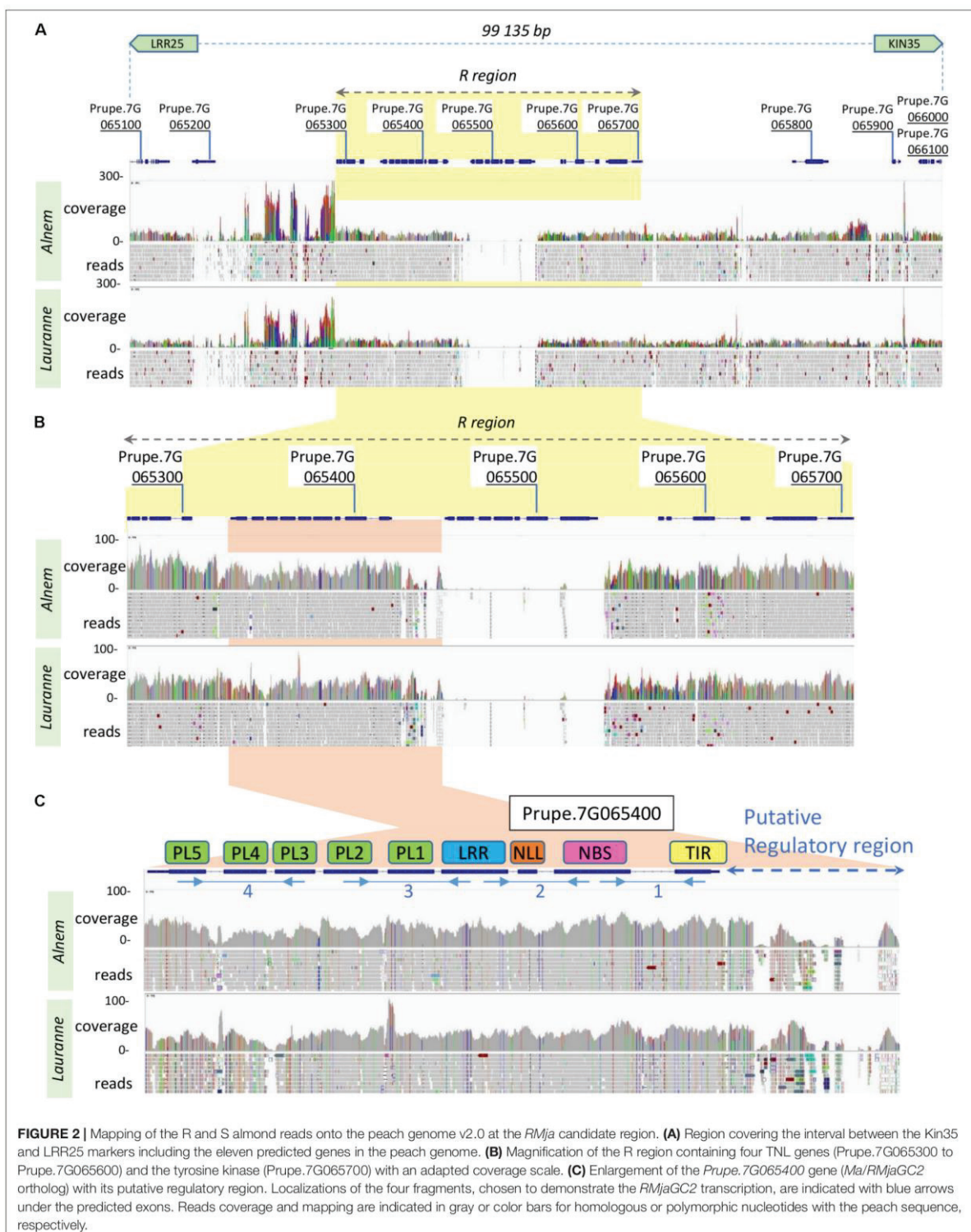
either the homozygote 'Alnem1' (R), or 'Lauranne' (S), or peach 'Lovell' accessions. Only the upstream putative regulatory region displays substantial divergences between both almonds and the peach accessions (Figure 2C).

BAC Sequences and Expression Data Strongly Suggest That *RMja* Is the Ortholog of the *Ma* Gene

We selected BAC clones that correspond to the region of interest from a genomic library constructed from the R accession 'Alnem1.' We sequenced two overlapping BAC clones ('17E21' and '43O01') [PRJNA448736] covering the KIN35-LRR25 interval, using long-read sequencing technology to overcome potential misassembly due to repeated sequences. These two sequences meet the quality criteria with a quality value $QV > 48$ meaning a high confidence in nucleotide identification. The BAC clone '43O01' contains an insert of 154 kb that can be aligned from 65 kb upstream the KIN35 marker to the *RMjaGC3* gene. The second BAC clone sequence '17E21' is 209 kb long and maps from the *RMjaGC1- RMjaGC2* intergenic region to 123 kb downstream the LRR25 marker (Figure 1A). The overlapping region between the two BAC clones covers a sequence of 27 kb that displays a similarity of 99.97%. Only eight nucleotide deletions were found and localized in five SSR regions in the 27 kb, whereas the remainder 190 kb identified SSRs were strictly identical. The fragment length between the KIN35 and LRR25 markers is 49 kb longer in almond (Supplementary Table S5) than in peach.

Gene finding analyses and functional annotations revealed 48 predicted ORFs including one lectin kinase, only three TNLs of different sizes, small peptides (potential artefacts) and TEs that explain the size variation in this region (Figure 1B and Supplementary Table S3). The lectin kinase that precedes the TNL cluster is highly conserved between plum, peach and almond ($>98\%$ amino acid identity). The protein is composed of a signal peptide, an extra-cellular carbohydrate-binding domain, a transmembrane portion and an intra-cellular tyrosine kinase domain (Supplementary Table S4). This gene is the ortholog (54.8% identity) of the *Arabidopsis thaliana* gene coding protein *At2g41890* assigned to the GO biological process of pollen recognition and self-incompatibility. A blast analysis revealed that the *Ma* and lectin kinase loci existed separately in Rosales species but only co-localized in *Prunoideae* species as there are only found physically close to each other in *Malus*, *Pyrus* and *Prunus* genomes (Supplementary Table S6).

Immediately after the kinase, we found a TIR domain-only protein (*RMjaGC1*). The analysis of the immediate downstream sequence revealed a disrupted NBS domain followed by parts of retrotransposons suggesting a progressive gene loss. Except the *RMjaGC2* gene, the candidate interval then only contains a last TNL, *RMjaGC3*. This gene is predicted to be split into a TIR domain-only protein followed by NBS-LRR protein in which, the LRR domain is truncated (232 aa) due to a frameshift mutation inducing a premature stop codon. Furthermore, a retrotransposon fragment, present on the minus strand, disrupts its putative C-terminal part (Figure 1B).



The *RMjaGC2* gene is the sole TNL in the R candidate-region to display a complete structure with full-length domains and to be constitutively transcribed. Indeed, we carried out transcript analysis in the ‘Alnem1’ R and ‘Lauranne’ S parental accessions and in several R F2 individuals. In Alnem1 as in R F2 individuals, we successfully amplified and sequenced four fragments from the *RMjaGC2* transcript, using primers designed across exons (**Supplementary Table S1** and **Figure 2C**) and cDNA from leaves and roots as illustrated in **Figure 3**. We also detected the *RMjaGC2* transcript in the single recombinant between KIN35 and EndKin markers, which was phenotypically resistant. Unlike R accessions, the same manipulation failed to detect any *RMjaGC2* transcripts in ‘Lauranne’ (S). In parallel, we failed to amplify any transcripts of *RMjaGC1* and *RMjaGC3* genes in ‘Alnem1.’ A similar approach detected the transcripts of the *Ma* and *Prupe.7G065400* genes in leaves of RKN-free plants of each of the accessions *P. cerasifera* cv. ‘P.2175’ (the *Ma* donor) and *P. persica* cv. ‘Lovell,’ respectively, suggesting a constitutive expression and an absence of tissue-specificity in the three genes (**Figure 3**). Moreover, a blast analysis using the *Ma*, *RMjaGC2* and *Prupe.7G065400* transcript sequences as queries against RNAseq and ESTs data from diverse studies, confirmed the transcription of this gene in (i) roots and leaves of *P. persica* and (ii) in various cultivars from different *Prunus* species (**Supplementary Table S6**).

An *in silico* analysis predicted a regulatory region approximately 1kb upstream the start codon of the *RMjaGC2* gene while no consistent *cis*-element binding sites were detected in the vicinity of the *RMjaGC1* and *RMjaGC3* genes. The predicted *RMjaGC2* regulatory region appeared to be very

similar to that of peach but quite different from that of plum where multiple regulatory regions are detected (**Supplementary Figure S1**). In parallel, we screened the upstream region of the *Ma*, *RMjaGC2* and *Prupe.7G65400* genes for putative W-boxes (WRKY binding site). Two independent predictions identified an enrichment (>40%) of putative W-boxes localized upstream the *Ma* and *RMjaGC2* genes compared to the single one present upstream the *Prupe.7G65400* (**Supplementary Figure S1**). In peach, an insertion of a TNL gene (*Prupe.7G65500*) has probably replaced a part of the W-boxes in this region.

The PL Domains Contain Substantial Variations Between *RMjaGC2* and Its Orthologs

To date, the *Ma* gene is the sole gene conferring a full resistance to all RKN species tested including *M. incognita*, *M. arenaria* and *M. javanica*. It belongs to a cluster of three TNL genes (of which two display structural defects) preceded by a tyrosine kinase. The three orthologous genes (*Ma*, *RMjaGC2*, *Prupe.7G65400*) are highly similar (>96% CDS identity). While the *Ma* coding sequence displays a similar identity with peach and almond, its predicted protein shows a slightly higher identity to *RMjaGC2* (94.41%) than to *Prupe.7G065400* (94.12%) (**Table 1**).

A gene structure analysis revealed a disparity localized in the first and last intronic regions of these genes. In *Ma*, the first intron between the TIR and NBS-coding exons displays an approximate 1.5 kb fragment that is absent in both *RMjaGC2* and *Prupe.7G065400* (**Figure 4A**). Each PL-coding exon is preceded by dinucleotide SSRs implying polymorphic areas. The final intron displays a differential pattern. Indeed, we noticed that in the *RMjaGC2* and *Prupe.7G065400* genes, the final intron exhibits a duplication of a fragment containing a part of the intron and 32 nucleotides of the last exon including the splicing acceptor site (**Figure 4A**). This singularity might induce alternative splicing and/or truncated mRNA. Using selective primers, PCR experiments carried out with Alnem1 cDNA, failed to amplify any alternative transcripts and only the predicted PLA-PL5 was detected (**Figure 4A**). Nevertheless, we cannot exclude that alternative forms might exist in low numbers or might be induced under specific conditions.

At the peptide level, we identified 110 amino acid mutations that are specific to *Ma*, 12 that are specific to *RMjaGC2* and 14 that are only present in *Prupe.7G065400*. Unlike the polymorphism detected between *RMjaGC2* and

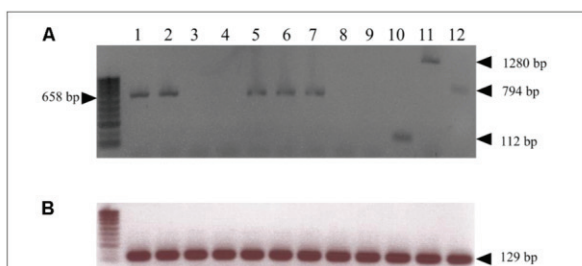


FIGURE 3 | Transcript detection of *Ma* homologous genes. **(A)** cDNA amplifications for the *Ma* orthologous genes from ‘Alnem1’ (*RMjaGC2*) (lanes 1 and 2, from leaves and roots, respectively), ‘Lauranne’ (S) (lanes 3 and 4, from leaves and roots, respectively), the ‘R recombinant’ (between markers KIN35 and EndKin) (lane 5 from leaves), ‘P.2175’ (*Ma*) (lane 6 from leaves), and ‘Lovell’ (*Prupe.7G065400*) (lane 7 from leaves) using primers ORF5F/LRRF1_RC amplifying the fragment 2 (658 bp) shown in **Figure 2C**. The corresponding genomic fragment size of *RMjaGC2* is 1042 bp. cDNA amplifications targeting *RMjaGC1* (lane 8) and *RMjaGC3* (lane 9) sequences in ‘Alnem1’ (leaves) (using the specific primers TIRRMjaGC1F/R, TIRRMjaGC3F/NBSRMjaGC3R, respectively) failed to amplify any fragments. Amplifications of the fragments 1 (TIRRMjaGC2F/NBSRMjaGC2R; 112 bp), 3 (primers EPISLRRF/ PL2F2_RC; 1280 bp) and 4 (primers PL3F/PL5R3; 794 bp) from *RMjaGC2* transcript in ‘Alnem1,’ shown in **Figure 2C** (lanes 10, 11 and 12, respectively). **(B)** cDNA amplifications of the actin gene (*Prupe.2G235000*; 129 bp) used as control.

TABLE 1 | Percent identity matrix between the three orthologous genes, CDS and proteins, *Ma*, *Prupe.7G065400* and *RMjaGC2*.

		<i>Prupe.7G065400</i>	<i>Ma</i>
Gene UTR incl.	<i>RMjaGC2</i>	98.42	91.41
	<i>Prupe.7G065400</i>		91.52
CDS	<i>RMjaGC2</i>	99.20	96.65
	<i>Prupe.7G065400</i>		96.67
Protein	<i>RMjaGC2</i>	98.78	94.41
	<i>Prupe.7G065400</i>		94.12

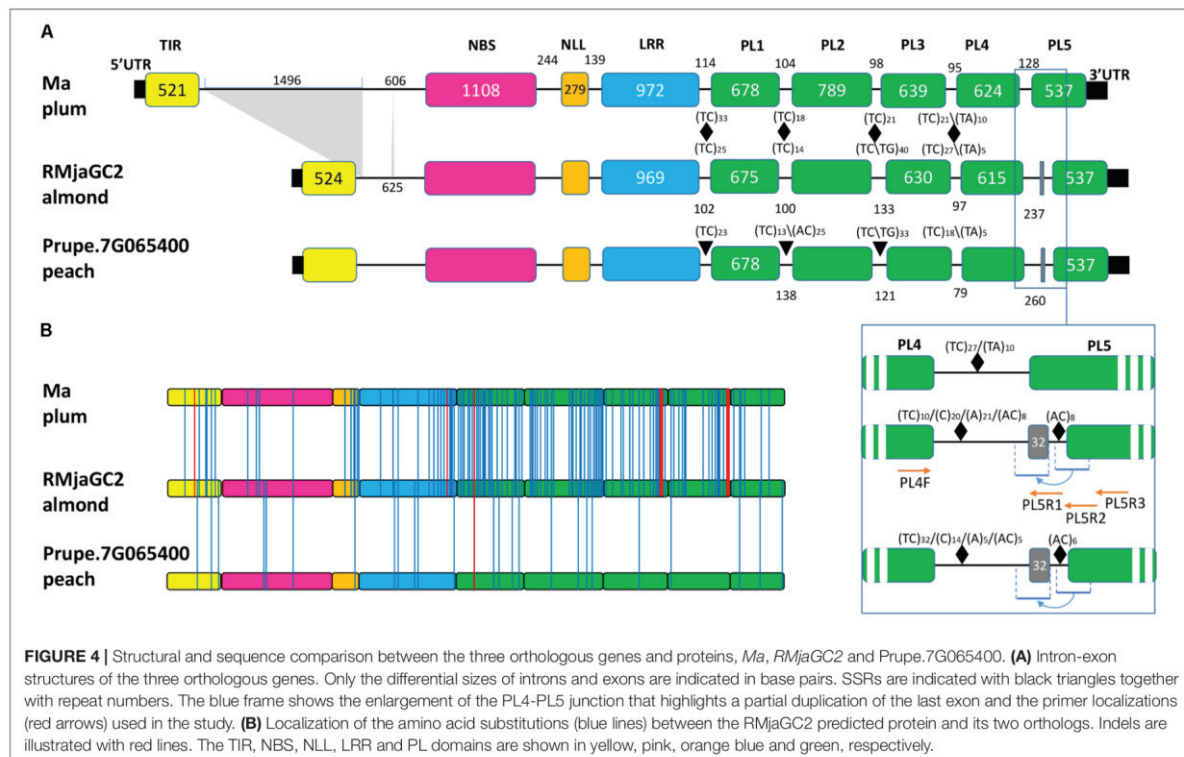


FIGURE 4 | Structural and sequence comparison between the three orthologous genes and proteins, *Ma*, *RMjaGC2* and *Prupe.7G065400*. **(A)** Intron-exon structures of the three orthologous genes. Only the differential sizes of introns and exons are indicated in base pairs. SSRs are indicated with black triangles together with repeat numbers. The blue frame shows the enlargement of the PL4-PL5 junction that highlights a partial duplication of the last exon and the primer localizations (red arrows) used in the study. **(B)** Localization of the amino acid substitutions (blue lines) between the *RMjaGC2* predicted protein and its two orthologs. Indels are illustrated with red lines. The TIR, NBS, NLL, LRR and PL domains are shown in yellow, pink, orange blue and green, respectively.

Prupe.7G065400, the mutations between *RMjaGC2* and *Ma* are unevenly distributed. Indeed, the variation are mostly assigned to specific domains or parts of domains such as the C-terminal parts of the NLL, LRR, PL2, PL3 and PL4 domains. The highest numbers of mutations are present in PL1 and PL2 and indels in PL3 and PL4 domains (Figure 4B).

We established a repertoire of mutations that are (1) only present in *Ma* (*Ma* specific) compared to both *RMjaGC2* and *Prupe.7G065400*, (2) specific to *RMjaGC2* (*RMjaGC2* specific) or (3) common to both *Ma* and *RMjaGC2* and so differing from *Prupe.7G065400* (*Ma-RMjaGC2* specific). This latter category might highlight crucial amino acid for resistance whereas the first two categories might be associated to spectrum determinants of the resistance. Each mutation is assigned to conservative, semi-conservative or non-conservative mutation subsets as such characteristics often have a differential impact on protein functioning (Table 2). In *Ma-RMjaGC2* specific mutations, we detected two non-conservative events in both PL1 (T1103R; G1150R) and PL2 (P1408Q; T1423R) domains that may influence the resistance mechanism. The non-conservative mutations that are specific to *Ma* are mostly localized in the LRR, PL1, PL2, PL3 and PL4 domains with a maximum of 12 non-conservative amino acid mutations in the PL2. *RMjaGC2* displays an alanine deletion in the PL1 and one non-conservative mutation in PL2, PL3 and PL4 (S1281L, Q1570P and T1668L, respectively) that may be in relation to its selective spectrum of resistance.

The non-synonymous (*Ka*) to synonymous (*Ks*) substitutions ratio measured on the entire CDS (0.413, 0.368, 0.278 for *Ma/Prupe.7G065400*, *Ma/RMjaGC2* and *RMjaGC2/Prupe.7G.065400*, respectively) suggest that a negative selection occurs. Nevertheless, an exon-based *Ka/Ks* ratio analysis revealed that the different domains show highly variable ratios (Figure 5). Indeed, the *Ka/Ks* ratios of the TIR, NBS and PL5 domains are $<<1$ suggesting that these domains are under strong negative selection. Conversely, the PL1 and PL2 domains display *Ka/Ks* ratios $>>1$ suggesting that adaptive evolution occurs in this region. The PL2 domain is atypical as the measured ratios are extremely different between the genes considered, thus preventing any definitive conclusion related to it. The remaining domains (NLL, LRR, PL3 and PL4) display variable *Ka/Ks* ratios, mainly between 0.2 and 0.7 (Figure 5).

Ma Cluster Evolution

The generation of the almond sequence, together with its syntenic regions in plum and in peach, offered us the opportunity to study the evolution of a cluster containing atypical and functional TNL genes. The pairwise analysis of the ten different TNL genes detected in the cluster revealed two conserved loci over the three species (Figure 6). The first conserved locus includes the *MaGC3*, *Prupe.7G065300* and *RMjaGC3* genes that display similar structure with intact TIR, NBS and NLL domains but a truncated LRR domain and no PL domains due to TEs insertion or frameshift mutations

TABLE 2 | Distribution of the conservative (C), semi-conservative (S) and non-conservative (N) amino acid mutations over the different domains.

Mutations	Total		TIR			NBS			NULL			LRR			PL1			PL2			PL3			PL4			PL5		
	C	S	C	S	N	C	S	N	C	S	N	C	S	N	C	S	N	C	S	N	C	S	N	C	S	N			
Ma specific	53	13	44	3	0	3	0	0	4	1	0	5	1	6	8	7	7	12	1	12	8	2	7	6	1	9	4	0	1
RMjaGC2 specific	2	5	5	0	0	1	0	0	1	0	1	2	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	1	0
RMjaGC2 & Ma specific	8	2	4	2	0	0	1	0	0	0	0	2	0	0	1	0	2	0	1	2	0	0	0	0	0	0	2	0	0

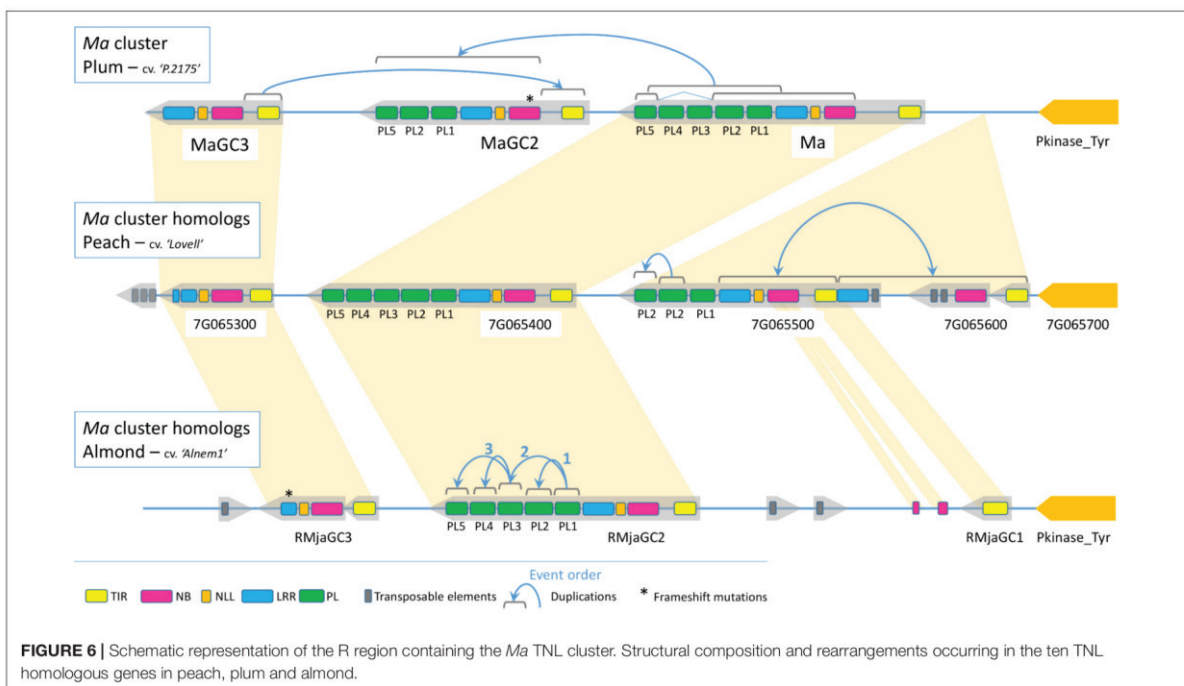
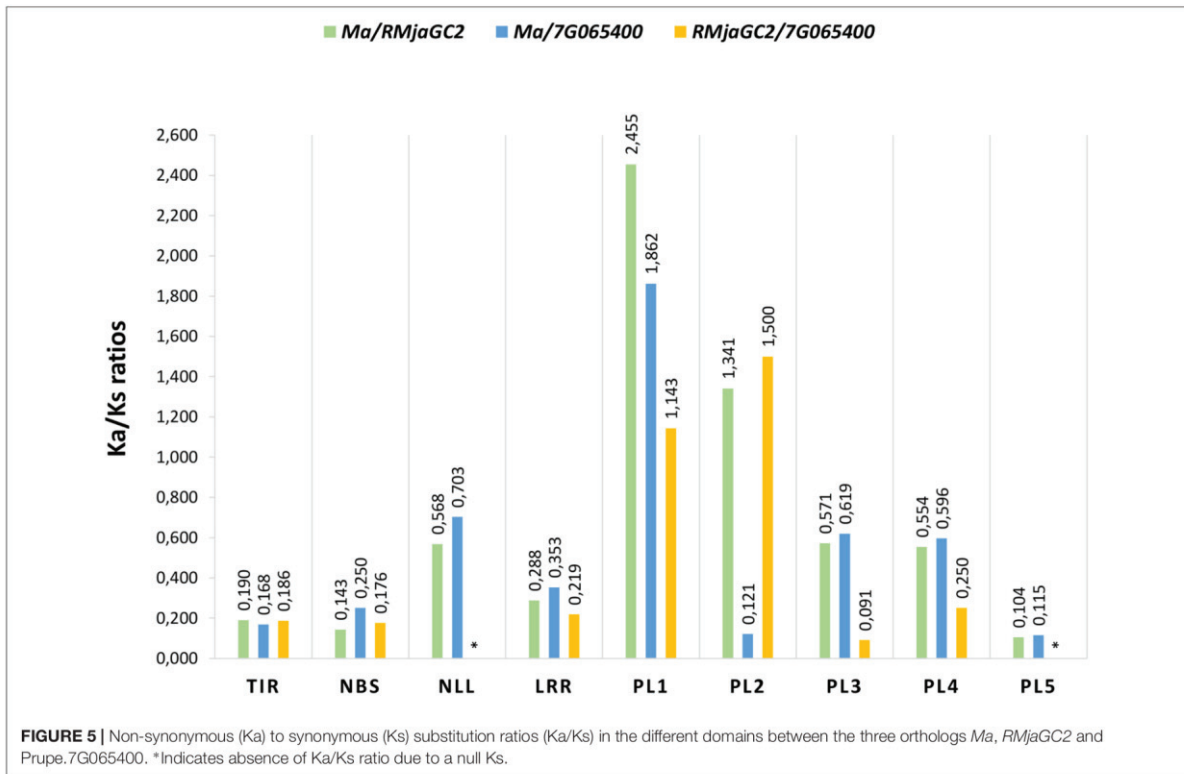
inducing premature stop codons. The second conserved locus includes the *Ma*, *RMjaGC2* and *Prupe.7G065400* genes, characterized by a complex structure. The *Prupe.7G065500*, *Prupe.7G065600* and *RMjaGC1* genes are related to each other and underwent severe reorganizations including TEs insertions or duplications leading to chimeric structures. They emerged either from the *Prunus* ancestor and have been lost in plum, or they are posterior to the split between plums and *Amygdalus*.

In plum, pairwise and phylogenetic analyses of the diverse domains revealed that the *MaGC2* gene recently emerged from partial duplications of the *Ma* and *MaGC3* genes with internal rearrangements. Indeed, the reorganization involves a TIR domain originating from *MaGC3*, and a remaining sequence originating from the *Ma* gene except two lost exons (PL3 and PL4). In *RMjaGC2* and its orthologs, the PL1 domain is the closest PL domain to the outgroup (the PL domain of the tobacco *N* gene). It probably duplicated to give birth to PL2, which then diverged. PL3, PL4 and PL5 domains display a significant identity between each other and are closer to PL1 than the PL2 domains (Figure 6 and Supplementary Figure S2). These results suggest that the PL1 domain underwent exon duplication to generate (i) PL2, a long time ago, and (ii) one of PL3, PL4 or PL5, which rapidly duplicated itself to produce this unique structure.

DISCUSSION

The *RMja* Gene Is Presumably the *Ma* Ortholog

Our high-resolution mapping together with the peach genome and *Ma* gene resources helped us to precisely locate the *RMja* gene. This approach, based on a large progeny, points out a limited window in which very few predicted ORFs are credible candidate genes to confer such a resistance. In ‘Alnem1,’ the precise annotation of the *RMjaGC2* gene and its environment together with expression data strongly suggest that the *RMjaGC2* gene is *RMja* and so confers resistance to *M. javanica*. Nevertheless, the poor *Prunus* ability to transformation, the peculiar recalcitrance of almond to produce roots spontaneously together with the large size of *RMjaGC2* are so many obstacles to engineer *RMja*-transformed roots and so validate the *RMjaGC2* gene. Its transfer in species with a more favorable genetic background would be an alternative. In this perspective, preliminary works to transfer *Ma* in tomato failed to provide any resistance, even though a constitutive transcription has been detected (H. Duval, unpublished). Missing components in *Solanaceae* probably explains the unsuccessful transfer of resistance, which seriously restrains this approach at least in these models. The dual transfer of *Arabidopsis* NBS-LRR genes, *RRS1* and *RPS4*, circumvented the restricted taxonomic functionality (Tai et al., 1999) in *Solanaceae* species (Narusaka et al., 2013). In the future, the identification of a *Ma/RMja* partner could help consequently to break down the technical barriers that are obstructing deeper functional analyses.



Insight Into the Molecular Determinants Involved in RKN Resistance

The *Ma* gene is not circumvented so far and confers resistance to mitotic species *M. incognita*, *M. javanica*, *M. arenaria*, *M. ethiopica*, *M. enterolobii* and the meiotic *M. floridensis* (Claverie et al., 2011). A recent study showed that *Ma* and *RMja* resistance spectra are closer than expected. Indeed, the sole species uncontrolled by *RMja* (*M. incognita* and *M. floridensis*) (D. Esmenjaud, unpublished) are also close in the phylogenetic tree of the *Meloidogynidae* family (Holterman et al., 2009). The weak polymorphism between *RMjaGC2* and its peach ortholog suggests that few mutations may confer resistance to most RKNs. The differential specificity between *Ma* and *RMja* may be due either to the gain of recognition of a ubiquitous effector from *Meloidogyne* spp. in *Ma* or to the loss of recognition of an effector specific to *M. incognita* and *M. floridensis* in *RMja*. In the case of a direct recognition of effectors by NBS-LRRs, few mutations may lead to extended recognition specificity or, on the contrary, to a failure of recognition due to a rapid co-evolutionary arms race (Karasov et al., 2014). In this context, we should observe among *Prunus* accessions a panel of alleles conferring variable resistance interactions to RKNs for species specificity and/or resistance strength.

Structurally, the three genes *RMjaGC2*, *Ma* and *Prupe.7G065400* display an identical repetition of five PL domains. The cost for the plant to produce and conserve such massive C-terminal ends in *Prunus* species is in line with an essential role of these repeats. PLs may function as baits mimicking more or less fitted RKN effector targets and thus may reveal quite original immune mechanisms in woody perennial plants. Screening identified RKN effectors [reviewed with their potential targets in Truong et al. (2015)] for *Ma* recognition is a strategy that should help to clarify the RKN resistance mechanism in *Prunus*. Even though the role of the PL domain is still unclear, mutation in a conserved PL-domain motif (i.e., the second motif defined in Van Ghelder and Esmenjaud, 2016) of the Arabidopsis *RPS4* gene, impaired the dual-*RPS4/RRS1* mediated resistance suggesting a role of this domain in immune signaling and/or partner interactions (Sohn et al., 2014).

Position and Role of the 5-PL-Containing Genes in the Resistance Mechanism

Ma and *RMja* are single homologous genes conferring resistance to different RKN species. Nevertheless, other partners could also be actively involved in this process. The orthologous genes *Rpi-blb2* and *Mi-1.2* from potato and tomato, respectively, conferred resistance to certain RKNs, aphids and white flies (van der Vossen et al., 2005). Recent studies showed that these two genes required other NBS-LRRs (NLR required for cell death-helpers or NRC-helpers) to function. Besides, transfers of diverse tomato NRC-helpers together with *Rpi-blb2/Mi* in NRC-silenced *N. benthamiana* showed contrasting rescues of *Rpi-blb2/Mi*-mediated cell death (Wu et al., 2017). This highlights the influence of the NBS-LRR sensor-helper affinity to transfer

successfully a complete resistance and interrogates about a model in which *Ma*, *RMja* or *Prupe.7G065400* sensors would confer distinct resistance spectra in accordance with a divergent affinity between sensors and helper(s) in plum, almond, and peach.

As crucial elements of PTI, kinase proteins are effector targets, which will modify and alter kinases activity favoring pathogen growth and development. The conserved lectin S-domain receptor kinase in the R region, which mostly segregates with *Ma* orthologs, is classified within the *Galanthus nivalis* agglutinin (GNA) homologs (Eggermont et al., 2017). Some members are known to be upregulated in presence of pathogens while others are presumably involved in calcium influx, which is one of the PTI characteristics. At the same time, many targets of RKN effectors are unknown, even in the well-described *M. incognita* calreticulin effector Mi-CRT, which is crucial for infection by possibly altering calcium homeostasis (Jaouannet et al., 2013). In this perspective, the lectin kinase might be targeted by RKN effectors and monitored by *Ma/RMja* as illustrated in the kinase/NBS-LRR (PBS1/ RPS5; Pto/Prf) models in which kinase modifications induced by bacterial effectors are detected by NBS-LRRs inducing immune responses (Zhang et al., 2010; Ntoukakis et al., 2014). Interestingly, the CNL *Prf* also lies in the middle of the kinase *Pto* cluster (Salmeron et al., 1996). As, effectors can target different kinases, some NBS-LRRs can monitor multiple kinases (Lewis et al., 2013). Unlike direct recognitions, the indirect strategy potentially provides a wider protection against pathogens (Cesari, 2017) and may be favored in long-lived organisms. Guard components will be potentially modified by effectors in different ways (e.g., phosphorylation, acetylation, cleavage, etc.) and consequently detected by NBS-LRRs proteins as illustrated by the RIN4 guardee/RPS2-RPM1 CNLs model (Mackey et al., 2003). In this case, the only way for the pathogen to escape recognition is a loss of function of the effector, which may be insurmountable and lead to durable resistance.

Expression and Regulatory Region

The predicted regulatory region of *Ma* is large and complex. Nevertheless, the transfer of *Ma* together with the 5.4 kb region localized immediately upstream the ATG in susceptible plants, was shown to be sufficient to confer the resistance (Claverie et al., 2011). The pattern of *cis*-acting binding sites is very similar between *RMjaGC2* and *Prupe.7G065400* putative promoter regions but quite different from the *Ma* region. The WRKY transcription factors, which bind to *cis*-regulatory W-boxes, respond to activation of abscisic acid-dependent pathway and belong to SA signaling that are both linked to NBS-LRR regulation (Eulgem et al., 2000). We found more W-boxes in the *Ma* regulatory region than in the *RMjaGC2* region. The *Prupe.7G065400* upstream region is clearly the least endowed with W-boxes. The insertion of the *Prupe.7G065500* gene that does not exist in plum and almond might disturb the regulatory region of the downstream *Prupe.7G065400*. Beyond the expression level, the sensitivity of the ON/OFF equilibrium of NBS-LRRs plays a crucial role to trigger or not the resistance. Indeed the switch toward an ON state of

the *L6/L7* alleles of the *L* locus is governed by the ligand affinity and the intramolecular constraints existing between TNL domains (Bernoux et al., 2016). These attributes defined how NBS-LRRs may cross the threshold conferring them the ability to produce, or not, cell death and/or resistance (Zhang et al., 2017).

We showed that *Ma* orthologs exhibit various and sizeable intronic microsatellites upstream each PL-coding exon with a particular feature in the last intron (PL4-PL5). This intriguing structural characteristic is present in 60% of the PL-containing TNLs in peach (Van Ghelder and Esmenjaud, 2016). We do not know yet if the presence of SSRs exclusively before PL-coding exons is only a consequence of genetic processes or if this structure plays a role in the modulation of TNL expression or splicing. Besides, several studies established a direct link between microsatellite size in intron and certain human diseases (Kerstring et al., 2008; Suryadevara et al., 2013). Although the precise mechanisms are not fully understood, the microsatellite expansion seems to disrupt the targeted gene transcription (Bidichandani et al., 1998).

The R Region Is a Combination of Dynamic and Preserved Loci

Transposable elements (TEs) are key factors in the birth-and-death model of R genes (Michelmore and Meyers, 1998). The increase of the KIN35-LRR25 interval size between peach and almond is due to TEs insertion. TEs disrupted the TNL environment inducing premature ends (*Prupe.7G065300*, *RMjaGC3*), quasi-total disruption (*RMjaGC1*) or NBS-LRR split (*Prupe.7G065500* - *Prupe.7G065600*) as also reported for *RMia* candidate genes in peach (Duval et al., 2014). The lectin tyrosine kinase is a highly preserved milestone contrasting with NBS-LRR cluster dynamism. Interestingly the kinase ortholog in *Arabidopsis* (*At2G41890*) lies in a region of the chromosome 2 that lacks the TNL cluster as the co-localization of the kinase and the *Ma* loci emerged in the *Prunoideae*. Our cluster study revealed two conserved TNL loci and other TNLs originated, disrupted or disappeared through recombination events. Evolution studies of R genes in the *Arabidopsis* genome proposed segmental and ectopic duplication as mode of NBS-LRR dispersion (Baumgarten et al., 2003; Leister, 2004). A ratio estimation of each mechanisms in three legume genomes revealed that local tandem duplication is the main mode of expansion (>75%) before ectopic or segmental duplications (Shao et al., 2014). Regardless of the type of mechanism, NBS-LRR duplications induce phenotypic variations through the diversification of sequences, and therefore in potential functionality, but also in gene expression (Wang et al., 2012). As illustrated in the *Ma* cluster, the modular structure of TNL genes together with a combination of partial duplications and deletions shape the TNL diversity. The chimeric *MaGC2* gene originated from an incomplete gene duplication of the *Ma* and *MaGC3* genes. Interestingly this gene includes a frameshift mutation close to the recombination site, making it prematurely truncated and perhaps preventing deleterious effects due to

auto activity. In this cluster, PL-coding exons are prone to intragenic duplications creating high sequence diversity. While *Prupe.7G065500* underwent PL2 duplication, the *Ma* orthologs underwent PL1 duplication followed by rapid PL3-PL4-PL5 duplications.

Finally, our mapping pointed out a restricted region conferring resistance to *M. javanica* in almond. The functional annotation, the genetic resources related to *Ma* and the peach genome, together with the transcript expression analysis strongly support that the *RMja* gene is the *Ma* ortholog. In a dynamic environment, the high conservation of the three orthologs, either conferring a complete-spectrum resistance (*Ma/plum*), or conferring an incomplete-spectrum resistance (*RMjaGC2/almond*), or conferring no resistance (*Prupe.7G065400/peach*), argues for a major role of these genes. Our study also identified the highest gene polymorphism in the atypical PL region that might be involved in the RKN resistance spectrum. These results pave the way to reveal the role of the PL domain and a putative original immune process for RKN control in *Prunus*.

DATA AVAILABILITY STATEMENT

All relevant data is contain in the manuscript or in **supplementary material**, except the almond genome sequencing reads and the BAC clone sequences.

These datasets are publicly available in the NCBI BioProject accession number [PRJNA448729] [<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA448729>] (BioSamples numbers [SAMN08863554] [<https://www.ncbi.nlm.nih.gov/biosample/SAMN08863554/>] and [SAMN08863555] [<https://www.ncbi.nlm.nih.gov/biosample/SAMN08863555/>]) and in the NCBI BioProject accession number [PRJNA448736] [<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA448736>], respectively.

AUTHOR CONTRIBUTIONS

HD, DE, and CVG designed the research project. HD, DE, MM, and CVG carried out and supervised the greenhouse and laboratory works. ED supervised and carried out the sequencing of almond genomes and produced the alignment files. CC supervised and carried out the almond genomic library, selected BAC clones and produced BAC clone sequences. HD supervised the data releases, CVG planned and designed the sequence analyses, analyzed the data, and drafted the manuscript. CVG and DE revised the manuscript. All authors commented and approved the manuscript.

FUNDING

This work was supported by the French Ministry of Agriculture and Food [CASDAR project 'Pyrédune' n°C2011-01

(2011–2014)], the INRA BAP department (projects IVD2 and IVD3), the Plant-Nematode Interaction team (UMR ISA, Sophia Antipolis, France), and the GAFL Unit (INRA Montfavet, France).

ACKNOWLEDGMENTS

The authors want to thank Emmanuel Botton, Laure Heurtevin, Nadia Therme, Ariane Fazari, and Ulysse Portier for their

contribution in tomato regeneration, molecular biology support, and phenotyping of almond progenies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01269/full#supplementary-material>

REFERENCES

- Baumgarten, A., Cannon, S., Spangler, R., and May, G. (2003). Genome-level evolution of resistance genes in *Arabidopsis thaliana*. *Genetics* 165, 309–319.
- Bernoux, M., Burdett, H., Williams, S. J., Zhang, X., Chen, C., Newell, K., et al. (2016). Comparative analysis of the flax immune receptors L6 and L7 suggests an equilibrium-based switch activation model. *Plant Cell* 28, 146–159. doi: 10.1105/tpc.15.00303
- Bidichandani, S. I., Ashizawa, T., and Patel, P. I. (1998). The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am. J. Hum. Genet.* 62, 111–121. doi: 10.1086/301680
- Blok, V. C., Jones, J. T., Phillips, M. S., and Trudgill, D. L. (2008). Parasitism genes and host range disparities in biotrophic nematodes: the conundrum of polyphagy versus specialisation. *Bioessays* 30, 249–259. doi: 10.1002/bies.20717
- Burch-Smith, T. M., Schiff, M., Caplan, J. L., Tsao, J., Czymbek, K., and Dinesh-Kumar, S. P. (2007). A novel role for the TIR domain in association with pathogen-derived elicitors. *PLoS Biol.* 5:e68. doi: 10.1371/journal.pbio.0050068
- Cesari, S. (2017). Multiple strategies for pathogen perception by plant immune receptors. *New phytol.* doi: 10.1111/nph.14877 [Epub ahead of print].
- Chalhoub, B., Belcram, H., and Caboche, M. (2004). Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size. *Plant Biotechnol. J.* 2, 181–188. doi: 10.1111/j.1467-7652.2004.00065.x
- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Chin, S. W., Shaw, J., Haberle, R., Wen, J., and Potter, D. (2014). Diversification of almonds, peaches, plums and cherries – Molecular systematics and biogeographic history of *Prunus* (*Rosaceae*). *Mol. Phylogenet. Evol.* 76, 34–48. doi: 10.1016/j.ympev.2014.02.024
- Claverie, M., Dirlewanger, E., Bosselut, N., Van Ghelder, C., Voisin, R., Kleinhentz, M., et al. (2011). The *Ma* gene for complete-spectrum resistance to *Meloidogyne* species in *Prunus* is a TNL with a huge repeated C-terminal post-LRR region. *Plant Physiol.* 156, 779–792. doi: 10.1104/pp.111.176230
- Duval, H., Hoerter, M., Polidori, J., Confolent, C., Masse, M., Moretti, A., et al. (2014). High-resolution mapping of the *RMia* gene for resistance to root-knot nematodes in peach. *Tree Genet. Genomes* 10, 297–306. doi: 10.1007/s11295-013-0683-z
- Eggermont, L., Verstraeten, B., and Van Damme, E. J. M. (2017). Genome-wide screening for lectin motifs in *Arabidopsis thaliana*. *Plant Genome* 10. doi: 10.3835/plantgenome2017.02.0010
- Elling, A. A. (2013). Major emerging problems with minor *Meloidogyne* species. *Phytopathology* 103, 1092–1102. doi: 10.1094/PHYTO-01-13-0019-RVW
- Esmenjaud, D., Voisin, R., Van Ghelder, C., Bosselut, N., Lafargue, B., Di Vito, M., et al. (2009). Genetic dissection of resistance to root-knot nematodes *Meloidogyne* spp. in plum, peach, almond, and apricot from various segregating interspecific *Prunus* progenies. *Tree Genet. Genomes* 5, 279–289. doi: 10.1007/s11295-008-0173-x
- Eulgem, T., Rushton, P. J., Robatzek, S., and Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5, 199–206. doi: 10.1016/S1360-1385(00)01600-9
- Frith, M. C., Hansen, U., and Weng, Z. (2001). Detection of cis-element clusters in higher eukaryotic DNA. *Bioinformatics* 17, 878–889. doi: 10.1093/bioinformatics/17.10.878
- Gassmann, W., Hinsch, M. E., and Staskawicz, B. J. (1999). The *Arabidopsis* *RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *Plant J.* 20, 265–277. doi: 10.1046/j.1365-313X.1999.00600.x
- Gonthier, L., Bellec, A., Blassiau, C., Prat, E., Helmstetter, N., Rambaud, C., et al. (2010). Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., *Asteraceae*). *BMC Res. Notes* 3:225. doi: 10.1186/1756-0500-3-225
- Heath, M. C. (2000). Hypersensitive response-related death. *Plant Mol. Biol.* 44, 321–334. doi: 10.1023/A:1026592509060
- Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27, 297–300. doi: 10.1093/nar/27.1.297
- Holterman, M., Karssen, G., van den Elsen, S., van Megen, H., Bakker, J., and Helder, J. (2009). Small subunit rDNA-based phylogeny of the *Tylenchida* sheds light on relationships among some high-impact plant-parasitic nematodes and the evolution of plant feeding. *Phytopathology* 99, 227–235. doi: 10.1094/PHYTO-99-3-0227
- Janati, A., Berge, J. B., Triantaphyllou, A. C., and Dalmasso, A. (1982). Nouvelles données sur l'utilisation des isoestérases pour l'identification des *Meloidogyne*. *Fundam. Appl. Nematol.* 5, 147–154.
- Jaouannet, M., Magliano, M., Arguel, M. J., Gourgues, M., Evangelisti, E., Abad, P., et al. (2013). The root-knot nematode calreticulin Mi-CRT is a key effector in plant defense suppression. *Mol. Plant Microbe Interact.* 26, 97–105. doi: 10.1094/MPMI-05-12-0130-R
- Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982
- Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286
- Jones, J. T., Haegeman, A., Danchin, E. G. J., Gaur, H. S., Helder, J., Jones, M. G. K., et al. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol. Plant Pathol.* 14, 946–961. doi: 10.1111/mpp.12057
- Karasov, T. L., Kniskern, J. M., Gao, L., DeYoung, B. J., Ding, J., Dubiella, U., et al. (2014). The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature* 512, 436–440. doi: 10.1038/nature13439
- Kersting, C., Agelopoulos, K., Schmidt, H., Korsching, E., August, C., Gosheger, G., et al. (2008). Biological importance of a polymorphic CA sequence within intron 1 of the epidermal growth factor receptor gene (EGFR) in high grade central osteosarcomas. *Genes Chromosomes Cancer* 47, 657–664. doi: 10.1002/gcc.20571
- Khallouk, S., Voisin, R., Portier, U., Polidori, J., Van Ghelder, C., and Esmenjaud, D. (2013). Multiyear evaluation of the durability of the resistance conferred by *Ma* and *RMia* genes to *Meloidogyne incognita* in *Prunus* under controlled conditions. *Phytopathology* 103, 833–840. doi: 10.1094/PHYTO-09-12-0228-R
- Kochba, J., and Spiegel-Roy, P. (1975). Inheritance of resistance to the root-knot nematode (*Meloidogyne javanica* Chitwood) in bitter almond progenies. *Euphytica* 24, 453–457. doi: 10.1007/BF00028213
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends Genet.* 20, 116–122. doi: 10.1016/j.tig.2004.01.007

- Lewis, J. D., Lee, A. H., Hassan, J. A., Wan, J., Hurley, B., Jhingree, J. R., et al. (2013). The *Arabidopsis* ZED1 pseudokinase is required for ZAR1-mediated immunity induced by the *Pseudomonas syringae* type III effector HopZ1a. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18722–18727. doi: 10.1073/pnas.1315520110
- Luo, S., Zhang, Y., Hu, Q., Chen, J., Li, K., Lu, C., et al. (2012). Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. *Plant Physiol.* 159, 197–210. doi: 10.1104/pp.111.192062
- Mackey, D., Belkadir, Y., Alonso, J. M., Ecker, J. R., and Dangl, J. L. (2003). *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell* 112, 379–389. doi: 10.1016/S0092-8674(03)00040-0
- Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., and Young, N. D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20, 317–332. doi: 10.1046/j.1365-313X.1999.00606.x
- Michelmore, R. W., and Meyers, B. C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* 8, 1113–1130. doi: 10.1101/gr.8.11.1113
- Milligan, S. B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P., and Williamson, V. M. (1998). The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *Plant Cell* 10, 1307–1319. doi: 10.2307/3870642
- Narusaka, M., Kubo, Y., Hatakeyama, K., Imamura, J., Ezura, H., Nanasato, Y., et al. (2013). Breaking restricted taxonomic functionality by dual resistance genes. *Plant Signal. Behav.* 8:e24244. doi: 10.4161/psb.24244
- Nicol, J. M., Turner, S. J., Coyne, D. L., den Nijs, L., Hockland, S., and Maafi, Z. T. (2011). "Current nematode threats to world agriculture," in *Genomics and Molecular Genetics of Plant-Nematode Interactions*, eds J. T. Jones, G. Gheysen, and C. Fenoll (Heidelberg: Springer), 21–44. doi: 10.1007/978-94-007-0434-3_2
- Ntoukakis, V., Saur, I. M., Conlan, B., and Rathjen, J. P. (2014). The changing of the guard: the Pto/Prf receptor complex of tomato and pathogen recognition. *Curr. Opin. Plant Biol.* 20, 69–74. doi: 10.1016/j.pbi.2014.04.002
- Paal, J., Henselewski, H., Muth, J., Meksem, K., Menéndez, C. M., Salamini, F., et al. (2004). Molecular cloning of the potato *Gro1-4* gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. *Plant J.* 38, 285–297. doi: 10.1111/j.1365-313X.2004.02047.x
- Peterson, D. G., Tomkins, J. P., Frisch, D. A., Wing, R. A., and Paterson, A. H. (2000). Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. *J. Agric. Genomics* 5, 1–100.
- Rairdan, G. J., and Moffett, P. (2006). Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation. *Plant Cell* 18, 2082–2093. doi: 10.1105/tpc.106.042747
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754
- Salmeron, J. M., Oldroyd, G. E., Rommens, C. M., Scofield, S. R., Kim, H. S., Lavelle, D. T., et al. (1996). Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. *Cell* 86, 123–133. doi: 10.1016/S0092-8674(00)80083-5
- Saucet, S. B., Van Ghelder, C., Abad, P., Duval, H., and Esmenjaud, D. (2016). Resistance to root-knot nematodes *Meloidogyne* spp. in woody plants. *New Phytol.* 211, 41–56. doi: 10.1111/nph.13933
- Shao, Z. Q., Zhang, Y. M., Hang, Y. Y., Xue, J. Y., Zhou, G. C., Wu, P., et al. (2014). Long-term evolution of nucleotide-binding site-leucine-rich repeat genes: understanding gained from and beyond the legume family. *Plant Physiol.* 166, 217–234. doi: 10.1104/pp.114.243626
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Sohn, K. H., Segonzac, C., Rallapalli, G., Sarris, P. F., Woo, J. Y., Williams, S. J., et al. (2014). The nuclear immune receptor RPS4 is required for RRS1SLH1-dependent constitutive defense activation in *Arabidopsis thaliana*. *PLoS Genet.* 10:e1004655. doi: 10.1371/journal.pgen.1004655
- Solovyev, V., Kosarev, P., Seledsov, I., and Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 1, 10.1–10.12. doi: 10.1186/gb-2006-7-s1-s10
- Suryadevara, N. C., Neela, V. S., Devalaraju, K. P., Jain, S., SivaSai, K. S., Valluri, V. L., et al. (2013). Influence of Intron II microsatellite polymorphism in human toll-like receptor 2 gene in leprosy. *Hum. Immunol.* 74, 1034–1040. doi: 10.1016/j.humimm.2013.04.009
- Tai, T. H., Dahlbeck, D., Clark, E. T., Gajiwala, P., Pasion, R., Whalen, M. C., et al. (1999). Expression of the *Bs2* pepper gene confers resistance to bacterial spot disease in tomato. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14153–14158. doi: 10.1073/pnas.96.24.14153
- Tong, Z., Qu, S., Zhang, J., Wang, F., Tao, J., Gao, Z., et al. (2012). A modified protocol for RNA extraction from different peach tissues suitable for gene isolation and real-time PCR analysis. *Mol. Biotechnol.* 50, 229–236. doi: 10.1007/s12033-011-9433-3
- Truong, N. M., Nguyen, C. N., Abad, P., Quentin, M., and Favery, B. (2015). Function of root-knot nematode effectors and their targets in plant parasitism. *Adv. Bot. Res.* 73, 293–324. doi: 10.1016/bs.abr.2014.12.010
- van der Hoorn, R. A., and Kamoun, S. (2008). From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell* 20, 2009–2017. doi: 10.1105/tpc.108.060194
- van der Vossen, E. A., Gros, J., Sikkema, A., Muskens, M., Wouters, D., Wolters, P., et al. (2005). The *Rpi-blb2* gene from *Solanum bulbocastanum* is an *Mi-1* gene homolog conferring broad-spectrum late blight resistance in potato. *Plant J.* 44, 208–222. doi: 10.1111/j.1365-313X.2005.02527.x
- Van Ghelder, C., and Esmenjaud, D. (2016). TNL genes in peach: insights into the post-LRR domain. *BMC Genomics* 17:317. doi: 10.1186/s12864-016-2635-0
- Van Ghelder, C., Lafargue, B., Dirlwanger, E., Ouassa, A., Voisin, R., Polidori, J., et al. (2010). Characterization of the *RMja* gene for resistance to root-knot nematodes in almond: spectrum, location, and interest for *Prunus* breeding. *Tree Genet. Genomes* 6, 503–511. doi: 10.1007/s11295-010-0268-z
- Verde, I., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., et al. (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat. Genet.* 45, 487–494. doi: 10.1038/ng.2586
- Wang, Y., Wang, X., and Paterson, A. H. (2012). Genome and gene duplications and gene expression divergence: a view from plants. *Ann. N. Y. Acad. Sci.* 1256, 1–14. doi: 10.1111/j.1749-6632.2011.06384.x
- Whitham, S., Dinesh-Kumar, S. P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994). The product of the *Tobacco mosaic virus* resistance gene *N*: similarity to toll and the interleukin-1 receptor. *Cell* 78, 1101–1115. doi: 10.1016/0092-8674(94)90283-6
- Wu, C. H., Abd-El-Halim, A., Bozkurt, T. O., Belhaj, K., Terauchi, R., Vossen, J. H., et al. (2017). NLR network mediates immunity to diverse plant pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 114, 8113–8118. doi: 10.1073/pnas.1702041114
- Zhang, J., Li, W., Xiang, T., Liu, Z., Laluk, K., Ding, X., et al. (2010). Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. *Cell Host Microbe* 7, 290–301. doi: 10.1016/j.chom.2010.03.007
- Zhang, X., Dodds, P. N., and Bernoux, M. (2017). What do we know about NOD-like receptors in plant immunity? *Annu. Rev. Phytopathol.* 55, 205–229. doi: 10.1146/annurev-phyto-080516-035250

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Van Ghelder, Esmenjaud, Callot, Dubois, Mazier and Duval. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CHAPITRE 6

**DONNEES COMPLEMENTAIRES
RELATIVES AUX MECANISMES ET
DETERMINANTS DE LA RESISTANCE
CONFEREES PAR LE GENE *MA***

Chapitre 6 : DONNEES COMPLEMENTAIRES RELATIVES AUX MECANISMES ET DETERMINANTS DE LA RESISTANCE CONFEREES PAR LE GENE *MA*

6.1 Introduction

Le locus *RMja/Ma* offre une opportunité de comprendre les mécanismes et les déterminants guidant un processus de résistance particulièrement robuste vis-à-vis des nématodes à galles. En effet, l'étude des séquences du gène *Ma* chez le prunier myrobolan (*P. cerasifera*) et de ses orthologues chez l'amandier (*P. dulcis*) et le pêcher (*P. persica*) a montré une grande conservation dans leur structure canonique de type TNL mais également dans la structure originale que représentent les cinq domaines PL répétés malgré des phénotypes très différents. L'identification des transcrits et l'exploitation de données d'expression de *Ma* ainsi que de ses allèles et orthologues, issus de différents clones et/ou espèces, associés à leurs phénotypes respectifs vis-à-vis des nématodes à galles, peuvent contribuer à une meilleure compréhension des facteurs impliqués dans la résistance. Hormis le polymorphisme de séquence et la présence d'éventuelles isoformes, le niveau d'expression peut également jouer un rôle sur l'établissement de la résistance.

Dans ce paragraphe nous traiterons des résultats qui n'ont pu être intégrés aux articles présentés précédemment mais apportent un complément d'information et ouvrent des perspectives de recherches et travaux futurs dans cette thématique.

6.2 Données d'expression du gène *Ma* chez le clone P.2175 (*P. cerasifera*) et en système hétérologue

Clone P.2175. Une étude préliminaire consacrée à l'étude de l'expression du gène *Ma* a été réalisée chez le prunier myrobolan P. 2175. Une vingtaine de jeunes plants issus d'*in vitro* ont été transférés pour moitié en boîtes de Pétri sur milieu gélosé (condition *in vitro*) et pour moitié en pots contenant du sol préalablement stérilisé (condition *in vivo*). Pour chacune des conditions, 5 plants ont été utilisés comme témoins non inoculés et 5 autres ont été inoculés avec environ 200 individus de l'espèce *M.*

incognita au stade infestant J2. Après 3 jours, racines et feuilles ont été prélevées pour en extraire l'ARN et synthétiser l'ADN complémentaire pour une analyse en PCR quantitative (qPCR).

Nos résultats montrent que le gène *Ma* est exprimé de façon constitutive dans les deux conditions de culture et aussi bien en racines (Figure 22A) qu'en feuilles (témoin positif ; T+). L'expérimentation complète en condition *in vitro*, bien que plus lourde, semble être plus reproductible. Ceci s'explique par un meilleur contrôle des facteurs de l'environnement mais surtout par un meilleur état physiologique des racines. En effet les racines cultivées dans du sol initient plus précocement qu'en condition *in vitro*, une lignification des tissus et synthétisent des composants réduisant l'efficacité des extractions d'acides nucléiques particulièrement délicate chez les *Prunus*. L'expression du gène *Ma* chez les plants non inoculés (expression basale) semble être relativement élevée pour un NBS-LRR (Figure 22B) et mériterait d'être comparée à l'expression d'autres NBS-LRRs chez ce clone. En effet les Ct mesurés pour *Ma* (Ct= 27.41 ; σ =1.06), bien que plus élevés que ceux mesurés pour l'actine (Ct= 23.97 ; σ =1.10), indiquent que les transcrits sont aisément détectables. Par ailleurs, la quantification relative entre témoins et plants inoculés ne semble pas indiquer d'expression différentielle. Après trois jours, le « Fold Change » (FC) mesuré n'est pas significativement différent de 1, quelle que soit la condition de culture (Figure 22C). Ce temps de trois jours a été choisi du fait (i) que les J2 pénètrent dans les racines dès les premières heures de l'infestation et (ii) qu'avant 2 jours aucune différence phénotypique dans le processus d'infestation n'est relevée entre plants résistants porteurs du gène *Ma* et plants sensibles non porteurs du gène. Ces éléments suggèrent qu'une différence d'expression du gène *Ma* pourrait être mise en évidence au-delà de 2 jours. Bien que préliminaires, ces travaux soulignent le caractère constitutif et le niveau significatif de transcription du gène *Ma*, ce qui laisse à penser que le gène assure un niveau de surveillance élevé au sein de la cellule. Ces premiers résultats ouvrent la voie à une étude plus large qui devra approfondir les données de cinétique d'expression sur des plants obtenus en condition *in vitro* en prenant en compte (i) des temps post-inoculation plus longs, (ii) un nombre plus élevé de réplicats et de contrôles (autres gènes du même type) et (iii) des plants de prunier myrobolan sensibles afin d'établir un éventuel différentiel entre plants R et S.

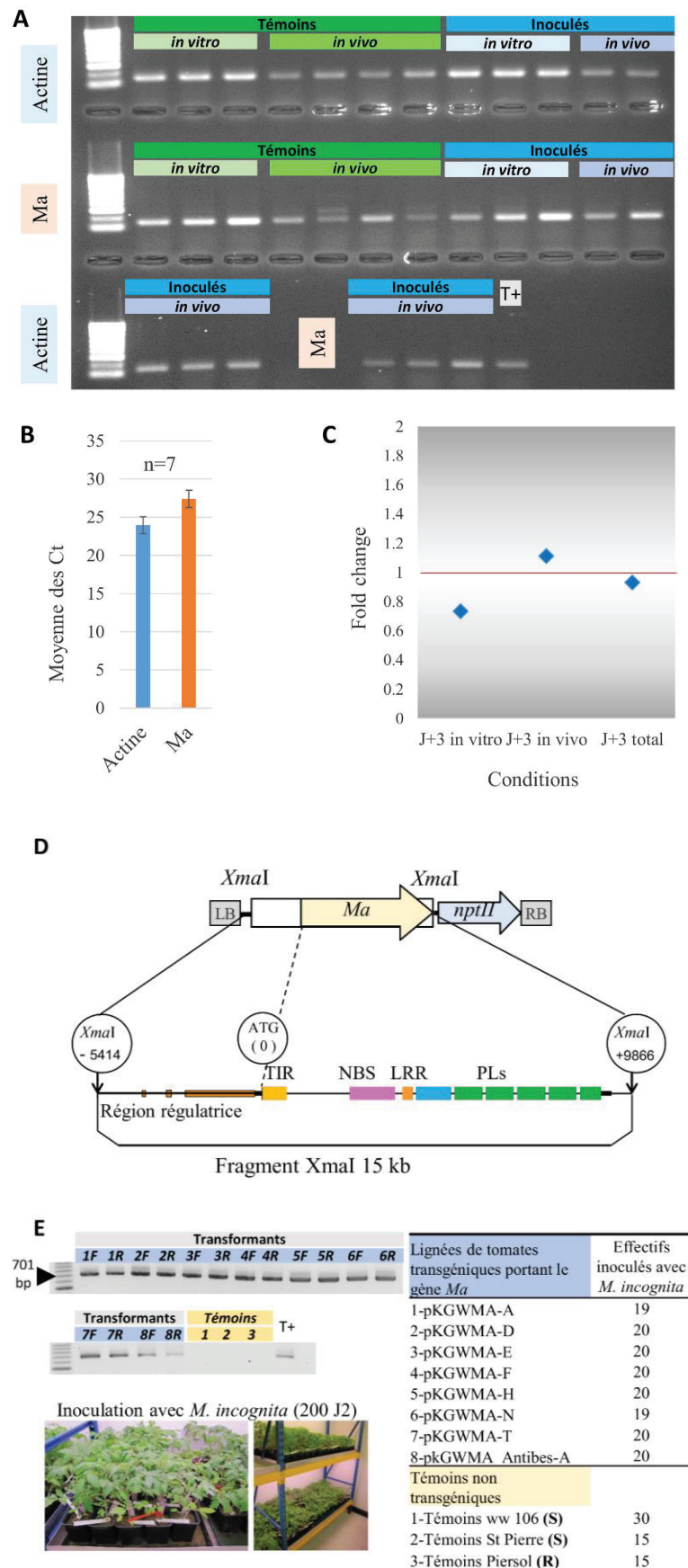


Figure 22 - Expression du gène *Ma* chez le prunier myrobolan P.2175 et en système hétérologue chez la tomate.

(A) Amplification des marqueurs utilisés en qPCR (actine, ACT235000 : 129 bp et Ma, QCT3 : 129 bp) sur ADNc issus de racines inoculées avec *M. incognita* et non inoculées. (B) Moyenne des Ct observés sur les échantillons non inoculés. (C) 'Fold change' ($=2^{-\Delta\Delta Ct}$) observé 3 jours après inoculation en conditions *in vitro* et *in vivo*. (D) Construction incluant le gène *Ma*, utilisée pour transformer des tomates sensibles aux *Meloidogyne*. (E) Effectifs des lignées de tomates (variété WVA106) transformées avec *Ma* et phénotypées vis-à-vis de *M. incognita*. L'expression du gène *Ma* est vérifiée pour chacune des 8 lignées sur feuilles et racines (1F/1R à 8F/8R) avec le marqueur spécifique CT3 (701 bp).

Système hétérologue. La création de tomates transformées avec le gène *Ma* a été réalisé par nos collègues de l'INRA d'Avignon (GAFL). Le gène *Ma* avec sa région promotrice, représentant un insert d'environ 15 kb, a été utilisé avec un vecteur classique Gateway™ pour réaliser les transformations avec *Agrobacterium tumefaciens* (Figure 22D). Huit lignées stables, représentant 8 évènements de transformation indépendants contenant une seule insertion chacune, ont été produites. Environ 20 plants par lignée, ainsi que deux variétés sensibles de tomates non transformées ('WVA106' utilisée pour les transformations et 'St Pierre') et une variété résistante portant le gène *Mi-1.2* ('Piersol'), ont ensuite été testées par notre laboratoire vis-à-vis de *M. incognita* (Figure 22E). Après deux mois, aucune des lignées testées n'a présenté de phénotype résistant à ce nématode. L'expression du gène a été vérifiée par un marqueur spécifique (CT3) à la fois sur feuilles et racines (Figure 22E). Similairement à son expression dans son fond prunier, le transcrit du gène *Ma* a pu être détecté aisément dans ces deux organes et de façon constitutive (Figure 22E).

De nombreuses hypothèses peuvent expliquer ce résultat comme l'absence ou la mauvaise association de la protéine Ma avec des partenaires moléculaires essentiels en amont ou aval de la cascade moléculaire. Du fait que la détection de la protéine Ma n'a pas été encore réalisée, une élimination post-traductionnelle est également possible. L'éloignement entre ces deux groupes taxonomiques (familles des Solanaceae pour la tomate et des Rosaceae pour les *Prunus*) peut aussi être un obstacle important.

Au cours de notre étude concernant la distribution des sous-familles de NBS-LRRs à travers les plantes, nous avons pu constater que la tomate ne possède qu'une trentaine de TNLs et que ceux-ci ne représentent que 12% environ du répertoire de NBS-LRRs chez cette dernière (Qian *et al.*, 2017). A l'inverse les *Prunus* possèdent près de 200 TNLs représentant environ 50% des effectifs des NBS-LRRs totaux (Jia *et al.*, 2015). Cette tendance, également partagée au niveau de leurs familles botaniques respectives (Solanaceae vs Rosaceae), a très certainement influencé la quantité et la diversité des partenaires privilégiés impliqués dans la voie de signalisation des TNLs (EDS1, PAD4, ...). De même, le contingent de RNLs est également réduit : 19 unités chez le pêcher contre 3 chez la tomate (Chapitre 4.2). Ces différentes pistes pourraient être intéressantes à explorer afin de mieux comprendre les mécanismes mis en jeu et contribuer à terme à la possibilité de s'affranchir de cette restriction taxonomique pour le transfert de la résistance.

6.3 Exploitation de la diversité du prunier myrobolan

Nous avons également initié une approche comparative visant à analyser les variations des séquences exprimées de *Ma* chez diverses espèces de *Prunus* mais surtout chez le prunier myrobolan au sein d'une collection de variétés établie en serre à l'ISA. Ces variétés ont été précédemment phénotypées pour leur résistance vis-à-vis des nématodes à galles (présence d'un allèle R ou S de *Ma*). Elles ont pour origine différentes régions d'Europe et du Moyen-Orient et incluent la zone environnant la Mer Noire, proche de l'ère native présumée de cette espèce et donc susceptible d'abriter un niveau élevé de diversité (**Figure 23**). Hormis ces variétés et le transcrite du gène *RMja* désormais connu, nous pourrions également nous appuyer sur l'espèce proche prunier japonais (*P. salicina*) dont l'accession 'J222' porte le gène *Rjap*, probablement orthologue du gène *Ma* et conférant une résistance à toutes les espèces de *Meloidogyne* testées. L'étude de cette diversité allélique devrait nous aider à identifier plus précisément les régions du gène indispensables au processus de résistance aux *Meloidogyne*.



Figure 23 – Carte illustrant la diversité géographique de la collection de pruniers myrobolans phénotypés vis-à-vis de *Meloidogyne* spp.

La carte indique également l'origine de *P. salicina* cv. 'J222' (Chine), porteur du gène *Rjap*, et *P. dulcis* cv. 'Alnem 1' (Israël) porteur du gène *RMja*, tous deux très probablement orthologues du gène *Ma*. Les clones de phénotype résistant sont en bleu ; ceux de phénotype sensible sont en rose. La collection contient également 4 autres variétés à l'origine inconnue (pissardi-B17, P.1254, P31 *P. cerasifera* x *P. salicina*, Korolkowii-B13). Image Google map ©.

Nous avons choisi de ne pas utiliser les séquences génomiques. En effet la taille du gène (>10 kb), la présence de séquences microsatellites dans les introns et la possible absence d'expression de certains allèles S, nous ont conduit à travailler sur les transcrits de ces allèles. La priorité étant d'obtenir des séquences fiables spécifiques de ce locus et de reconstituer le ou les allèles exprimés, des fragments d'environ 1kb amplifiés par marqueurs spécifiques et ensuite clonés ont été séquencés dans les deux sens (**Figure 24**). Les difficultés techniques les plus importantes résidaient dans la possibilité que certains marqueurs soient trop spécifiques de l'allèle *Mal* cloné et séquencé (pas d'amplification) ou ne le soient pas assez (amplification de paralogues présents dans le cluster de *Ma*).

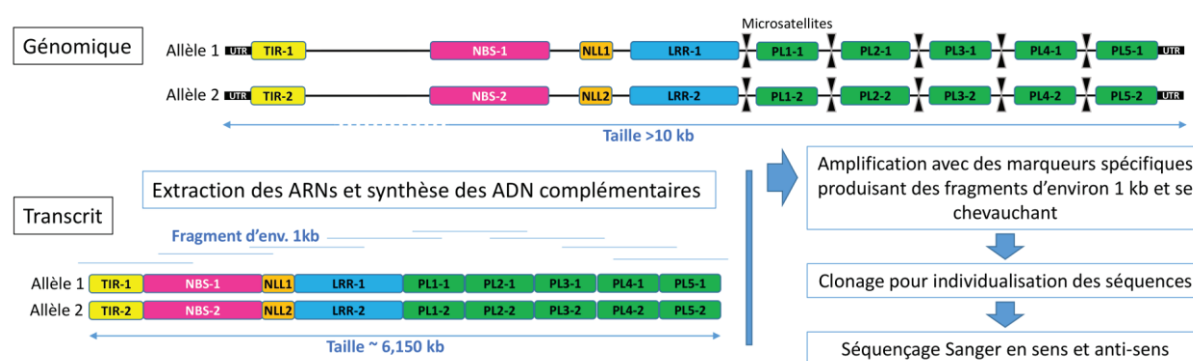


Figure 24 - Méthodologie d'analyse de la diversité allélique du gène *Ma*.

Les caractéristiques typiques des allèles au niveau génomique et transcriptomique sont indiquées avec la méthodologie d'obtention des séquences transcrites.

Les premiers résultats indiquent que toutes les variétés testées et phénotypées *R* jusqu'à présent, possèdent au moins un allèle transcrit constitutivement du gène. Le résultat est plus hétérogène chez les variétés sensibles car nous n'avons pu mettre en évidence les transcrits que chez deux des six variétés testées. Dans un premier temps, nous avons choisi de nous concentrer sur la zone que nous avons identifiée comme étant la plus polymorphe lors de l'étude du gène *RMja*, à savoir les indels situés en région C-terminale des domaines PL3 et PL4. Ces deux indels de neuf nucléotides chacun, sont situés à proximité des sites donneurs d'épissage en position C-terminale des exons codant pour les domaines PL3 et PL4. Chez les huit variétés résistantes testées, nous n'avons mis en évidence qu'un seul et même allèle de résistance ; celui-ci est sans délétion dans les domaines PL3 et PL4 et identique à celui exprimé chez P.2175 (**Figure 25**). Il est intéressant de noter que, malgré son origine éloignée, le clone J222 (*P. salicina*), de spectre de résistance équivalent à celui conféré par *Ma*, ne semble également exprimer que cet allèle.

A l'inverse, les variétés sensibles présentent des profils plus diversifiés avec tous les cas représentés (aucun ou un ou deux allèles exprimés). Si le résultat concernant les allèles résistants suggère que la présence des PL3 et PL4 'sans délétion' est associée au large spectre de résistance conféré par *Ma*, d'autres hypothèses doivent être mises en avant pour expliquer les données acquises chez les variétés sensibles. Seule l'obtention des séquences intégrales chez ces variétés complétées éventuellement par leur profil d'expression pourra permettre d'interpréter ces résultats. Chez les pêcheurs 'Lovell' (variété

utilisée pour le génome du pêcher) et 'Nemared' (variété portant le gène *RMia* situé sur le chromosome 2), seuls des allèles contenant les délétions dans les PL3 et 4 sont présents. Mais dans ces deux variétés homozygotes au locus *Ma*, une isoforme contenant une délétion plus importante dans le PL3 (**Figure 25**) a également été mise en évidence par PCR et séquençage, et confirmée par l'analyse de 'reads' générés sur le pêcher 'Lovell' par RNAseq (SRP013437). Il est possible que la délétion située dans le PL3 proche du site donneur d'épissage puisse être liée à l'apparition de cette isoforme absente chez P.2175. De plus, l'alignement des protéines prédites *Ma*, *RMja* et Prupe.7G65400 avec les structures secondaires prédites montre que la zone localisée en position C-terminale du PL3, abritant l'indel de 3 acides aminés Val-Asn-Pro (V1007-N1008-P1009) ainsi que plusieurs mutations (N1012H, H1014P, E1020K), est également le siège d'une structure secondaire prédite en hélice α (**Annexe 5**).

Par ailleurs, le séquençage des génomes de diverses variétés de pêcheurs (Pamir, Montclar, Zephir, Wfp, Malokonare, Lovell, Summergrand et Nemared) indique un très faible polymorphisme du locus *Ma* (identité > 99.87%) (Henri Duval, GAFL INRA, com. pers.). Cette caractéristique, due à l'autogamie préférentielle des pêcheurs, et le fait qu'aucune résistance n'ait été détectée au locus *Ma* chez ces derniers, sont des facteurs limitant l'intérêt de cette espèce pour notre étude.

A court terme, une approche de capture et séquençage de grands fragments pourrait être une stratégie ciblée plus robuste et moins chronophage. Par ailleurs, l'utilisation d'amandiers pourrait également être un atout compte tenu de leur niveau d'hétérozygotie plus élevé que celui des pêcheurs. Ces résultats contrastés doivent aussi nous conduire à réfléchir sur le rôle du polymorphisme sur le mécanisme de résistance et notamment le spectre restreint du gène *RMja*. Le dosage des différents allèles et l'influence d'éventuelles isoformes sont des facteurs qui ont également été décrits comme prépondérants chez certains modèles comme *N* (Dinesh-Kumar & Baker, 2000) mais qui peuvent s'avérer délicats à révéler comme chez RPP5 (Jordan *et al.*, 2002). Enfin des essais de transfert de la résistance *Ma* par transformation du pêcher pourraient apporter des informations sur l'effet du fond génétique du prunier.

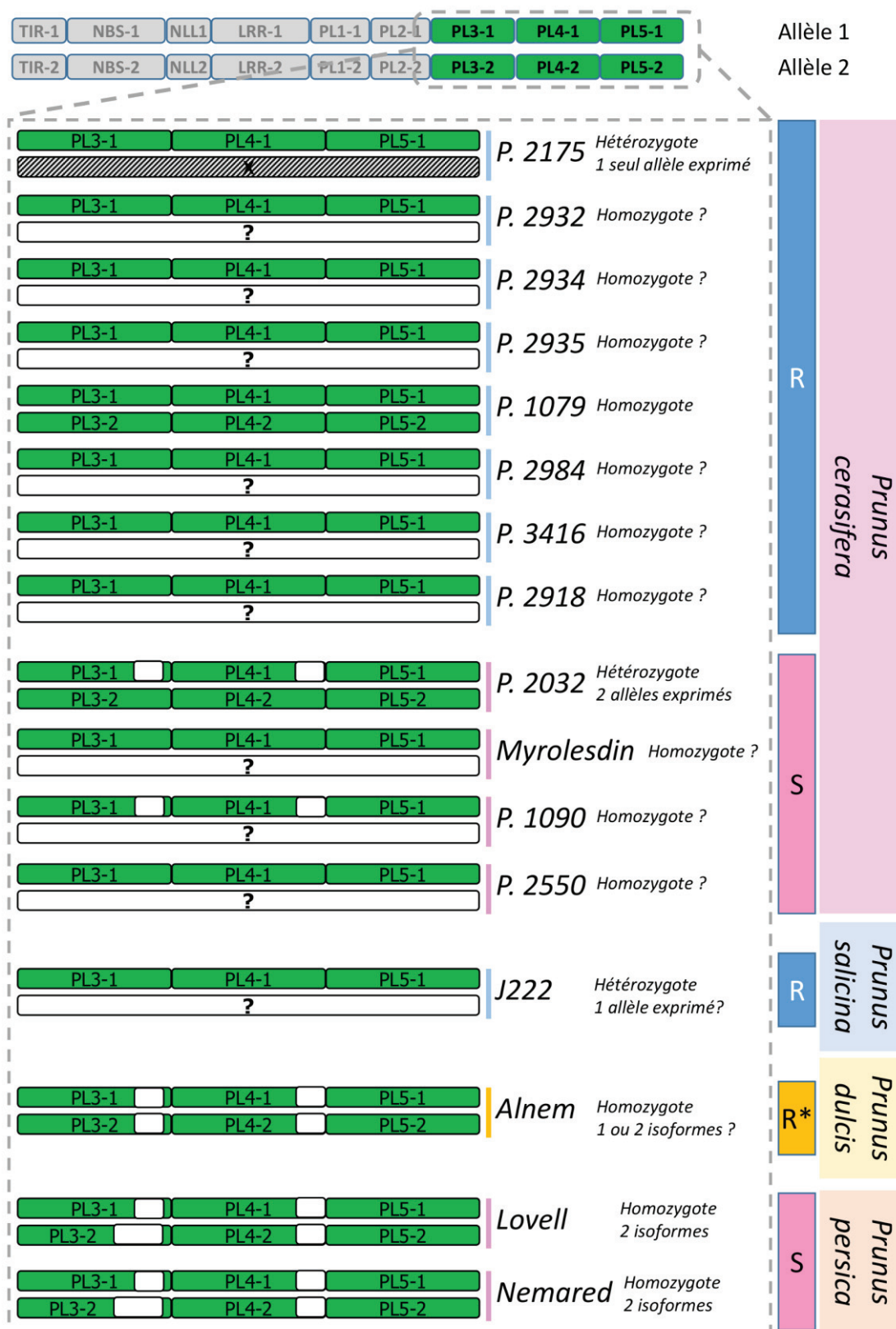


Figure 25 - Polymorphisme observé au niveau des domaines PL3, PL4 et PL5 au sein de quelques variétés de la collection allélique du prunier myrobolan et de variétés d'autres espèces de *Prunus* diploïdes.

Les positions des indels sont indiquées sur les séquences par les rectangles blancs. Sur le panneau de droite, les rectangles indiquent le spectre de résistance aux *Meloidogyne* : la couleur bleue correspond à un spectre complet (R), la couleur jaune correspond au spectre restreint (R*) du gène *RMja* et la couleur rose (S) indique la sensibilité à toutes les espèces.

CHAPITRE 7

CONCLUSION GENERALE ET PERSPECTIVES

Chapitre 7 : CONCLUSION GENERALE ET PERSPECTIVES

Les propriétés remarquables du gène *Ma* chez le prunier myrobolan sont tout d'abord d'ordre biologique : elles sont liées à son spectre complet à ce jour et au haut niveau de résistance qu'il confère. Notamment aucun contournement n'a encore été observé tant en conditions naturelles que contrôlées et *Ma* n'est pas affecté par un facteur environnemental majeur tel que la température. Cependant la robustesse de cette résistance n'a pas été testée en présence d'autres stress biotiques tels que les virus ou phytoplasmes. La mise en évidence de ce gène chez les *Prunus* est particulièrement intéressante car la résistance aux nématodes à galles avait été très peu étudiée auparavant dans les espèces pérennes. Chez ces dernières, *Ma* reste le seul gène de résistance aux *Meloidogyne* cloné à ce jour et constitue donc un modèle d'étude.

Le gène *Ma* est également remarquable par sa structure. En aval de sa conformation canonique TIR-NBS-LRR, il possède une région C-terminale constituée de 5 domaines PL répétés qui en fait à notre connaissance le plus grand gène cloné à ce jour parmi les NBS-LRRs. Cette particularité a offert l'opportunité de développer une étude originale, par une approche bio-informatique, sur la distribution et la fréquence d'un tel type de structure C-terminale chez les plantes.

Nos travaux de thèse ont également bénéficié d'un atout supplémentaire grâce à la mise en évidence du gène *RMja* de résistance aux nématodes à galles chez l'amandier, gène qui confère un spectre différent de *Ma* mais qui lui est orthologue. En effet la disponibilité de *Ma*, de *RMja* et de l'allèle sensible porté par le pêcher Lovell (variété utilisée pour le génome du pêcher), a ouvert la voie à une approche comparative basée sur des séquences à spectre variable.

Dans ce chapitre, les principaux résultats de nos travaux seront synthétisés et discutés afin d'en tirer les conclusions majeures puis nous présenterons les perspectives qui peuvent être mises en avant aux niveaux fondamental et appliqué (agronomique).

7.1 Conclusion générale

7.1.1 Originalité du domaine PL au sein des NBS-LRRs

L'originalité structurale du gène *Ma* a justifié nos travaux de caractérisation fine du domaine PL localisé en position C-terminale chez les TNLs, travaux qui ont montré qu'il s'agit d'un domaine à part entière. A partir des signatures spécifiques de ce domaine, notre étude a montré que les répétitions de ce domaine étaient rares chez les plantes. La présence de PLs multiples chez certaines Rosaceae est

vraisemblablement due à des duplications successives d'exons initiées à partir de PL1. Cette particularité évolutive est probablement apparue tôt dans l'ordre des Rosales. La structure constituée de 5 PLs répétés apparaît comme unique à *Ma* et ses gènes orthologues chez les *Prunus* et chez les Rosaceae apparentées et a été héritée de leur ancêtre commun dans la sous-famille des Spiraeoideae.

La distribution du domaine PL que nous avons mise en lumière chez les Dicotylédones suggère que ce composant est utile au fonctionnement d'une grande partie des TNLs. Le fait qu'il soit retrouvé dans au moins un TNL sur deux chez la plupart des Dicotylédones testées laisse à penser qu'il leur apporte un avantage adaptatif dans le mécanisme de réponse déclenché par les TNLs.

Nos travaux ont également établi l'absence du domaine PL chez les conifères. Nous avons révélé que ce groupe taxonomique porte un effectif élevé de TNLs même si ce domaine n'est pas requis dans leur mode de fonctionnement.

7.1.2 Implication du (des) domaine(s) PL(s) dans la résistance

Apports des approches de mutagenèse chez les gènes R de type TNL portant un seul domaine PL

Notre étude sur le domaine PL chez le pêcher a mis en lumière trois motifs principaux. Le polymorphisme observé entre les motifs de taille réduite et les structures conservées en brins β semblent favoriser une hypothèse selon laquelle ce sont vraisemblablement les structures secondaires du domaine plutôt que la nature proprement dite de ses résidus qui sont importants pour donner un atout fonctionnel au domaine PL. Mais la conservation chez les différentes plantes testées de certains résidus notamment dans le premier motif (résidus proline, tryptophane-phénylalanine ou tyrosine-phénylalanine) délimitant la frontière avec le domaine LRR est très certainement requise pour le fonctionnement des TNLs disposant d'un domaine PL. Par ailleurs, un crible génétique mené chez *Arabidopsis* a mis en évidence des mutants intéressants sur les TNLs en paire, RPS4/RRS1. En effet certains mutants sushi (suppressor of slh1 immunity) présentaient des mutations dans la partie C-terminale de RPS4, Sushi41 (S914F), sushi58 (G952E) et sushi32 (G997E) (Sohn *et al.*, 2014). Cette dernière mutation est particulièrement intéressante car elle se situe exactement sur un résidu glycine très conservé du deuxième motif que nous avons défini. L'utilisation de ces mutants en expression transitoire avec RRS1 et les effecteurs PopP2 ou AvrRps4 ne conduit pas à une réaction de HR ou à la reconnaissance des effecteurs. Ces données suggèrent donc fortement que ce domaine joue un rôle dans la réponse immune, potentiellement par des interactions avec RRS1 ou d'autres partenaires (Sohn *et al.*, 2014). Plus récemment, le domaine PL de RPS4, nommé CTD pour 'C-terminal domain', a fait l'objet d'expériences d'échanges entre les paires de TNLs RPS4/RRS1 et RPS4B/RRS1B (Ma *et al.*, 2018). Cette étude confirme l'implication directe ou indirecte des résidus identifiés chez les mutants sushi dans le déclenchement de la réponse induite par l'un ou l'autre des effecteurs. Il apparaît que la complémentarité entre le domaine PL de RPS4 et la

région correspondante chez RRS1 est importante mais aussi que l'interface créée s'inscrit dans une configuration plus large à respecter pour préserver la réponse de la plante aux effecteurs. Le domaine PL serait donc *a minima* impliqué dans les mécanismes de dimérisation des TNLs dont on sait qu'ils sont essentiels au déclenchement de la réponse immune de la plante (Mestre & Baulcombe, 2006; Schreiber *et al.*, 2016).

En examinant les huit couples différents de TNLs fonctionnant en paire et arrangés en face à face dans le génome d'*Arabidopsis* tel RRS1/RPS4 (Narusaka *et al.*, 2009), nous pouvons dégager un profil particulier. En effet dans chaque paire, le TNL disposant d'une structure et des motifs canoniques avec un domaine TIR capable de déclencher à lui seul une HR possède toujours un domaine PL avec des motifs très conservés. A l'inverse, l'autre TNL de chaque paire, portant des structures atypiques et/ou un ID, ne possède pas de domaine PL ou possède un PL avec des résidus, habituellement conservés dans les signatures, absents. Dans ce contexte d'hétérodimérisation, il semble qu'un domaine PL canonique (avec les signatures spécifiques) soit toujours requis chez le premier TNL de la paire.

En complément de la suppression totale ou de l'échange du domaine entier, il serait utile de développer une stratégie plus ciblée de mutation dirigée vers les résidus précédemment évoqués dans un ou plusieurs TNLs modèles. L'importance du PL serait confirmée si ces mutants par exemple n'induisent plus ou que partiellement le déclenchement de HR ou au contraire induisent des phénomènes d'auto-activation.

Apports de l'étude de polymorphisme entre espèces et clones de *Prunus* de la région post-LRR à PLs multiples (locus *Ma*)

La localisation puis l'identification du gène *RMja* nous ont permis de progresser dans l'identification des déterminants impliqués dans la résistance aux nématodes à galles. En effet, l'obtention de la séquence de l'amandier au locus *Ma*, ajoutée à celles déjà disponibles du pêcher et du prunier myrobolan, permet de révéler le polymorphisme entre trois accessions répondant différemment vis-à-vis de ces ravageurs. D'une façon surprenante, la structure globale de ce gène reste très conservée chez les trois espèces, produisant un transcrite de taille supérieure à 6 kb soit environ 2 fois la taille d'un transcrite de TNL standard. Ce n'est donc pas un polymorphisme de taille marqué, avec par exemple des domaines PL tronqués ou absents, qui explique des spectres de résistance aussi différents entre formes alléliques des gènes chez ces trois espèces. Le faible polymorphisme observé est en accord avec des travaux qui montrent qu'une seule mutation dans le domaine LRR peut entraîner une perte ou une amélioration de la reconnaissance de l'effecteur (Bakker *et al.*, 2014; Segretin *et al.*, 2014). Même si notre analyse a permis de localiser quelques mutations dans les domaines TIR, NBS et NLL, l'essentiel du polymorphisme entre *Ma* et *RMja* se trouve dans la région entre la partie C-terminale du domaine LRR et le domaine PL4. En croisant les données relatives au polymorphisme entre *Ma* et ses orthologues et aux pressions de sélection (ratios Ka/Ks) opérant sur les différents domaines, la zone PL3-PL4 apparaît comme celle qui a le plus de chances d'être impliquée dans la résistance aux nématodes à galles.

De plus, l'épissage alternatif mis en lumière uniquement chez les deux clones de pêcheurs et les structures secondaires identifiés dans cette zone tendent à confirmer l'intérêt d'étudier plus en détail cette région. Il faut noter que les variétés résistantes *Ma* de prunier myrobolan et *Rjap* du prunier japonais possèdent un allèle particulier qui ne porte aucune des deux délétions fréquemment observées chez PL3 et PL4 et concernant deux indels de trois acides aminés (résidus valine-asparagine-proline et aspartate-glutamate-lysine). Des résultats de génotypage et phénotypage d'un grand nombre de variétés permettraient de répondre à l'hypothèse d'une implication de ce(s) motif(s) dans la résistance complète des pruniers. A titre d'exemple, Mi-1.2 conférant la résistance à certaines espèces de *Meloidogyne* chez la tomate, possède un homologue identique à 91%, Mi-1.1, mais qui ne confère pas de résistance (Milligan *et al.*, 1998). Parmi les mutations entre ces deux homologues, les auteurs suggèrent que les trois acides aminés consécutifs (résidus asparagine-lysine-glycine) localisés dans le LRR de Mi-1.2 pourraient être impliqués dans la reconnaissance du nématode (Hwang & Williamson, 2003). Dans notre étude il est vraisemblable que la résistance à spectre incomplet, qui différencie l'amandier Alnem 1 du pêcheur sensible Lovell, ne dépend que de quelques mutations (98.78% de similarité entre RMja et Prupe.7G65400). Il n'est toutefois pas exclu que cette résistance soit également influencée par d'autres facteurs tels la présence de partenaires moléculaires ou le niveau d'expression des isoformes présentes. Nos résultats vont dans le sens d'une hypothèse selon laquelle i) les déterminants de la résistance seraient localisés dans la région post-LRR et ii) le spectre de résistance dépendrait d'un polymorphisme relativement modéré dans cette dernière. Une zone plus ciblée, située au niveau des domaines PL3 et PL4, pourrait être liée au spectre complet de la résistance portée par les pruniers.

Les TNLs contenant plusieurs domaines PLs

Si la distribution du domaine PL plaide en faveur d'un rôle générique chez les plantes à fleur, apportant probablement un atout fonctionnel, le caractère exceptionnel des gènes dotés de répétitions de PLs suggère qu'ils correspondent davantage au profil des domaines « leurres » ('Decoy/Integrated domains') (van der Hoorn & Kamoun, 2008). Alors que les PLs répétés de *Ma* et *RMja* ne portent pas tous les motifs conservés de ce domaine, il est possible que ces répétitions permettent la configuration d'une structure impliquée dans les propriétés remarquables de résistance aux nématodes à galles. Le polymorphisme observé sur certains motifs des PLs de *Ma* ou *RMja* par rapport aux signatures décrites dans notre étude semble aller dans ce sens. Par ailleurs, la conservation de cinq PLs avec des dimensions très similaires chez *Ma* et ses orthologues à travers diverses espèces de *Prunus* et au-delà sont des éléments en accord avec une structure où chaque élément individuel aurait son importance. Chez *Arabidopsis* certains TNLs impliqués dans des mécanismes de leurre possèdent des structures atypiques et/ou des domaines intégrés tels RRS1 (Le Roux *et al.*, 2015), RRS1B, CHS3 ou WRKY19 mais ne possèdent pas de domaine PL. Chez cette même espèce un seul TNL, VICTL (AT5G46510), possède deux domaines PL. Ce gène est retrouvé dans un cluster de quatre TNLs dont le plus proche paralogue

ou gène tandem, *VICTR* ou *ACQOS* (*AT5G46520*), est impliqué à la fois dans l'arrêt de la croissance racinaire et la fermeture des stomates induite par l'acide abscissique, et dans un mécanisme de compromis entre résistance à un stress biotique (bactérien) et de sensibilité à un stress abiotique (stress osmotique) (Kim *et al.*, 2012; Ariga *et al.*, 2017). Même si le rôle de *VICTL* n'est pas très clair dans ce système complexe, il nous renvoie à la possible implication de paralogues de *Ma/RMja* dans le processus de résistance mais également dans la probabilité que ce locus soit impliqué dans d'autres mécanismes physiologiques (adaptation aux stress abiotiques, balance hormonale).

7.1.3 Origine et évolution des NBS-LRRs et du domaine PL

L'origine et l'évolution des NBS-LRRs font depuis quelques années l'objet de nombreuses études (Shao *et al.*, 2016, 2018; Urbach & Ausubel, 2017). Les études phylogénétiques ont permis d'identifier leur origine (Gao *et al.*, 2018), leur processus de formation et d'évolution (Yue *et al.*, 2012) mais aussi de situer cette famille multigénique par rapport aux récepteurs immuns (NOD-like) chez les animaux (Urbach & Ausubel, 2017). Les TNLs apparaissent très tôt dans la lignée végétale (Gao *et al.*, 2018). En recherchant la distribution du domaine PL, nous avons pu mettre en évidence l'absence de ce domaine chez la plupart des plantes ayant divergé avant les Angiospermes. L'étude d'un jeu de données conséquent chez les conifères, dont les NBS-LRRs avaient été jusqu'ici peu étudiés, a été cruciale pour mettre en lumière cette caractéristique. Nos résultats ont apporté des éléments pour identifier le point d'apparition de ce domaine. Il semble que toutes les Dicotylédones ayant des TNLs, même en effectif réduit, disposent de domaines PL. A l'inverse, parmi les plantes ayant divergé avant les Angiospermes, seul le *G. biloba*, qui possède un répertoire relativement important de TNLs (environ 130) (Gao *et al.*, 2018), présente quelques domaines PL (**Annexe 7**). Cette donnée illustre une acquisition partielle du domaine chez certaines plantes anciennes mais surtout une expansion adaptative précoce chez les Angiospermes Dicotylédones. Du fait de la disparition des TNLs chez les Monocotylédones (Tarr & Alexander, 2009), le domaine PL est également absent de ce groupe taxonomique.

La question de l'apparition du domaine PL, formé soit *de novo* (issu de séquences non codantes puis sélectionné) ou soit par intégration d'un domaine préexistant, reste ouverte. Chez certaines protéines LRR-kinase, nous avons retrouvé, en position N-terminale du domaine LRR, une similarité avec le motif 1 défini chez le domaine PL (**Annexe 8**). Ces protéines présentes dans des plantes archaïques (*Aquilegia caerulea*: Aqcoe6G120200.1.p, *Physcomitrella patens*: Pp3c6_19730V3.1.p, *Sphagnum fallax*: Sphfalx0065s0013.1.p) dont les TNLs ne possèdent pas de domaine PL, pourraient avoir initié l'acquisition du premier motif par recombinaison entre domaines LRR de ces kinases et certains TNLs. Cet événement aurait précédé une élongation du domaine et lui aurait ainsi conféré un avantage adaptatif. Cependant cela reste spéculatif en raison de la taille et de la spécificité réduite du motif en

question. L'analyse future de données transcriptomiques et génomiques chez d'autres groupes taxonomiques tels les fougères ou d'autres Gymnospermes (Cycadaceae, Gnétales) pourrait nous permettre d'affiner le portrait évolutif de ce domaine. Cependant, si aucun domaine PL n'est retrouvé dans ces groupes, la question restera ouverte du fait que l'espèce *G. biloba* est une impasse évolutive.

Les conclusions d'ordres structurel, évolutif et fonctionnel liées au domaine PL sont résumées dans la **Figure 26**.

A l'image de ce qui est observé chez les algues vertes, les effectifs des différentes sous-familles de NBS-LRRs sont variables selon les groupes taxonomiques, incluant même la disparition de certaines sous-familles. L'analyse des répertoires de ces sous-familles nous a permis de remarquer des effectifs importants de TNLs et RNLs chez les deux groupes taxonomiques très éloignés des conifères et des Rosaceae. Ces données nous ont conduit à proposer l'hypothèse d'une corrélation quantitative entre les effectifs de RNLs et TNLs chez de nombreuses plantes. Au regard du dynamisme de cette famille de gènes, les RNLs et les TNLs ont donc subi des expansions et contractions vraisemblablement coordonnées ayant abouti à un ratio approximatif de 1:10. Cette relation génomique est à mettre en perspective avec les études fonctionnelles récentes qui ont montré le lien entre TNLs en amont et RNLs en aval pour déclencher une réponse immune adaptée de la plante chez des modèles bien décrits tels que N ou RPS4/RRS1 (Qi *et al.*, 2018; Castel *et al.*, 2019). Les auteurs soulignent l'implication des RNLs ADR1, déjà identifié comme intervenant dans la tolérance à la sécheresse (Chini *et al.*, 2004), et NRG1, comme plateforme agissant en aval des NBS-LRRs pour induire une réaction immune. Notre travail chez les conifères a mis en lumière certaines séquences RNL précocement réactives au stress occasionné par un déficit en eau et potentiellement impliquées dans l'adaptation à ce stress abiotique. A l'image d'ADR1, ces séquences pourraient également être associées à la réponse générale des NBS-LRRs aux stress biotiques. Ces résultats complètent ainsi un peu plus le portrait des partenaires essentiels au fonctionnement de la réponse déclenchée par les TNLs et laissent entrevoir des possibilités nouvelles de transfert de la résistance à des espèces jusque-là trop éloignées taxonomiquement. Plus spécifiquement dans le cas du gène *Ma*, un ou plusieurs RNLs, dont une grande partie se trouve également sur le chromosome 7, pourraient être des partenaires privilégiés dans la réponse déclenchée par ce gène.

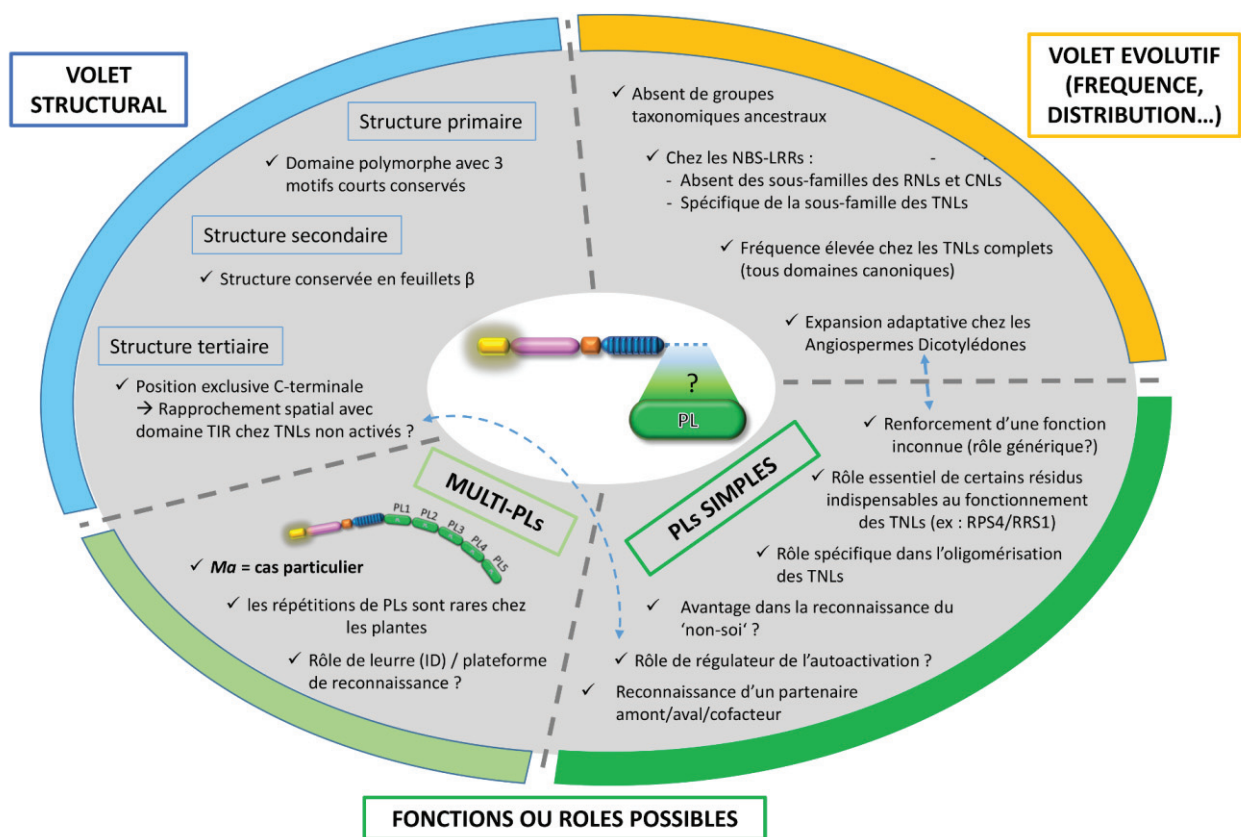


Figure 26 - Diagramme regroupant les données et hypothèses concernant le domaine PL chez les TNLs.

7.2 Perspectives

Nos travaux de caractérisation du domaine PL ont permis d'apporter de nouvelles connaissances (**Figure 26**) même si on ignore toujours le rôle et/ou la fonction précis de cette structure. Les établir est une perspective prioritaire de nos travaux futurs et ne sera possible qu'en menant à bien plusieurs approches complémentaires. Une première approche est une étude comparative complétant celle que nous avons initiée dans notre travail (Chapitre 6) mais basée sur un nombre élevé de séquences alléliques provenant d'accessions de phénotypes de résistance connus. Une autre approche est mécanistique et basée sur des échanges, délétions ou modifications de séquences chez des formes alléliques résistantes et sensibles au locus *Ma*.

7.2.1 *Exploitation des séquences alléliques et des données d'expression*

L'acquisition d'une cinétique précise d'expression du gène *Ma* par PCR quantitative en regard de l'infection par le nématode pourrait permettre de préciser l'étape où le gène agit préférentiellement. La connaissance de cette étape couplée aux données transcriptomiques disponibles chez les différents stades parasitaires de *Meloidogyne* orienterait dans un second temps la recherche d'un type d'effecteur agissant de façon plus ou moins précoce dans l'induction du site nourricier.

La détermination des séquences alléliques/orthologues de *Ma*, résistantes et sensibles, associées aux données d'expression permettrait de préciser les déterminants pouvant être impliqués dans la résistance aux nématodes à galles. Avec le recul des données obtenues, une approche de type RNAseq sur les clones P2175 (R) et P2032 (S) de *P. cerasifera* sous deux conditions (inoculés avec *M. incognita* et non inoculés) nous aurait donné les profils d'expression des allèles résistants et sensibles mais aussi des différentes isoformes et des autres partenaires potentiels chez les *Prunus* (kinases, RNLs, EDS1, ...). Les modifications engendrées par l'infestation avec *M. incognita* pourraient ainsi être détectées pour la première fois chez une espèce pérenne ligneuse et mises en perspective avec les profils d'expression disponibles chez d'autres espèces de plantes d'intérêt agronomique (Solanaceae) infectées par ce nématode (Bagnaresi *et al.*, 2013; Shukla *et al.*, 2018).

7.2.2 *Modification de la région PL chez Ma (approche mécanistique)*

Afin de mieux comprendre les raisons fonctionnelles qui sous-tendent la structure atypique du gène *Ma*, une stratégie de modification des domaines PL apparaît pertinente. Même si la transformation chez les *Prunus*, l'insertion d'un gène de cette dimension et le phénotypage par inoculation *in vitro* de nématodes

restent des opérations délicates, cette approche est sans doute la plus fiable pour tester différentes constructions dans un fond génétique idoine. Une approche, initiée dans notre laboratoire et basée sur du clonage modulaire facilitant l'assemblage de fragments d'ADN dans l'ordre désiré (Engler *et al.*, 2014), doit permettre de produire des constructions alternatives du gène *Ma* pour leur test ultérieur après transformation et production de 'hairy roots' chez un *Prunus* sensible. Cela permet notamment de réaliser des inversions entre des domaines ou régions régulatrices chez des allèles résistant et sensible *Ma*. La délétion d'un (ou plusieurs) domaine(s) PL(s) peut également permettre de savoir si ces domaines sont requis pour la résistance même si une réponse négative serait surprenante en raison de la conservation de cette structure. Une approche plus ciblée centrée sur le domaine PL3 et sa partie C-terminale contribuerait à identifier si cette zone joue un rôle dans la détermination du spectre de résistance, qu'il soit complet ou plus étroit.

7.2.3 Recherche des partenaires moléculaires du gène *Ma*

Une stratégie de 'Knock-out' par 'genome editing' (Ilardi & Tavazza, 2015) ou de 'silencing' par VIGS (Cui & Wang, 2017) chez le clone P.2175 de *P. cerasifera*, porteur du gène *Ma*, en ciblant des partenaires candidats tels la lectine kinase située en amont du cluster de *Ma* ou certains RNLs, nous permettrait de tester l'obtention d'un éventuel phénotype sensible chez cette accession. Dans un deuxième temps, l'identification de tels partenaires ouvrirait la voie à une étude d'interaction entre ceux-ci et la protéine *Ma* ou plus spécifiquement une région de cette protéine. Une approche de double hybride / BiFC devrait conduire à la détection de telles interactions.

L'étude du gène *Ma* dans un système hétérologue faciliterait également la compréhension de ce système. L'absence de résistance aux nématodes à galles chez la tomate transformée avec *Ma* suggère de travailler sur des plantes sensibles aux nématodes à galles qui soient taxonomiquement plus proches des *Prunus*. De telles plantes doivent imposer des contraintes techniques moindres que les *Prunus* et posséder un répertoire de TNLs plus développé que celui des Solanaceae. *Arabidopsis*, ou l'espèce *Medicago truncatula*, encore plus proche botaniquement des *Prunus*, semblent répondre à ces critères.

7.2.4 Utilisation des TNLs modèles afin de comprendre le rôle du domaine PL

D'un point de vue plus fondamental, le décryptage du (ou des) rôle(s) du domaine PL chez les TNLs serait une avancée décisive dans la compréhension du fonctionnement global de ces récepteurs cytoplasmiques. Les travaux réalisés sur la paire RPS4/RRS1 (Ma *et al.*, 2018) soutiennent l'hypothèse d'une interaction du domaine PL de RPS4 avec son vis-à-vis chez RRS1 contribuant à la dimérisation

des deux protéines. Cependant, il est possible que ce ne soit pas le rôle principal du domaine PL à l'image du domaine TIR, également engagé dans la dimérisation et remplissant d'autres rôles indépendants (Bernoux *et al.*, 2011). Mais contrairement aux domaines TIR, NBS et LRR qui ont pu bénéficier des travaux conduits chez les animaux (respectivement les récepteurs Toll de la drosophile, les domaines NACHT chez les métazoaires ou les récepteurs membranaires chez de nombreux modèles), la présence exclusive du domaine PL chez les TNLs impose une approche mécanistique plus spécifique. Générer des mutants ciblés au niveau des résidus présents dans les motifs 1, 2 et 3 (**paragraphe 7.1.2**) chez des TNLs modèles apparait comme une stratégie permettant d'amorcer cette thématique. La paire RPS4/RRS1 est un système hétérogène de TNLs bien étudié impliquant l'intervention d'au moins deux effecteurs différents. L'agro-infiltration en feuille de tabac de RPS4 doté d'un PL muté, en même temps que le partenaire RRS1 et les effecteurs PopP2 et AvrRps4, est en mesure de nous aider à identifier les résidus requis pour le déclenchement d'une HR ou éventuellement ceux prévenant par une réaction d'auto-immunité l'implication de ces motifs. La mutation de ces mêmes résidus chez N, autre TNL modèle chez le tabac, est susceptible de contribuer à une meilleure interprétation des résultats sur un système sensiblement différent. En effet N s'oligomérisent mais existe sous deux isoformes, toutes deux essentielles à la réalisation du processus de résistance, dont une seule dispose du domaine PL (Dinesh-Kumar & Baker, 2000).

Cependant, si l'on veut élucider plus précisément le rôle du domaine PL, la stratégie ambitieuse à privilégier serait, dans un premier temps, la recherche d'interactants intra-moléculaires (domaines TIR, NBS ou LRR) par double hybride par exemple et, dans un deuxième temps, la recherche d'interactants extra-moléculaires par co-IP ou pull down en couplage avec une spectrométrie de masse.

7.2.5 Applications

Dans une optique plus finalisée, transférer le gène *Ma* chez le pêcher, espèce de grande importance économique, et connaître son activité dans ce fond génétique serait une avancée significative. En effet cela permettrait notamment de savoir si un autre facteur appartenant au prunier est requis pour conférer cette résistance au pêcher. A plus long terme, l'identification de partenaires du gène *Ma* pourrait contribuer à la réussite du transfert de la résistance portée par ce gène dans des espèces éloignées taxonomiquement à l'image du transfert de la résistance conférée par RPS4/RRS1 réalisé avec succès chez des Solanaceae (Narusaka *et al.*, 2013). En effet la co-transformation des divers intervenants dans les processus de résistance devrait permettre à terme de franchir certaines barrières taxonomiques. L'identification des déterminants de la résistance chez *Ma* ouvrirait la voie à la conception et la création d'une nouvelle génération de gènes *R*, un objectif dont les premières applications pourraient être obtenues très prochainement (Kourelis & van der Hoorn, 2018).

Sur le plan strictement agronomique, il convient de signaler que nos résultats, suggérant de façon quasi-certaine pour *Ma* et *RMja* le statut de gènes orthologues, apportent des éléments nouveaux pour la lutte

contre les *Meloidogyne* chez les arbres fruitiers à noyau. En effet les nouveaux porte-greffes de *Prunus* sont souvent obtenus par hybridation interspécifique. La création de porte-greffe à résistance forte et élevée vis-à-vis des nématodes à galles peut ainsi s'appuyer sur le pyramidage des gènes issus des différentes espèces utilisées. Mais dans cet objectif, le gène *RMja*, en tant que membre du cluster *Ma*, ne constitue pas une ressource aussi intéressante que l'on pouvait l'imaginer *a priori* car il a une même origine et met en jeu vraisemblablement un mécanisme de reconnaissance identique ou proche. Cette stratégie de pyramidage reste toutefois justifiée pour les couples de gènes génétiquement indépendants, *Ma+RMia* et *RMja+RMia*. Dans ce contexte, les hybrides prunier-pêcher *Ma+RMia* confèrent une double protection vis-à-vis de toutes les espèces à l'exception de *M. javanica*, *M. floridensis* et *M. mayaguensis* (contrôlés uniquement par *Ma*). Les hybrides amandier-pêcher *RMja+RMia* portent quant à eux une double protection vis-à-vis de *M. arenaria* et *M. ethiopica*, une protection simple vis-à-vis de *M. incognita*, *M. javanica* et *M. mayaguensis* mais restent sensibles à *M. floridensis*. A terme, la structure unique de *Ma* devrait vraisemblablement permettre de mettre en évidence un mécanisme inédit de défense des plantes vis-à-vis de ravageurs évolués tels que les nématodes à galles. Les caractéristiques biologiques remarquables que cette structure confère à *Ma* sont un atout essentiel dans la lutte contre ces nématodes. Il est probable que la connaissance de ce mécanisme ouvrira des perspectives de transfert de la résistance portée par ce gène à d'autres espèces végétales et contribuera à une gestion durable de ces bioagresseurs.

REFERENCES BIBLIOGRAPHIQUES

- Aarts N, Metz M, Holub E, Staskawicz BJ, Daniels MJ, Parker JE. 1998. Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 10306–11.
- Abad P, Gouzy J, Aury J-MM, Castagnone-Sereno P, Danchin EGJJ, Deleury E, Perfus-Barbeoch L, Anthouard V, Artiguenave F, Blok VC, *et al.* 2008. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology* **26**: 909–915.
- Abad P, Williamson VM. 2010. Plant nematode interaction: a sophisticated dialogue. In: Delseny M, Kader J-C, eds. *Advances in Botanical Research*, Volume 53. 147–192.
- Ade J, DeYoung BJ, Golstein C, Innes RW. 2007. Indirect activation of a plant nucleotide binding site-leucine-rich repeat protein by a bacterial protease. *Proceedings of the National Academy of Sciences* **104**: 2531–2536.
- Ali GS, Prasad KVSK, Day I, Reddy ASN. 2007. Ligand-dependent reduction in the membrane mobility of FLAGELLIN SENSITIVE2, an arabidopsis receptor-like kinase. *Plant & cell physiology* **48**: 1601–11.
- Ameline-Torregrosa C, Wang B-B, O’Blens MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB. 2007. Identification and Characterization of Nucleotide-Binding Site-Leucine-Rich Repeat Genes in the Model Plant *Medicago truncatula*. *Plant Physiology* **146**: 5–21.
- Andolfo G, Jupe F, Witek K, Etherington GJ, Ercolano MR, Jones JDGG. 2014. Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. *BMC plant biology* **14**: 120.
- APG IV, Moore M, Olmstead R, Perret M, Skog L, Smith J, Tank D, Vorontsova M, Weber A. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**: 1–20.
- Ariga H, Katori T, Tsuchimatsu T, Hirase T, Tajima Y, Parker JE, Alcázar R, Koornneef M, Hoekenga O, Lipka AE, *et al.* 2017. NLR locus-mediated trade-off between abiotic and biotic stress adaptation in Arabidopsis. *Nature plants* **3**: 17072.
- Baek S, Choi K, Kim GB, Yu HJ, Cho A, Jang H, Kim C, Kim HJ, Chang KS, Kim JH, *et al.* 2018. Draft genome sequence of wild *Prunus yedoensis* reveals massive inter-specific hybridization between sympatric flowering cherries. *Genome Biology* **19**: 1–17.
- Bagnaresi P, Sala T, Irdani T, Scotto C, Lamontanara A, Beretta M, Rotino GL, Sestili S, Cattivelli L, Sabatini E. 2013. *Solanum torvum* responses to the root-knot nematode *Meloidogyne incognita*. *BMC genomics* **14**: 540.
- Bakker EH, Roosien J, Slootweg EJ, Butterbach PBE, Spiridon LN, Smant G, Petrescu AJ, Bakker J, Goverse A. 2014. A single amino acid substitution in the LRR domain is sufficient to compromise effector recognition by the nematode immune receptor GPA2. In: XVI International Congress on Molecular Plant-Microbe Interactions. Rhodos, Greece.
- Barker KR, Carter CC, Sasser JN. 1985. *Advanced Treatise on Meloidogyne, Volume I: Biology and Control*. Department of Plant Pathology, North Carolina State University.
- Baumgarten A, Cannon S, Spangler R, May G. 2003. Genome-level evolution of resistance genes in *Arabidopsis thaliana*. *Genetics* **165**: 309–19.
- Bellaïfiore S, Shen Z, Rosso MN, Abad P, Shih P, Briggs SP. 2008. Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. *PLoS Pathogens* **4**.
- Bendahmane A, Farnham G, Moffett P, Baulcombe DC. 2002. Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the Rx locus of potato. *The Plant journal : for cell and molecular biology* **32**: 195–204.
- Bernoux M, Ve T, Williams S, Warren C, Hatters D, Valkov E, Zhang X, Ellis JG, Kobe B, Dodds PN. 2011. Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell Host and Microbe* **9**: 200–211.
- Bhattacharjee S, Halane MK, Kim SH, Gassmann W. 2011. Pathogen Effectors Target Arabidopsis EDS1 and Alter Its Interactions with Immune Regulators. *Science* **334**: 1405–1408.
- van der Biezen EA, Jones JD. 1998. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Current biology : CB* **8**: R226-7.

- Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL. 2011.** Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proceedings of the National Academy of Sciences* **108**: 16463–16468.
- Borrelli GM, Mazzucotelli E, Marone D, Crosatti C, Michelotti V, Valè G, Mastrangelo AM. 2018.** Regulation and Evolution of NLR Genes : A Close Interconnection for Plant Immunity. : 1–26.
- Brendolise C, Montefiori M, Dinis R, Peeters N, Storey RD, Rikkerink EH. 2017.** A novel hairpin library-based approach to identify NBS-LRR genes required for effector-triggered hypersensitive response in *Nicotiana benthamiana*. *Plant Methods* **13**: 1–10.
- Burch-Smith TM, Schiff M, Caplan JL, Tsao J, Czymmek K, Dinesh-Kumar SP. 2007.** A novel role for the TIR domain in association with pathogen-derived elicitors. *PLoS Biology* **5**: 0501–0514.
- Caldwell KS, Michelmore RW. 2009.** Arabidopsis thaliana genes encoding defense signaling and recognition proteins exhibit contrasting evolutionary dynamics. *Genetics* **181**: 671–684.
- Cao K, Wang L, Zhao P, Zhu G, Fang W, Chen C, Wang X. 2014.** Identification of a candidate gene for resistance to root-knot nematode in a wild peach and screening of its polymorphisms. *Plant Breeding* **133**: 530–535.
- Cardoso-Moreira M, Arguello JR, Gottipati S, Harshman LG, Grenier JK, Clark AG. 2016.** Evidence for the fixation of gene duplications by positive selection in *Drosophila*. *Genome Research* **26**: 787–798.
- Castagnone-Sereno P, Bongiovanni M, Dalmasso A. 1994.** Reproduction of Virulent Isolates of *Meloidogyne incognita* on Susceptible and Mi-resistant Tomato. *J Nematol* **26**: 324–328.
- Castagnone-Sereno P, Danchin EGJ. 2014.** Parasitic success without sex - the nematode experience. *Journal of Evolutionary Biology* **27**: 1323–1333.
- Castagnone-Sereno P, Danchin EGJ, Perfus-Barbeoch L, Abad P. 2013.** Diversity and Evolution of Root-Knot Nematodes, Genus *Meloidogyne* : New Insights from the Genomic Era. *Annual Review of Phytopathology* **51**: 203–220.
- Castel B, Ngou P-MM, Cevik V, Redkar A, Kim D-SS, Yang Y, Ding P, Jones JDGG. 2019.** Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *The New phytologist* **222**: 966–980.
- Cesari S. 2017.** Multiple strategies for pathogen perception by plant immune receptors. *New Phytologist*.
- Césari S, Kanzaki H, Fujiwara T, Bernoux M, Chalvon V, Kawano Y, Shimamoto K, Dodds P, Terauchi R, Kroj T. 2014.** The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *The EMBO Journal* **33**: 1941–1959.
- Cheng YT, Germain H, Wiermer M, Bi D, Xu F, Garcia A V., Wirthmueller L, Despres C, Parker JE, Zhang Y, et al. 2009.** Nuclear Pore Complex Component MOS7/Nup88 Is Required for Innate Immunity and Nuclear Accumulation of Defense Regulators in Arabidopsis. *the Plant Cell Online* **21**: 2503–2516.
- Chin SW, Shaw J, Haberle R, Wen J, Potter D. 2014.** Diversification of almonds, peaches, plums and cherries - Molecular systematics and biogeographic history of *Prunus* (Rosaceae). *Molecular Phylogenetics and Evolution* **76**: 34–48.
- Chini A, Grant JJ, Seki M, Shinozaki K, Loake GJ. 2004.** Drought tolerance established by enhanced expression of the CC-NBS-LRR gene, ADR1, requires salicylic acid, EDS1 and ABI1. *The Plant journal : for cell and molecular biology* **38**: 810–22.
- Chisholm ST, Coaker G, Day B, Staskawicz BJ. 2006.** Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response. *Cell* **124**: 803–814.
- Claverie M, Bosselut N, Lecouls AC, Voisin R, Lafargue B, Poizat C, Kleinhentz M, Laigret F, Dirlwanger E, Esmenjaud D. 2004a.** Location of independent root-knot nematode resistance genes in plum and peach. *Theoretical and Applied Genetics* **108**: 765–773.
- Claverie M, Dirlwanger E, Bosselut N, Van Ghelder C, Voisin R, Kleinhentz M, Lafargue B, Abad P, Rosso M-NM-NM-NM-N, Chalhoub B, et al. 2011.** The Ma Gene for Complete-Spectrum Resistance to *Meloidogyne* Species in *Prunus* Is a TNL with a Huge Repeated C-Terminal Post-LRR Region. *Plant Physiology* **156**: 779–92.
- Claverie M, Dirlwanger E, Cosson P, Bosselut N, Lecouls AC, Voisin R, Kleinhentz M, Lafargue B, Caboche M, Chalhoub B, et al. 2004b.** High-resolution mapping and chromosome landing at the root-knot nematode resistance locus Ma from Myrobalan plum using a large-insert BAC DNA library. *Theoretical and Applied Genetics* **109**: 1318–1327.
- Collier SM, Hamel L-P, Moffett P. 2011.** Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Molecular plant-microbe interactions : MPMI* **24**: 918–31.

- Crepet WL, Nixon KC, Gandolfo MA. 2004.** Fossil evidence and phylogeny: the age of major angiosperm clades based on mesofossil and macrofossil evidence from Cretaceous deposits. *American Journal of Botany* **91**: 1666–1682.
- Cui H, Gobbato E, Kracher B, Qiu J, Bautor J, Parker JE. 2017.** A core function of EDS1 with PAD4 is to protect the salicylic acid defense sector in Arabidopsis immunity. *New Phytologist* **213**: 1802–1817.
- Cui H, Wang A. 2017.** An efficient viral vector for functional genomic studies of Prunus fruit trees and its induced resistance to Plum pox virus via silencing of a host factor gene. *Plant Biotechnology Journal* **15**: 344–356.
- Danchin EGJ, Arguel MJ, Campan-Fournier A, Perfus-Barbeoch L, Magliano M, Rosso MN, Da Rocha M, Da Silva C, Nottet N, Labadie K, et al. 2013.** Identification of Novel Target Genes for Safer and More Specific Control of Root-Knot Nematodes from a Pan-Genome Mining. *PLoS Pathogens* **9**.
- Danchin EGJ, Rosso M-N, Vieira P, de Almeida-Engler J, Coutinho PM, Henrissat B, Abad P. 2010.** Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences* **107**: 17651–17656.
- Decraemer W (UGent), Hunt DJ. 2006.** Structure and classification. In: Moens RNP and M, ed. Plant nematology. CABI, 3–32.
- Deisenhofer J, Kobe B. 1995.** A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* **374**: 183–186.
- Demangeat G, Komar V, Van-Ghelder C, Voisin R, Lemaire O, Esmenjaud D, Fuchs M. 2010.** Transmission competency of single-female Xiphinema index lines for Grapevine fanleaf virus. *Phytopathology* **100**: 384–389.
- Deslandes L, Olivier J, Theulieres F, Hirsch J, Feng DX, Bittner-Eddy P, Beynon J, Marco Y. 2002.** Resistance to *Ralstonia solanacearum* in Arabidopsis thaliana is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 2404–9.
- Dinesh-Kumar SP, Baker BJ. 2000.** Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 1908–13.
- Djian-caporalino C, Navarrete M, Fazari A, Baily-bechet M, Marteu N, Dufils A, Tchamitchian M, Lefèvre A, Pares L, Mateille T, et al. 2019.** Conception et évaluation de systèmes de culture maraîchers méditerranéens innovants pour gérer les nématodes à galles. *Biotechnology, Agronomy, Society and Environment* **23**: 7–21.
- Djian-Caporalino C, Palloix A, Fazari A, Marteu N, Barbary A, Abad P, Sage-Palloix AM, Mateille T, Risso S, Lanza R, et al. 2014.** Pyramiding, alternating or mixing: Comparative performances of deployment strategies of nematode resistance genes to promote plant resistance efficiency and durability. *BMC Plant Biology* **14**: 1–13.
- Dodds PN, Lawrence GJ, Catanzariti A-M, Teh T, Wang C-IA, Ayliffe MA, Kobe B, Ellis JG. 2006.** Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proceedings of the National Academy of Sciences* **103**: 8888–8893.
- Dodds PN, Lawrence GJ, Ellis JG. 2001.** Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. *The Plant Cell* **13**: 163–78.
- Dosba F, Bernhard R, Zanetto A. 1994.** Importance des ressources génétiques des Prunus. *Comptes rendus de l'Académie d'Agriculture de France* **80**: 45–57.
- Dropkin VH. 1969.** The necrotic reaction of tomatoes and other hosts resistant to Meloidogyne: reversal by temperature. *Phytopathology* **59**: 1632–1637.
- Duval H, Van Ghelder C, Portier U, Confolent C, Meza P, Esmenjaud D. 2019.** New Data Completing the Spectrum of the Ma, RMia, and RMja Genes for Resistance to Root-Knot Nematodes (Meloidogyne spp.) in Prunus. *Phytopathology* **109**: 615–622.
- Duval H, Hoerter M, Polidori J, Confolent C, Masse M, Moretti A, Van Ghelder C, Esmenjaud D. 2014.** High-resolution mapping of the RMia gene for resistance to root-knot nematodes in peach. *Tree Genetics and Genomes* **10**: 297–306.
- Engler C, Youles M, Gruetzner R, Ehnert T-M, Werner S, Jones JDG, Patron NJ, Marillonnet S. 2014.** A golden gate modular cloning toolbox for plants. *ACS synthetic biology* **3**: 839–43.
- Eremin VG. 1978.** GENETIC POTENTIAL OF SPECIES PRUNUS CERASIFERA EHRH AND ITS USE IN BREEDING. *Acta Horticulturae*: 61–66.
- Esmenjaud D, Minot JC, Voisin R, Pinochet J, Salesses G. 1994.** Interspecific and Intraspecific Resistance Variability in Myrobalan Plum, Peach, and Peach Almond Rootstocks Using 22 Root-Knot Nematode Populations. *Journal of the American Society for Horticultural Science* **119**: 94–100.

- Esmenjaud D, Minot JC, Voisin R, Pinochet J, Simard MH, Salesses G. 1997.** Differential response to root-knot nematodes in prunus species and correlative genetic implications. *Journal of nematology* **29**: 370–80.
- Esmenjaud D, Voisin R, Van Ghelder C, Bosselut N, Lafargue B, Di Vito M, Dirlewanger E, Poëssel JL, Kleinhentz M. 2009.** Genetic dissection of resistance to root-knot nematodes *Meloidogyne* spp. in plum, peach, almond, and apricot from various segregating interspecific *Prunus* progenies. *Tree Genetics and Genomes* **5**: 279–289.
- Flor HH. 1955.** Host-parasite interactions in flax Gene for Gene. *Phytopathology* **45**: 680–685.
- Gao Z, Chung E-H, Eitas TK, Dangl JL. 2011.** Plant intracellular innate immune receptor Resistance to *Pseudomonas syringae* pv. *maculicola* 1 (RPM1) is activated at, and functions on, the plasma membrane. *Proceedings of the National Academy of Sciences* **108**: 7619–7624.
- Gao Y, Wang W, Zhang T, Gong Z, Zhao H, Han G-Z. 2018.** Out of Water: The Origin and Early Diversification of Plant R-Genes. *Plant physiology* **177**: 82–89.
- Gassmann W, Hinsch ME, Staskawicz BJ. 1999.** The Arabidopsis RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *The Plant journal : for cell and molecular biology* **20**: 265–77.
- Van Ghelder C, Esmenjaud D. 2016.** TNL genes in peach: insights into the post-LRR domain. *BMC genomics* **17**: 317.
- Van Ghelder C, Lafargue B, Dirlewanger E, Ouassa A, Voisin R, Polidori J, Kleinhentz M, Esmenjaud D. 2010.** Characterization of the RMja gene for resistance to root-knot nematodes in almond: Spectrum, location, and interest for *Prunus* breeding. *Tree Genetics and Genomes* **6**: 503–511.
- Goodey JB, Franklin MT, Hooper DJ. 1965.** *The nematode parasites of plants catalogued under their hosts* (F Royal, Ed.). Commonwealth Agricultural Bureaux.
- Gou M, Shi Z, Zhu Y, Bao Z, Wang G, Hua J. 2012.** The F-box protein CPR1/CPR30 negatively regulates R protein SNC1 accumulation. *The Plant Journal* **69**: 411–420.
- Gu L, Si W, Zhao L, Yang S, Zhang X. 2015.** Dynamic evolution of NBS–LRR genes in bread wheat and its progenitors. *Molecular Genetics and Genomics* **290**: 727–738.
- Guo Y-L, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D. 2011.** Genome-Wide Comparison of Nucleotide-Binding Site-Leucine-Rich Repeat-Encoding Genes in Arabidopsis. *Plant Physiology* **157**: 757–769.
- Hammond-Kosack KE, Parker JE. 2003.** Deciphering plant-pathogen communication: Fresh perspectives for molecular resistance breeding. *Current Opinion in Biotechnology* **14**: 177–193.
- Hanson PI, Whiteheart SW. 2005.** AAA+ proteins: have engine, will work. *Nature reviews. Molecular cell biology* **6**: 519–29.
- Heath MC. 2000a.** Nonhost resistance and nonspecific plant defenses. *Current Opinion in Plant Biology* **3**: 315–319.
- Heath MC. 2000b.** Hypersensitive response-related death. *Plant molecular biology* **44**: 321–34.
- Heguy A, Baldari CT, Macchia G, Telford JL, Melli M. 1992.** Amino acids conserved in interleukin-1 receptors (IL-1Rs) and the *Drosophila* toll protein are essential for IL-1R signal transduction. *The Journal of biological chemistry* **267**: 2605–9.
- Hofberger JA, Zhou B, Tang H, Jones JDG, Schranz ME. 2014.** A novel approach for multi-domain and multi-gene family identification provides insights into evolutionary dynamics of disease resistance genes in core eudicot plants. *BMC Genomics* **15**.
- Hohmann U, Lau K, Hothorn M. 2017.** The Structural Basis of Ligand Perception and Signal Activation by Receptor Kinases. *Annual Review of Plant Biology* **68**: 109–137.
- Holterman M, Karssen G, van den Elsen S, van Megen H, Bakker J, Helder J. 2009.** Small Subunit rDNA-Based Phylogeny of the Tylenchida Sheds Light on Relationships Among Some High-Impact Plant-Parasitic Nematodes and the Evolution of Plant Feeding. *Phytopathology* **99**: 227–235.
- van der Hoorn RAL, Kamoun S. 2008.** From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors. *the Plant Cell Online* **20**: 2009–2017.
- Huang G, Dong R, Allen R, Davis EL, Baum TJ, Hussey RS. 2006.** A Root-Knot Nematode Secretory Peptide Functions as a Ligand for a Plant Transcription Factor. *Molecular Plant-Microbe Interactions* **19**: 463–470.
- Huang CS, Maggenti AR. 1969.** Wall modifications in developing giant cells of *Vicia faba* and *Cucumis sativus* induced by root knot nematode, *Meloidogyne javanica*. *Phytopathology* **59**: 931–7.
- Hugot J-P, Baujard P, Morand S. 2001.** Biodiversity in helminths and nematodes as a field of study: an overview. *Nematology* **3**: 199–208.

- Hwang CF, Williamson VM. 2003.** Leucine-rich repeat-mediated intramolecular interactions in nematode recognition and cell death signaling by the tomato resistance protein Mi. *Plant Journal* **34**: 585–593.
- Iberkleid I, Vieira P, de Almeida Engler J, Firester K, Spiegel Y, Horowitz SB. 2013.** Fatty Acid-and Retinol-Binding Protein, Mj-FAR-1 Induces Tomato Host Susceptibility to Root-Knot Nematodes. *PLoS ONE* **8**.
- Iardi V, Tavazza M. 2015.** Biotechnological strategies and tools for Plum pox virus resistance: trans-, intra-, cis-genesis, and beyond. *Frontiers in plant science* **6**: 379.
- International Rice Genome Sequencing Project. 2005.** The map-based sequence of the rice genome. *Nature* **436**: 793–800.
- Janati A, Berger JB, Triantaphyllou AC, Dalmaso A. 1982.** Nouvelles données sur l' utilisation des isoestérasés pour l' identification des Meloidogyne. *Revue de Nematologie* **5**: 147–154.
- Jaouannet M, Magliano M, Arguel MJ, Gourgues M, Evangelisti E, Abad P, Rosso MN, Institut UMR, Agrobiotech S, Nice U De, et al. 2013.** The Root-Knot Nematode Calreticulin Mi-CRT Is a Key Effector in Plant Defense Suppression. **26**: 97–105.
- Jaubert S, Laffaire JB, Ledger TN, Escoubas P, Amri EZ, Abad P, Rosso MN. 2004.** Comparative analysis of two 14-3-3 homologues and their expression pattern in the root-knot nematode Meloidogyne incognita. *International Journal for Parasitology* **34**: 873–880.
- Jaubert S, Ledger TN, Laffaire JB, Piotte C, Abad P, Rosso MN. 2002.** Direct identification of stylet secreted proteins from root-knot nematodes by a proteomic approach. *Molecular and Biochemical Parasitology* **121**: 205–211.
- Jia Y, Yuan Y, Zhang Y, Yang S, Zhang X. 2015.** Extreme expansion of NBS-encoding genes in Rosaceae. *BMC genetics* **16**: 48.
- Jones JDG, Dangl JL. 2006.** The plant immune system. *Nature* **444**: 323–9.
- Jones JT, Haegeman A, Danchin EGJ, Gaur HS, Helder J, Jones MGK, Kikuchi T, Manzanilla-López R, Palomares-Rius JE, Wesemael WML, et al. 2013.** Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology* **14**: 946–961.
- Jones DA, Jones JDG. 1997.** The Role of Leucine-Rich Repeat Proteins in Plant Defences. In: 89–167.
- Jordan T, Schornack S, Lahaye T. 2002.** Alternative splicing of transcripts encoding Toll-like plant resistance proteins – what's the functional relevance to innate immunity? *Trends in Plant Science* **7**: 392–398.
- Jung S, Lee T, Cheng C, Buble K, Zheng P, Yu J, Humann J, Ficklin SP, Gasic K, Scott K, et al. 2019.** 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Research* **47**: D1137–D1145.
- Jupe F, Pritchard L, Etherington GJ, Mackenzie K, Cock PJA, Wright F, Sharma SK, Bolser D, Bryan GJ, Jones JDG, et al. 2012.** Identification and localisation of the NB-LRR gene family within the potato genome. *BMC genomics* **13**: 75.
- Kadota Y, Shirasu K, Guerois R. 2010.** NLR sensors meet at the SGT1–HSP90 crossroad. *Trends in Biochemical Sciences* **35**: 199–207.
- Kale SM, Pardeshi VC, Barvkar VT, Gupta VS, Kadoo NY. 2013.** Genome-wide identification and characterization of nucleotide binding site leucine-rich repeat genes in linseed reveal distinct patterns of gene structure. *Genome* **56**: 91–9.
- Kaloshian I, Williamson VM, Miyao G, Lawn DA, Westerdahl BB. 1996.** “Resistance-breaking” nematodes identified in California tomatoes. *California Agriculture* **50**: 18–19.
- Kang YJ, Kim KH, Shim S, Yoon MY, Sun S, Kim MY, Van K, Lee S-H. 2012.** Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC plant biology* **12**: 139.
- Khallouk S, Voisin R, Van Ghelder C, Engler G, Amiri S, Esmenjaud D. 2011.** Histological mechanisms of the resistance conferred by the Ma gene against Meloidogyne incognita in Prunus spp. *Phytopathology* **101**: 945–951.
- Khallouk S, Voisin R, Portier U, Polidori J, Van Ghelder C, Esmenjaud D. 2013.** Multiyear evaluation of the durability of the resistance conferred by Ma and RMia genes to Meloidogyne incognita in Prunus under controlled conditions. *Phytopathology* **103**: 833–840.
- Kim T-H, Kunz H-H, Bhattacharjee S, Hauser F, Park J, Engineer C, Liu A, Ha T, Parker JE, Gassmann W, et al. 2012.** Natural variation in small molecule-induced TIR-NB-LRR signaling induces root growth arrest via EDS1- and PAD4-complexed R protein VICTR in Arabidopsis. *The Plant cell* **24**: 5177–92.
- Knepper C, Savory EA, Day B. 2011.** The role of NDR1 in pathogen perception and plant defense signaling. *Plant Signaling and Behavior* **6**: 1114–1116.

- Kobe B, Deisenhofer J. 1994.** The leucine-rich repeat: a versatile binding motif. *Review Literature And Arts Of The Americas*.
- Kochba J, Spiegel-Roy P. 1975.** Inheritance of resistance to the root-knot nematode (*Meloidogyne javanica* Chitwood) in bitter almond progenies. *Euphytica* **24**: 453–457.
- Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W, Martin F. 2008.** Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant molecular biology* **66**: 619–36.
- Kourelis J, van der Hoorn RAL. 2018.** Defended to the Nines: 25 years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. *The Plant Cell* **30**: tpc.00579.2017.
- Krasileva K V., Dahlbeck D, Staskawicz BJ. 2010.** Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. *The Plant cell* **22**: 2444–58.
- Kroj T, Chancelud 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. *The New phytologist* **210**: 618–26.
- Lee D, Bourdais G, Yu G, Robatzek S, Coaker G. 2015.** Phosphorylation of the Plant Immune Regulator RPM1-INTERACTING PROTEIN4 Enhances Plant Plasma Membrane H⁺-ATPase Activity and Inhibits Flagellin-Triggered Immune Responses in Arabidopsis. *The Plant Cell* **27**: 2042–2056.
- Leipe DD, Koonin E V., Aravind L. 2004.** STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *Journal of molecular biology* **343**: 1–28.
- Leister D. 2004.** Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. *Trends in genetics : TIG* **20**: 116–22.
- Levy AA, Feldman M. 2002.** The impact of polyploidy on grass genome evolution. *Plant physiology* **130**: 1587–93.
- Lewis JD, Lee AH-Y, Hassan JA, Wan J, Hurley B, Jhingree JR, Wang PW, Lo T, Youn J-Y, Guttman DS, et al. 2013.** The Arabidopsis ZED1 pseudokinase is required for ZAR1-mediated immunity induced by the *Pseudomonas syringae* type III effector HopZ1a. *Proceedings of the National Academy of Sciences* **110**: 18722–18727.
- Li J, Ding J, Zhang W, Zhang Y, Tang P, Chen JQ, Tian D, Yang S. 2010.** Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Molecular Genetics and Genomics* **283**: 427–438.
- Lu Y, Ran J-H, Guo D-M, Yang Z-Y, Wang X-Q. 2014.** Phylogeny and Divergence Times of Gymnosperms Inferred from Single-Copy Nuclear Genes (S Buerki, Ed.). *PLoS ONE* **9**: e107679.
- Luo S, Zhang Y, Hu Q, Chen J, Li K, Lu C, Liu H, Wang W, Kuang H. 2012.** Dynamic Nucleotide-Binding Site and Leucine-Rich Repeat-Encoding Genes in the Grass Family. *PLANT PHYSIOLOGY* **159**: 197–210.
- Ma Y, Guo H, Hu L, Martinez PP, Moschou PN, Cevik V, Ding P, Duxbury Z, Sarris PF, Jones JDG. 2018.** Distinct modes of derepression of an Arabidopsis immune receptor complex by two different bacterial effectors. *Proceedings of the National Academy of Sciences of the United States of America* **115**: 10218–10227.
- Mandadi KK, Scholthof K-BG. 2015.** Genome-wide analysis of alternative splicing landscapes modulated during plant-virus interactions in *Brachypodium distachyon*. *The Plant cell* **27**: 71–85.
- Mastrangelo AM, Marone D, Laidò G, De Leonardis AM, De Vita P. 2012.** Alternative splicing: Enhancing ability to cope with stress via transcriptome plasticity. *Plant Science* **185–186**: 40–49.
- McHale L, Tan X, Koehl P, Michelmore RW. 2006.** Plant NBS-LRR proteins: Adaptable guards. *Genome Biology* **7**: 212.
- Mestre P, Baulcombe DC. 2006.** Elicitor-mediated oligomerization of the tobacco N disease resistance protein. *The Plant Cell Online* **18**: 491–501.
- Meunier E, Broz P. 2017.** Evolutionary Convergence and Divergence in NLR Function and Structure. *Trends in immunology* **38**: 744–757.
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. 1999.** Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *The Plant journal : for cell and molecular biology* **20**: 317–32.
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. 2003.** Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *The Plant cell* **15**: 809–34.

- Meziadi C, Richard MMS, Derquennes A, Thareau V, Blanchet S, Gratiás A, Pflieger S, Geffroy V. 2016.** Development of molecular markers linked to disease resistance genes in common bean based on whole genome sequence. *Plant Science* **242**: 351–357.
- Michael Weaver L, Swiderski MR, Li Y, Jones JDG. 2006.** The Arabidopsis thaliana TIR-NB-LRR R-protein, RPP1A; protein localization and constitutive activation of defence by truncated alleles in tobacco and Arabidopsis. *The Plant Journal* **47**: 829–840.
- Michelmore RW, Meyers BC, Young ND. 1998.** Cluster of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Current Opinion in Plant Biology* **3**: 285–90.
- Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM. 1998.** The root knot nematode resistance gene Mi from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *The Plant cell* **10**: 1307–19.
- Mondragón-Palomino M, Meyers BC, Michelmore RW, Gaut BS. 2002.** Patterns of positive selection in the complete NBS-LRR gene family of Arabidopsis thaliana. *Genome Research* **12**: 1305–1315.
- Mota MM, Takemoto S, Takeuchi Y, Hara N, Futai K. 2006.** Comparative Studies between Portuguese and Japanese Isolates of the Pinewood Nematode, *Bursaphelenchus xylophilus*. *Journal of nematology* **38**: 429–433.
- Mucyn TS, Clemente A, Andriotis VME, Balmuth AL, Oldroyd GED, Staskawicz BJ, Rathjen JP. 2006.** The Tomato NBARC-LRR Protein Prf Interacts with Pto Kinase in Vivo to Regulate Specific Plant Immunity. *the Plant Cell Online* **18**: 2792–2806.
- Mun J-H, Yu H-J, Park S, Park B-S. 2009.** Genome-wide identification of NBS-encoding resistance genes in Brassica rapa. *Molecular genetics and genomics : MGG* **282**: 617–31.
- Munch D, Gupta V, Bachmann A, Busch W, Kelly S, Mun T, Andersen SU. 2018.** The Brassicaceae Family Displays Divergent, Shoot-Skewed NLR Resistance Gene Expression. *Plant Physiology* **176**: 1598–1609.
- Narusaka M, Kubo Y, Hatakeyama K, Imamura J, Ezura H, Nanasato Y, Tabei Y, Takano Y, Shirasu K, Narusaka Y. 2013.** Breaking restricted taxonomic functionality by dual resistance genes. *Plant Signaling and Behavior* **8**: 2–4.
- 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. *Plant Journal* **60**: 218–226.
- Nicol JM, Turner SJ, Coyne DL, Nijls L Den, Hockland S, Maafi ZT. 2011.** Current Nematode Threats to World Agriculture. In: Genomics and Molecular Genetics of Plant-Nematode Interactions. Dordrecht: Springer Netherlands, 21–43.
- Nishimura MT, Anderson RG, Cherkis KA, Law TF, Liu QL, Machius M, Nimchuk ZL, Yang L, Chung E-H, El Kasmi F, et al. 2017.** TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in Arabidopsis. *Proceedings of the National Academy of Sciences* **114**: E2053–E2062.
- Nyczepir A, Esmenjaud D. 2007.** Peach nematodes. In: Wayne D, Bassi D, eds. The peach : botany, production and uses. New York, NY, USA, 505–535.
- Van Ooijen G, Mayr G, Albrecht M, Cornelissen BJC, Takken FLW. 2008.** Transcomplementation, but not physical association of the CC-NB-ARC and LRR domains of tomato R protein Mi-1.2 is altered by mutations in the ARC2 subdomain. *Molecular Plant* **1**: 401–410.
- van Ooijen G, Mayr G, Kasiem MMAA, Albrecht M, Cornelissen BJCC, Takken FLWW. 2008.** Structure–function analysis of the NB-ARC domain of plant disease resistance proteins. *Journal of Experimental Botany* **59**: 1383–1397.
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al. 2009.** The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**: 551–556.
- Peart JR, Mestre P, Lu R, Malcuit I, Baulcombe DC. 2005.** NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Current Biology* **15**: 968–973.
- Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ. 2009.** Genome-wide analysis of Carica papaya reveals a small NBS resistance gene family. *Molecular genetics and genomics : MGG* **281**: 609–26.
- Potter D, Eriksson T, Evans RC, Oh S, Smedmark JEE, Morgan DR, Kerr M, Robertson KR, Arsenault M, Dickinson TA, et al. 2007.** Phylogeny and classification of Rosaceae.
- Prot JC. 1984.** A naturally occurring resistance breaking biotype of Meloidogyne arenaria on tomato. Reproduction and pathogenicity on tomato cultivars Roma and Rossol. *Revue Nématol* **7**: 23–28.
- Qi T, Seong K, Thomazella DPT, Kim JR, Pham J, Seo E, Cho M-J, Schultink A, Staskawicz BJ. 2018.** NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in Nicotiana benthamiana. *Proceedings of the National Academy of Sciences of the United States of America* **115**: E10979–E10987.

- Qian L-H, Zhou G-C, Sun X-Q, Lei Z, Zhang Y-M, Xue J-Y, Hang Y-Y. 2017.** Distinct Patterns of Gene Gain and Loss: Diverse Evolutionary Modes of NBS-Encoding Genes in Three Solanaceae Crop Species. *G3 (Bethesda, Md.)* **7**: 1577–1585.
- Quist CW, Smant G, Helder J. 2015.** Evolution of Plant Parasitism in the Phylum Nematoda. *Annual Review of Phytopathology* **53**: 289–310.
- Radwan O, Gandhi S, Heesacker A, Whitaker B, Taylor C, Plocik A, Kesseli R, Kozik A, Michelmore RW, Knapp SJ. 2008.** Genetic diversity and genomic distribution of homologs encoding NBS-LRR disease resistance proteins in sunflower. *Molecular Genetics and Genomics* **280**: 111–125.
- Rathjen JP, Dodds PN. 2017.** Dancing with the Stars: An Asterid NLR Family. *Trends in plant science* **22**: 1003–1005.
- Rayson S, Arciga-Reyes L, Wootton L, De Torres Zabala M, Truman W, Graham N, Grant M, Davies B. 2012.** A Role for Nonsense-Mediated mRNA Decay in Plants: Pathogen Responses Are Induced in Arabidopsis thaliana NMD Mutants (E Newbigin, Ed.). *PLoS ONE* **7**: e31917.
- Rehder A. 1940.** *Manual of Cultivated Trees and Shrubs Hardy in North America Exclusive of the Subtropical and Warmer Temperate Regions* (Collier Macmillan Ltd, Ed.). New York, The Macmillan Company.
- Riehs-Kearnan N, Gloggnitzer J, Dekrout B, Jonak C, Riha K. 2012.** Aberrant growth and lethality of Arabidopsis deficient in nonsense-mediated RNA decay factors is caused by autoimmune-like response. *Nucleic Acids Research* **40**: 5615–5624.
- Rosso M-N, Favery B, Piotte C, Arthaud L, De Boer JM, Hussey RS, Bakker J, Baum TJ, Abad P. 1999.** Isolation of a cDNA Encoding a β -1,4-endoglucanase in the Root-Knot Nematode *Meloidogyne incognita* and Expression Analysis During Plant Parasitism. *Molecular Plant-Microbe Interactions* **12**: 585–591.
- Le Roux C, Huet G, Jauneau A, Camborde L, Trémousaygue D, Kraut A, Zhou B, Levaillant M, Adachi H, Yoshioka H, et al. 2015.** A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* **161**: 1074–1088.
- Rubio-Cabetas MJ, Minot JC, Voisin R, Esmenjaud D. 2001.** Interaction of root-knot nematodes (RKN) and the bacterium *Agrobacterium tumefaciens* in roots of *Prunus cerasifera*: Evidence of the protective effect of the Ma RKN resistance genes against expression of crown gall symptoms. *European Journal of Plant Pathology* **107**: 433–441.
- Sarris PF, Cevik V, Dagdas G, Jones JDG, Krasileva K V. 2016.** Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biology* **14**.
- Saucet SB, Van Ghelder C, Abad P, Duval H, Esmenjaud D. 2016.** Resistance to root-knot nematodes *Meloidogyne* spp. in woody plants. *New Phytologist* **211**: 41–56.
- Schorneck S, Ballvora A, Gürlebeck D, Peart J, Ganai M, Baker B, Bonas U, Lahaye T. 2004.** The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. *Plant Journal* **37**: 46–60.
- Schreiber KJ, Bentham A, Williams SJ, Kobe B, Staskawicz BJ. 2016.** Multiple Domain Associations within the Arabidopsis Immune Receptor RPP1 Regulate the Activation of Programmed Cell Death. *PLoS Pathogens* **12**: 1–26.
- Schuler MA, Werck-Reichhart D. 2003.** Functional Genomics of P450S. *Annual Review of Plant Biology* **54**: 629–667.
- Segretin ME, Pais M, Franceschetti M, Chaparro-Garcia A, Bos JIB, Banfield MJ, Kamoun S. 2014.** Single Amino Acid Mutations in the Potato Immune Receptor R3a Expand Response to Phytophthora Effectors. *Molecular Plant-Microbe Interactions* **27**: 624–637.
- Shao ZQ, Xue JY, Wang Q, Wang B, Chen JQ. 2018.** Revisiting the Origin of Plant NBS-LRR Genes. *Trends in Plant Science* **24**: 9–12.
- Shao Z-Q, Xue J-Y, Wu P, Zhang Y-M, Wu Y, Hang Y-Y, Wang B, Chen J-Q. 2016.** Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. *Plant physiology* **170**: 2095–109.
- Shao Z-Q, Zhang Y-M, Hang Y-Y, Xue J-Y, Zhou G-C, Wu P, Wu X-Y, Wu X-Z, Wang Q, Wang B, et al. 2014.** Long-term evolution of nucleotide-binding site-leucine-rich repeat genes: understanding gained from and beyond the legume family. *Plant physiology* **166**: 217–34.
- Sharpe RH, Hesse CO, Lownsbery BF, Perry VG, Hansen CJ. 1969.** Breeding peaches for the root-knot nematode resistance. *Journal of the American Society for Horticultural Science* **94**: 209–212.
- Shirasawa K, Isuzugawa K, Ikenaga M, Saito Y, Yamamoto T, Hirakawa H, Isobe S. 2017.** The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. *DNA Research* **24**: 499–508.

- Shiu S, Bleecker AB. 2003.** Expansion of the Receptor-Like Kinase / Pelle Gene Family and Receptor-Like Proteins in Arabidopsis. *Plant physiology* **132**: 530–543.
- Shukla N, Yadav R, Kaur P, Rasmussen S, Goel S, Agarwal M, Jagannath A, Gupta R, Kumar A. 2018.** Transcriptome analysis of root-knot nematode (*Meloidogyne incognita*)-infected tomato (*Solanum lycopersicum*) roots reveals complex gene expression profiles and metabolic networks of both host and nematode during susceptible and resistance responses. *Molecular Plant Pathology* **19**: 615–633.
- Singh S, Singh B, Singh AP. 2015.** Nematodes: A Threat to Sustainability of Agriculture. *Procedia Environmental Sciences* **29**: 215–216.
- Slootweg E, Roosien J, Spiridon LN, Petrescu A-J, Tameling W, Joosten M, Pomp R, van Schaik C, Dees R, Borst JW, et al. 2010.** Nucleocytoplasmic Distribution Is Required for Activation of Resistance by the Potato NB-LRR Receptor Rx1 and Is Balanced by Its Functional Domains. *The Plant Cell* **22**: 4195–4215.
- Slootweg EJ, Spiridon LN, Roosien J, Butterbach P, Pomp R, Westerhof L, Wilbers R, Bakker E, Bakker J, Petrescu A-J, et al. 2013.** Structural Determinants at the Interface of the ARC2 and Leucine-Rich Repeat Domains Control the Activation of the Plant Immune Receptors Rx1 and Gpa2. *Plant Physiology* **162**: 1510–1528.
- Sohn KH, Segonzac C, Rallapalli G, Sarris PF, Woo JY, Williams SJ, Newman TE, Paek KH, Kobe B, Jones JDG. 2014.** The Nuclear Immune Receptor RPS4 Is Required for RRS1SLH1-Dependent Constitutive Defense Activation in Arabidopsis thaliana. *PLoS Genetics* **10**.
- Steele JF, Hughes RK, Banfield MJ. 2019.** Structural and biochemical studies of an NB-ARC domain from a plant NLR immune receptor. *bioRxiv* **557280**: 1–29.
- Tabata S, Kaneko T, Nakamura Y, Kotani H, Kato T, Asamizu E, Miyajima N, Sasamoto S, Kimura T, Hosouchi T, et al. 2000.** Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana. *Nature* **408**: 823–6.
- Takemoto D, Rafiqi M, Hurley U, Lawrence GJ, Bernoux M, Hardham AR, Ellis JG, Dodds PN, Jones DA. 2012.** N-Terminal Motifs in Some Plant Disease Resistance Proteins Function in Membrane Attachment and Contribute to Disease Resistance. *Molecular Plant-Microbe Interactions* **25**: 379–392.
- Takken FL, Albrecht M, Tameling W IL. 2006.** Resistance proteins: molecular switches of plant defence. *Current Opinion in Plant Biology* **9**: 383–390.
- Tameling WIL, Elzinga SDJ, Darmin PS, Vossen JH, Takken FLW, Haring MA, Cornelissen BJC. 2002.** The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. *The Plant cell* **14**: 2929–39.
- Tarr DEK, Alexander HM. 2009.** TIR-NBS-LRR genes are rare in monocots: Evidence from diverse monocot orders. *BMC Research Notes* **2**: 1–10.
- Thomas VP, Fudali SL, Schaff JE, Liu Q, Scholl EH, Opperman CH, Bird DM, Williamson VM. 2012.** A Sequence-Anchored Linkage Map of the Plant–Parasitic Nematode *Meloidogyne hapla* Reveals Exceptionally High Genome-Wide Recombination. *G3: Genes|Genomes|Genetics* **2**: 815–824.
- Triantaphyllou AC. 1973.** Environmental Sex Differentiation of Nematodes in Relation to Pest Management. *Annual Review of Phytopathology* **11**: 441–462.
- Truong NM, Nguyen C-N, Abad P, Quentin M, Favery B. 2015.** Function of Root-Knot Nematode Effectors and Their Targets in Plant Parasitism. In: 293–324.
- Urbach JM, Ausubel FM. 2017.** The NBS-LRR architectures of plant R-proteins and metazoan NLRs evolved in independent events. *Proceedings of the National Academy of Sciences of the United States of America* **114**: 1063–1068.
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al. 2010.** The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nature Genetics* **42**: 833–839.
- Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, et al. 2013.** The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics* **45**: 487–494.
- Voisin R, Rubio-Cabetas MJ, Minot JC, Esmenjaud D. 1999.** Penetration, development and emigration of juveniles of the nematode *Meloidogyne arenaria* in Myrobalan plum (*Prunus cerasifera*) clones bearing the Ma resistance genes. *European Journal of Plant Pathology* **105**: 103–108.
- Wagner S, Stuttmann J, Rietz S, Guerois R, Brunstein E, Bautor J, Niefind K, Parker JE. 2013.** Structural Basis for Signaling by Exclusive EDS1 Heteromeric Complexes with SAG101 or PAD4 in Plant Innate Immunity. *Cell Host & Microbe* **14**: 619–630.

- Wan H, Yuan W, Bo K, Shen J, Pang X, Chen J. 2013. Genome-wide analysis of NBS-encoding disease resistance genes in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. *BMC genomics* 14: 109.
- Wang J, Hu M, Wang J, Qi J, Han Z, Wang G, Qi Y, Wang H-W, Zhou J-M, Chai J. 2019a. Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science (New York, N.Y.)* 364.
- Wang G, Roux B, Feng F, Guy E, Li L, Li N, Zhang X, Lautier M, Jardinaud MF, Chabannes M, et al. 2015. The Decoy Substrate of a Pathogen Effector and a Pseudokinase Specify Pathogen-Induced Modified-Self Recognition and Immunity in Plants. *Cell Host and Microbe* 18: 285–295.
- Wang J, Wang J, Hu M, Wu S, Qi J, Wang G, Han Z, Qi Y, Gao N, Wang H-W, et al. 2019b. Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science (New York, N.Y.)* 364.
- Watkins R. 1976. Cherry, plum, peach, apricot and almond. *Prunus* spp. (Rosaceae). In: Evolution of Crop Plants. 242–247.
- Wei X, Wang R, Cao L, Yuan N, Huang J, Qiao W, Zhang W, Zeng H, Yang Q. 2012. Origin of *Oryza sativa* in China Inferred by Nucleotide Polymorphisms of Organelle DNA. *PLoS ONE* 7: 1–9.
- Whitham S, Dinesh-Kumar SP, Choi D, Hehl R, Corr C, Baker B. 1994. The product of the tobacco mosaic virus resistance gene N: similarity to toll and the interleukin-1 receptor. *Cell* 78: 1101–15.
- Williams SJ, Sohn KH, Wan L, Bernoux M, Sarris PF, Segonzac C, Ve T, Ma Y, Saucet SB, Ericsson DJ, et al. 2014. Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science* 344: 299–303.
- Williams SJ, Sornaraj P, deCourcy-Ireland E, Menz RI, Kobe B, Ellis JG, Dodds PN, Anderson PA. 2011. An Autoactive Mutant of the M Flax Rust Resistance Protein Has a Preference for Binding ATP, Whereas Wild-Type M Protein Binds ADP. *Molecular Plant-Microbe Interactions* 24: 897–906.
- Williams SJ, Yin L, Foley G, Casey LW, Outram MA, Ericsson DJ, Lu J, Boden M, Dry IB, Kobe B. 2016. Structure and Function of the TIR Domain from the Grape NLR Protein RPV1. *Frontiers in Plant Science* 7: 1–13.
- Wróblewski T, Spiridon L, Martin EC, Petrescu A-J, Cavanaugh K, Truco MJ, Xu H, Gozdowski D, Pawlowski K, Michelmore RW, et al. 2018. Genome-wide functional analyses of plant coiled-coil NLR-type pathogen receptors reveal essential roles of their N-terminal domain in oligomerization, networking, and immunity. *PLoS biology* 16: e2005821.
- Wu C-H, Abd-El-Halim A, Bozkurt TO, Belhaj K, Terauchi R, Vossen JH, Kamoun S. 2017. NLR network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences of the United States of America* 114: 8113–8118.
- Xiao S, Calis O, Patrick E, Zhang G, Charoenwattana P, Muskett P, Parker JE, Turner JG. 2005. The atypical resistance gene, RPW8, recruits components of basal defence for powdery mildew resistance in *Arabidopsis*. *Plant Journal* 42: 95–110.
- Xiao S, Ellwood S, Calis O, Patrick E, Li T, Coleman M, Turner JG. 2001. Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by RPW8. *Science (New York, N.Y.)* 291: 118–20.
- Xu Y, Wang J, Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, et al. 2013. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Genetics* 45: 51–58.
- Xue B, Hamamouch N, Li C, Huang G, Hussey RS, Baum TJ, Davis EL. 2013. The 8D05 parasitism gene of *Meloidogyne incognita* is required for successful infection of host roots. *Phytopathology* 103: 175–81.
- Xue J-Y, Wang Y, Wu P, Wang Q, Yang L-T, Pan X-H, Wang B, Chen J-Q. 2012. A primary survey on bryophyte species reveals two novel classes of nucleotide-binding site (NBS) genes. *PloS one* 7: e36700.
- Yang S, Tang F, Zhu H. 2014. Alternative splicing in plant immunity. *International Journal of Molecular Sciences* 15: 10424–10445.
- Yang S, Zhang X, Yue J-X, Tian D, Chen J-Q. 2008. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Molecular genetics and genomics : MGG* 280: 187–98.
- Yue J, Meyers BC, Chen J, Tian D, Yang S. 2012. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *The New phytologist* 193: 1049–63.
- Zhai J, Jeong D-H, De Paoli E, Park S, Rosen BD, Li Y, Gonzalez AJ, Yan Z, Kitto SL, Grusak MA, et al. 2011. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes & Development* 25: 2540–2553.
- Zhang X, Bernoux M, Bentham AR, Newman TE, Ve T, Casey LW, Raaymakers TM, Hu J, Croll TI, Schreiber KJ, et al. 2017a. Multiple functional self-association interfaces in plant TIR domains. *Proceedings of the National Academy of Sciences* 114: E2046–E2052.

- Zhang Q, Chen W, Sun L, Zhao F, Huang B, Wang J, Yang W, Tao Y, Yuan Z, Fan G, et al. 2012.** The genome of *Prunus mume*. *Nature Communications* **3**: 1318.
- Zhang X, Dodds PN, Bernoux M. 2017b.** What Do We Know About NOD-Like Receptors in Plant Immunity? *Annual review of phytopathology* **55**: 205–229.
- Zhang Y, Dorey S, Swiderski M, Jones JDG. 2004.** Expression of RPS4 in tobacco induces an AvrRps4-independent HR that requires EDS1, SGT1 and HSP90. *The Plant Journal* **40**: 213–224.
- Zhang X, Feng Y, Cheng H, Tian D, Yang S, Chen JQ. 2011.** Relative evolutionary rates of NBS-encoding genes revealed by soybean segmental duplication. *Molecular genetics and genomics : MGG* **285**: 79–90.
- Zhang J, Zhou JM. 2010.** Plant immunity triggered by microbial molecular signatures. *Molecular Plant* **3**: 783–793.
- Zhao T, Rui L, Li J, Nishimura MT, Vogel JP, Liu N, Liu S, Zhao Y, Dangl JL, Tang D. 2015.** A Truncated NLR Protein, TIR-NBS2, Is Required for Activated Defense Responses in the *exo70B1* Mutant. *PLoS Genetics* **11**: 1–28.
- Zhebentyayeva T, Shankar V, Scorza R, Callahan A, Ravelonandro M, Castro S, DeJong T, Saski CA, Dardick C. 2019.** Genetic characterization of worldwide *Prunus domestica* (plum) germplasm using sequence-based genotyping. *Horticulture Research* **6**: 12.
- Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D. 2004.** Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular Genetics and Genomics* **271**: 402–415.
- Zhu X, Xiao K, Cui H, Hu J. 2017.** Overexpression of the *Prunus sogdiana* NBS-LRR subgroup gene *PsoRPM2* promotes resistance to the root-knot nematode *Meloidogyne incognita* in tobacco. *Frontiers in Microbiology* **8**: 1–13.

ANNEXE 1

Liste des différents marqueurs moléculaires (amorces) utilisés dans le marquage fin du gène *RMja*.

NOM	TYPE	SEQUENCE	Fluorochrome
CPSCT004-F CPSCT004-R	Microsatellite	GCTCTGAAGCTCTGCATTGA TTTGAAATGGCTATGGAGTACG	NED
CPPCT039-F CPPCT039-R	Microsatellite	GCACCAGTTCTTCGTCATCTC GCATGCATAAAACCTTTATTGG	PET
CPPCT050-F CPPCT050-R	Microsatellite	CAGGACTAGCAAAGCCGAAG CGGAAATGGAAACGGTTTAG	FAM
LRR65-F LRR65-R	Microsatellite	AGAGGATAGGGAGCAACTCATT CGGAAGAATCTAGAGAGACTGACA	FAM
LRR25-F1 LRR25-R	Microsatellite	GGCTTGCAATCAAATCAACA GTGAATTGAATGCGACCACA	FAM
LRR5-F LRR5-R	Microsatellite	CGGCAGCAGAGGACAATTAT TTCTCAACGTCTCAGGGTGT	FAM
TIRRMjaGC3F NBSRMjaGC3R	SCAR	GCCAACCTCTCGGATATGAT AGTCCCTATACGACTCATGTTGGAT	-
PL3F PL4F PL5R1 PL5R2 PL5R3	SCAR	GGTAGGATCATCTAGTATAACTGGC GGTATCATCTAGTAGGCCTAGCAG AAGAGAGACAAGGTATCAATGTATT GTTATATATGGAGTGACGATCAAAG AGATTGTTGCCGAGGAGTCAT	-
EPISLRRF PL2F2-RC	SCAR	TACCTGAAAGCATCTCTCAGC AATAATGAGGAGTTCGATGTGA	-
ORF5F LRRF1-RC	SCAR	AAGAACAGCAGGACATGTTTC AGGATTAGCTGCTCAAGGTT	-
TIRRMjaGC2F NBSRMjaGC2R	SCAR	GCCGAAGGATTTGATCTGAC CCTTTAAATCATCATTTGAGAATG	-
TIRRMjaGC1F TIRRMjaGC1R	SCAR	CCATTTGACACTCTTGAAAGC ACCTCTTACATCAAATCCTGAGAGG	-
ENDKIN-F ENDKIN-R	SCAR	CCACTTGAGGAAGAAGAGGAG AGGAGGTTTGGGTGGTCAT	-
KIN35-F1 KIN35-R1	Microsatellite	ACGCTTGTTGCGCTCTTAAT CGCGCAAATCATAAAGACA	PET
CPPCT022-F CPPCT022-R	Microsatellite	CAATTAGCTAGAGAGAATTATTG GACAAGAAGCAAGTAGTTTG	VIC
CPSCT026-F CPSCT026-R	Microsatellite	TCTCACACGCTTTCGTCAAC AAAAAGCCAAAAGGGGTTGT	NED
UDP98-405F UDP98-405R	Microsatellite	ACGTGATGAACTGACACCCA GAGTCTTTGCTCTGCCATCC	PET
ACTPrupe.2G235000F1 ACTPrupe.2G235000R1	qPCR	GAAGATCAAGGTGGTGGCAC GATGGCCCAGACTCGTCATA	
CT3F CT3R	SCAR	AGCAACATTGTAAATCCTGTAAATC CGATCAAAGTCCCATTTCTCATC	
QCT3F QCT3R	qPCR	AGCAACATTGTAAATCCTGTAAATC AAACCACTCCAGAGTGATGCTTGAA	

ANNEXE 2

Liste des logiciels utilisés dans les différentes études présentées dans ce manuscrit.

Nom	Type	Lien	Référence
BLAST	Recherche de séquences par similarité	https://blast.ncbi.nlm.nih.gov/Blast.cgi et LOCAL	Altschul, S.F., 1997
Clustal Omega MUSCLE T-Coffee	Alignement de séquences	https://www.ebi.ac.uk/Tools/msa/	The EMBL-EBI bioinformatics web and programmatic tools framework. 2015
Dendroscope 3	Créateur d'arbres phylogénétiques	LOCAL	Huson, D.H. & Scornavacca, C. 2012
FGENESH	Prédicteur de gènes	http://www.softberry.com/	Solovyev V, et al. 2006
HMMscan	Annotation fonctionnelle de protéines	https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan	Potter, S.C. et al., 2018
IGV	Visualisation de génome (génome browser)	LOCAL	Robinson, J.T., et al., 2011
InterProScan	Annotation fonctionnelle de protéines	https://www.ebi.ac.uk/interpro/search/sequence-search	Mitchell et al. 2019
i-tasser	Prédicteur de structure 3D	https://zhanglab.ccmb.med.umich.edu/I-TASSER/	Yang, J., et al. 2015
Jpred 4	Prédicteur de structure secondaire	http://www.compbio.dundee.ac.uk/jpred4/	Drozdetskiy A, et al., 2015
MEGA7	Phylogénie et alignement	LOCAL	Kumar, S. et al., 2015
MEME suite	Recherche de motifs	http://meme-suite.org/index.html	Timothy L. et al., 2009
Pfam	Annotation fonctionnelle de protéines	http://pfam.xfam.org/	El-Gebali, S., et al. 2019
Phylogeny.fr	Phylogénie	http://www.phylogeny.fr/	Dereeper A., et al., 2008
SPDBV v4.1.0	Visualisation de structure protéique 3D	LOCAL	Guex, N., et al., 1996
SUPERFAMILY	Annotation fonctionnelle de protéines	http://beta.supfam.org/sequence/search	Pandurangan, A.P. et al., 2018.

ANNEXE 3

Liste détaillée des bases de données dont sont issus les données additionnelles utilisées dans les différentes études du manuscrit.

Espèces et versions des données	Bases de données	Liens
<i>Amaranthus hypochondriacus</i> v1.0	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Amborella trichopoda</i> v1.0	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Ananas comosus</i> v3	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Aquilegia coerulea</i> v3.1	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Arabidopsis lyrata</i> v2.1	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Arabidopsis thaliana</i> Araport11	TAIR	https://www.arabidopsis.org/
<i>Arachis duranensis</i>	LIS - legume Information System	https://legumeinfo.org
<i>Artocarpus camansi</i> Acamansi1.0	NCBI Genome information	https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/
<i>Boehmeria nivea</i> ASM293701v1	NCBI Genome information	https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/
<i>Brachypodium distachyon</i> v3.1	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Brassica rapa</i> FPsc v1.3	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Cajanus cajan</i>	LIS - legume Information System	https://legumeinfo.org
<i>Cannabis sativa</i> ASM341772v2	NCBI Genome information	https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/
<i>Capsella rubella</i> v1.0	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Carica papaya</i> ASGPBv0.4	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Chara braunii</i> Cbr_1.0	NCBI Genome information	https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/
<i>Citrullus lanatus</i> 97103 v2	CuGenDB	http://cucurbitgenomics.org/
<i>Citrus sinensis</i> v1.1	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Coffea canephora</i> v1.0	Sol Genomics Network	https://solgenomics.net/

<i>Cpsicum annuum</i> cv CM334 1.55	Sol Genomics Network	https://solgenomics.net/
<i>Cucumis melo</i> DHL92 v4.0	CuGenDB	http://cucurbitgenomics.org/
<i>Cucumis sativus</i> v1.0	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Cycas revoluta</i> PRJNA255888 TSA GBJU01.1	NCBI Sequence Set Browser	https://trace.ncbi.nlm.nih.gov/Traces/wgs/
<i>Ephedra trifurca</i> PRJNA255888 TSA GBKT01.1	NCBI Sequence Set Browser	https://trace.ncbi.nlm.nih.gov/Traces/wgs/
<i>Ficus carica</i> <i>F.carica_assembly01</i>	NCBI Genome information	https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/
<i>Fragaria vesca</i> v1.0 <i>ab initio gene proteins</i>	GDR - Genome Database for Rosaceae	https://www.rosaceae.org/
<i>Ginkgo biloba</i> <i>Gb.pep.fa TM011</i>	Guan et al. 2016	http://gigadb.org/dataset/100209
<i>Glycine max</i> <i>Wm82.a2.v1</i>	LIS - legume Information System	https://legumeinfo.org
<i>Kalanchoe laxiflora</i> v1.1	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Klebsormidium nitens</i> ASM70883v1	NCBI Genome information	https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/
<i>Linum usitatissimum</i> v1.0	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Malus domestica</i> V1.0 <i>predicted peptides</i>	GDR - Genome Database for Rosaceae	https://www.rosaceae.org/
<i>Marchantia</i> <i>polymorpha</i> v3.1	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Medicago truncatula</i> <i>Mt4.0v1</i>	LIS - legume Information System	https://legumeinfo.org
<i>Morus notabilis</i> ASM41409v2	NCBI Genome information	https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/
<i>Musa acuminata</i> v1	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Nicotiana</i> <i>benthamiana</i> v1.0.1	Sol Genomics Network	https://solgenomics.net/
<i>Oryza sativa</i> v7_JGI	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Phaseolus vulgaris</i>	LIS - legume Information System	https://legumeinfo.org
<i>Phoenix dactylifera</i> DPV01	NCBI Genome information	https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/

<i>Physcomitrella patens</i> v3.3	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Picea abies</i> v1.0 UPSC	CONGENIE	http://congenie.org/
<i>Pinus taeda</i> v1.0 TreeGenes	CONGENIE	http://congenie.org/
<i>Populus trichocarpa</i> v3.0	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Prunus avium</i> genome v1.0 proteins	GDR - Genome Database for Rosaceae	https://www.rosaceae.org/
<i>Prunus persica</i> v2.0.a1 all transcript peptide	GDR - Genome Database for Rosaceae	https://www.rosaceae.org/
<i>Pyrus bretschneideri</i> Rehd v1.0 hybrid proteins	GDR - Genome Database for Rosaceae	https://www.rosaceae.org/
<i>Salix purpurea</i> v1.0	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Selaginella</i> <i>moellendorffii</i> v1.0	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Solanum lycopersicum</i> ITAG release 3.20	Sol Genomics Network	https://solgenomics.net/
<i>Solanum</i> <i>pimpinellifolium</i> LA1589	Sol Genomics Network	https://solgenomics.net/
<i>Solanum tuberosum</i> PGSC DM v3.4	Sol Genomics Network	https://solgenomics.net/
<i>Sorghum bicolor</i> v3.1.1	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Sphagnum fallax</i> v0.5	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Vitis vinifera</i> Genoscope.12X	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Zea mays</i> Ensembl-18	PHYTOZOME v12.1	https://phytozome.jgi.doe.gov/pz/portal.html
<i>Ziziphus jujuba</i> ZizJuj_1.1	NCBI Genome information	https://www.ncbi.nlm.nih.gov/genome/browses/#!/overview/

ANNEXE 4

Liste détaillée des tailles de génomes et des effectifs de NBS-LRRs et de leur taux de clustérisation et références associées. Conifères (C), Monocotylédones (M), Dicotylédones (D), Rosides (R), Astérides (A).

Taxonomie	Nom de l'espèce	Génome (Gb)	NBS-LRRs	TNLs	% de clusterisation	Références
Algue verte	<i>Klebsormidium nitens</i>	-	7	3	-	(Gao <i>et al.</i> , 2018)
Algue verte	<i>Nitella mirabilis</i>	-	10	0	-	(Gao <i>et al.</i> , 2018)
Algue verte	<i>Coleochaete orbicularis</i>	-	28	0	-	(Gao <i>et al.</i> , 2018)
Marchantiophyta	<i>Marchantia polymorpha</i> (Hépatique des fontaines)	0.28	33	0	-	(Gao <i>et al.</i> , 2018)
Bryophyte	<i>Sphagnum fallax</i> (Sphaigne)	0.45	243	38	-	(Gao <i>et al.</i> , 2018)
Bryophyte	<i>Physcomitrella patens</i> (Mousse)	0.47	123	10	-	(Gao <i>et al.</i> , 2018)
Tracheophyte	<i>Selaginella moellendorffii</i>	0.10	17	0	-	(Gao <i>et al.</i> , 2018)
Gymnosperme	<i>Ginkgo biloba</i>	11.75	255	127	-	(Gao <i>et al.</i> , 2018)
Gymnosperme C	<i>Thuja occidentalis</i> (Thuja occidental)	-	486	317	-	(Van Ghelder <i>et al.</i> , 2019)
Gymnosperme C	<i>Larix laricina</i> (Mélèze laricin)	-	633	343	-	(Van Ghelder <i>et al.</i> , 2019)
Gymnosperme C	<i>Picea glauca</i> (Epinette blanche)	20.00	506	279	-	(Van Ghelder <i>et al.</i> , 2019)
Gymnosperme C	<i>Picea mariana</i> (Epinette noire)	17.50	725	379	-	(Van Ghelder <i>et al.</i> , 2019)
Gymnosperme C	<i>Pinus strobus</i> (Pin blanc)	31.20	560	339	-	(Van Ghelder <i>et al.</i> , 2019)
Gymnosperme C	<i>Pinus taeda</i> (Pin à torches)	22.00	679	247	-	(Gao <i>et al.</i> , 2018)
Gymnosperme C	<i>Pinus banksiana</i> (Pin gris)	21.50	486	250	-	(Van Ghelder <i>et al.</i> , 2019)
Gymnosperme C	<i>Abies balsamea</i> (Sapin baumier)	-	338	150	-	(Van Ghelder <i>et al.</i> , 2019)
Angiosperme	<i>Amborella trichopoda</i>	0.87	106	10	-	(Gao <i>et al.</i> , 2018)
Angiosperme M	<i>Spirodela polyrhiza</i>	0.16	81	0	-	(Gao <i>et al.</i> , 2018)

Angiosperme M	<i>Zea mays</i> (Maïs)	2.00	149	0	-	(Gao <i>et al.</i> , 2018)
Angiosperme M	<i>Zea mays</i> (Maïs)	2.30	129	0	-	(Li <i>et al.</i> , 2010)
Angiosperme M	<i>Sorghum bicolor</i> (Sorgo commun)	0.73	245	0	-	(Velasco <i>et al.</i> , 2010)
Angiosperme M	<i>Sorghum bicolor</i> (Sorgo commun)	0.73	211	0	-	(Paterson <i>et al.</i> , 2009)
Angiosperme M	<i>Sorghum bicolor</i> (Sorgo commun)	0.73	245	0	-	(Li <i>et al.</i> , 2010)
Angiosperme M	<i>Oryza sativa</i> (Riz)	0.39	508	0	-	(Li <i>et al.</i> , 2010)
Angiosperme M	<i>Oryza sativa</i> (Riz)	0.39	486	0	-	(Gao <i>et al.</i> , 2018)
Angiosperme M	<i>Oryza sativa</i> (Riz)	0.39	519	0	-	(International Rice Genome Sequencing Project, 2005)
Angiosperme M	<i>Oryza sativa</i> (Riz)	0.39	535	0	-	(Zhou <i>et al.</i> , 2004)
Angiosperme M	<i>Brachypodium distachyon</i> (Brachypode à deux épis)	0.36	316	0	-	(Gu <i>et al.</i> , 2015)
Angiosperme M	<i>Brachypodium distachyon</i> (Brachypode à deux épis)	0.36	239	0	-	(Li <i>et al.</i> , 2010)
Angiosperme M	<i>Aegilops tauschii</i> (Eglope)	4.30	842	0	-	(Gu <i>et al.</i> , 2015)
Angiosperme M	<i>Triticum urartu</i> (Blé rouge sauvage)	4.94	570	0	-	(Gu <i>et al.</i> , 2015)
Angiosperme M	<i>Triticum aestivum</i> (Blé)	17.00	2237	0	-	(Gu <i>et al.</i> , 2015)
Angiosperme M	<i>Hordeum vulgare</i> (Orge commune)	5.30	409	0	-	(Gao <i>et al.</i> , 2018)
Angiosperme M	<i>Hordeum vulgare</i> (Orge commune)	5.30	420	0	-	(Gu <i>et al.</i> , 2015)
Angiosperme D	<i>Aquilegia coerulea</i> (Ancolie bleue)	0.31	230	0	-	(Gao <i>et al.</i> , 2018)
Angiosperme D A	<i>Carica papaya</i> (Papaye)	0.37	54	10	-	(Porter <i>et al.</i> , 2009)
Angiosperme D A	<i>Carica papaya</i> (Papaye)	0.37	44	-	72	(Hofberger <i>et al.</i> , 2014)
Angiosperme D A	<i>Helianthus annuus</i> (Tournesol)	3.60	630	-	-	(Radwan <i>et al.</i> , 2008)
Angiosperme D A	<i>Nicotiana benthamiana</i> (Tabac benthamiana)	3.10	345	-	-	(Brendolise <i>et al.</i> , 2017)
Angiosperme D A	<i>Nicotiana benthamiana</i> (Tabac benthamiana)	3.10	233	-	25	(Hofberger <i>et al.</i> , 2014)

Angiosperme D A	<i>Solanum tuberosum</i> (Pomme de terre)	0.84	447	70	82.1	(Qian <i>et al.</i> , 2017)
Angiosperme D A	<i>Solanum tuberosum</i> (Pomme de terre)	0.84	438	77	73.24	(Jupe <i>et al.</i> , 2012)
Angiosperme D A	<i>Solanum tuberosum</i> (Pomme de terre)	0.84	402	-	59	(Hofberger <i>et al.</i> , 2014)
Angiosperme D A	<i>Solanum lycopersicum</i> (Tomate)	0.95	255	31	68.77	(Qian <i>et al.</i> , 2017)
Angiosperme D A	<i>Solanum lycopersicum</i> (Tomate)	0.95	326	40	-	(Andolfo <i>et al.</i> , 2014)
Angiosperme D A	<i>Solanum lycopersicum</i> (Tomate)	0.95	219	-	57	(Hofberger <i>et al.</i> , 2014)
Angiosperme D A	<i>Solanum pimpinellifolium</i> (Tomate groseille)	0.81	355	45	-	(Andolfo <i>et al.</i> , 2014)
Angiosperme D A	<i>Capsicum annuum</i> (Piment)	3.50	306	16	66.15	(Qian <i>et al.</i> , 2017)
Angiosperme D R	<i>Vitis vinifera</i> (Vigne)	0.49	315	26	-	(Gao <i>et al.</i> , 2018)
Angiosperme D R	<i>Vitis vinifera</i> (Vigne)	0.49	535	111	-	(Yang <i>et al.</i> , 2008)
Angiosperme D R	<i>Vitis vinifera</i> (Vigne)	0.49	341	-	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Vitis vinifera</i> (Vigne)	0.49	294	-	70	(Hofberger <i>et al.</i> , 2014)
Angiosperme D R	<i>Lotus japonicus</i> (Lotus)	0.48	158	60	-	(Zhang <i>et al.</i> , 2011)
Angiosperme D R	<i>Medicago truncatula</i> (Luzerne)	0.50	571	275	88.21	(Shao <i>et al.</i> , 2014)
Angiosperme D R	<i>Medicago truncatula</i> (Luzerne)	0.50	333	156	-	(Ameline-Torregrosa <i>et al.</i> , 2007)
Angiosperme D R	<i>Phaseolus vulgaris</i> (Haricot commun)	0.59	337	103	83.69	(Shao <i>et al.</i> , 2014)
Angiosperme D R	<i>Phaseolus vulgaris</i> (Haricot commun)	0.59	376	106	-	(Meziadi <i>et al.</i> , 2016)
Angiosperme D R	<i>Glycine max</i> (Soja)	1.12	465	178	74.73	(Shao <i>et al.</i> , 2014)
Angiosperme D R	<i>Glycine max</i> (Soja)	1.12	429	146	-	(Zhang <i>et al.</i> , 2011)
Angiosperme D R	<i>Glycine max</i> (Soja)	1.12	392	-	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Glycine max</i> (Soja)	1.12	319	116	-	(Kang <i>et al.</i> , 2012)
Angiosperme D R	<i>Glycine max</i> (soja)	1.12	372	-	-	(Velasco <i>et al.</i> , 2010)

Angiosperme D R	<i>Cajanus cajan</i> (Pois d'Angole)	0.83	289	112	71.01	(Shao <i>et al.</i> , 2014)
Angiosperme D R	<i>Fragaria vesca</i> (Fraisier)	0.24	346	153	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Prunus mume</i> (Abricotier japonais)	0.28	475	224	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Prunus persica</i> (Pêcher)	0.27	443	194	-	(Jia <i>et al.</i> , 2015; Van Ghelder & Esmenjaud, 2016)
Angiosperme D R	<i>Pyrus bretschneideri</i> (Poirier)	0.53	617	342	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Malus domestica</i> (Pommier)	0.74	992	268	-	(Velasco <i>et al.</i> , 2010)
Angiosperme D R	<i>Malus domestica</i> (Pommier)	0.74	1303	664	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Linum usitatissimum</i> (Lin)	0.37	147	98	-	(Kale <i>et al.</i> , 2013)
Angiosperme D R	<i>Cannabis sativa</i> (Cannabis)	0.83	234	-	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Ricinus communis</i> (Ricin commun)	0.35	130	34	-	(Zhang <i>et al.</i> , 2011)
Angiosperme D R	<i>Populus trichocarpa</i> (Peuplier de l'ouest)	0.49	402	125	-	(Kohler <i>et al.</i> , 2008)
Angiosperme D R	<i>Populus trichocarpa</i> (Peuplier de l'ouest)	0.49	416	88	-	(Yang <i>et al.</i> , 2008)
Angiosperme D R	<i>Populus trichocarpa</i> (Peuplier de l'ouest)	0.55	402	-	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Populus tremula</i> (Tremble)	0.44	419	99	-	(Gao <i>et al.</i> , 2018)
Angiosperme D R	<i>Cucumis sativus</i> (Concombre)	0.37	57	23	-	(Wan <i>et al.</i> , 2013)
Angiosperme D R	<i>Cucumis sativus</i> (Concombre)	0.37	71	33	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Cucumis melo</i> (Melon)	0.45	80	35	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Citrullus lanatus</i> (Pastèque)	0.43	45	20	-	(Jia <i>et al.</i> , 2015)
Angiosperme D R	<i>Citrullus lanatus</i> (Pastèque)	0.43	44	18	-	(Xu <i>et al.</i> , 2013)
Angiosperme D R	<i>Tarenaya hassleriana</i> (Cleome)	0.29	59	-	44	(Hofberger <i>et al.</i> , 2014)
Angiosperme D R	<i>Aethionema arabicum</i> (Aethionème)	0.20	112	49	63	(Hofberger <i>et al.</i> , 2014)
Angiosperme D R	<i>Eutrema parvulum</i> (Thellungiella)	0.14	72	-	51	(Hofberger <i>et al.</i> , 2014)
Angiosperme D R	<i>Brassica rapa</i> (Moutarde)	0.53	92	52	-	(Mun <i>et al.</i> , 2009)

Angiosperme D R	<i>Brassica rapa</i> (Moutarde)	0.53	167	-	55	(Hofberger <i>et al.</i> , 2014)
Angiosperme D R	<i>Arabidopsis lyrata</i> (Arabette lyrée)	0.23	185	123	63.8	(Guo <i>et al.</i> , 2011)
Angiosperme D R	<i>Arabidopsis lyrata</i> (Arabette lyrée)	0.23	166	-	43	(Hofberger <i>et al.</i> , 2014)
Angiosperme D R	<i>Arabidopsis thaliana</i> (Arabette des dames)	0.13	166	102	-	(Gao <i>et al.</i> , 2018)
Angiosperme D R	<i>Arabidopsis thaliana</i> (Arabette des dames)	0.13	174	-	-	(Tabata <i>et al.</i> , 2000)
Angiosperme D R	<i>Arabidopsis thaliana</i> (Arabette des dames)	0.13	209	147	-	(Meyers <i>et al.</i> , 2003)
Angiosperme D R	<i>Arabidopsis thaliana</i> (Arabette des dames)	0.13	159	98	71.1	(Guo <i>et al.</i> , 2011)
Angiosperme D R	<i>Arabidopsis thaliana</i> (Arabette des dames)	0.13	140	-	67	(Hofberger <i>et al.</i> , 2014)
Angiosperme D R	<i>Citrus sinensis</i> (Oranger)	0.37	455	-	30	(Hofberger <i>et al.</i> , 2014)

ANNEXE 5

Alignement de la zone allant du domaine LRR (bleu) au domaine PL5 chez Ma, RMja et deux isoformes détectées chez Prupe.7G65400. Les structures secondaires sont indiquées au-dessus (brins β , flèches vertes ; hélices α , tubes rouges). Les régions contenant les deux indels des PL3 et 4 sont encadrées en bleu.



S2 nd p	
Ma-LRR-5PLs	KFVRKLNKHYSYCYHGYQPEQEEDGMRSI LSTSGIQTEQEVEEQEETTSTRIVGQLRRN
RMja-LRR-5PLs	KFVRKLNKHYSYCYHGYQIEQEEDGMRSI PSTSGVQTEQEELQEETTSTNRIAGQLRRN
7G065400-LRR-5PLs-iso1	KFVQKLNKHYSYCYHGYQIEQEEDGMRSI PSTSRVQTEQEELQEETTSTRIAGQLRRN
7G065400-LRR-5PLs-iso2	KFVQKLNKHYSYCYHGYQIEQEEDGMRSI PSTSRVQTEQEELQEETTSTRIAGQLRRN
	.** ***** ** :*****:*****.*.*****
S2 nd p	
Ma-LRR-5PLs	VESLLEKLFEGLQQGLPNLYDYGFI FCLRERLPWFSEQSSTPACTVNLRLLPNLHNNEKW
RMja-LRR-5PLs	VESLLEKLFEGLQQGLPNLYDYGFI FCLRERLLWFSEQSSTPACTVNLRLLPNLHNDEKW
7G065400-LRR-5PLs-iso1	VESLLEKLFEGLQQGLPNLYDYGFI FCLRERLLWFSEQSSTPACTVNLRLLPNLHNDEKW
7G065400-LRR-5PLs-iso2	VESLLEKLFEGLQQGLPNLYDYGFI FCLRERLLWFSEQSSTPACTVNLRLLPNLHNDEKW
	*****.***** :***** *****:***
S2 nd p	
Ma-LRR-5PLs	AGLSLYVVCALPPGLPLFRFTFYECHLYTPIEAVGHEQLMHRMLLSSSWDDNVGSHRLLII
RMja-LRR-5PLs	AGLSLYVVYALPPGLPLFRFTFYECQLYTPVEAVGHEQLMHRMLLSPWDDNVGSHRLLII
7G065400-LRR-5PLs-iso1	AGLSLYVVCALPPGLPLFRFTFYECQLYTPVEAVGHEQLMHRMLLSPWDDNVGSHRLLII
7G065400-LRR-5PLs-iso2	AGLSLYVVYALPPGLPLFRFTFYECQLYTPVEAVGHEQLMHRMLLSPWDDNVGSHRLLII
	***** *****:****:***** ***** *.*****
S2 nd p	
Ma-LRR-5PLs	HIPRVRFPERLNRCHFIAQLFGCRTPGVEVKMCGMRLVYNQDLKGLIQTITRCTNDRPAY
RMja-LRR-5PLs	HIPRVRFPERLNRCHFIAQLFGCRTPGVEVEMCGMRLVYNQDLKGLIQTITRCTNDRPAY
7G065400-LRR-5PLs-iso1	HIPRVRFPERLNRCHFIAQLFGCRTPGVEVEMCGMRLVYNQDLKGLIQTITRCTNDRPAY
7G065400-LRR-5PLs-iso2	HIPRVRFPERLNRCHFIAQLFGCRTPGVEVEMCGMRLVYNQDLKGLIQTITRCTNDRPAY
	*****.*****:*****:***** *****:*****
S2 nd p	
Ma-LRR-5PLs	YGTGDFTEKTKYKGISLGATSLLTNLVEAAKSSEHSSVSEVCPFPMPIDFRPEPYSAEYS
RMja-LRR-5PLs	YGTGDFTDTKKYKGISLGATSLLTNLLAAKSSEHSSVSEVCPFCMPIDFRPETYSAEYS
7G065400-LRR-5PLs-iso1	YGTGDFTDTKKYKGISLGVTSLLTNLLAAKSSEHSSVSEVCPFCMPIDFRQETYSAEYS
7G065400-LRR-5PLs-iso2	YGTGDFTDTKKYKGISLGVTSLLTNLLAAKSSEHSSVSEVCPFCMPIDFRQETYSAEYS
	*****.*****.*****:*****:***** ***** ***** *
S2 nd p	
Ma-LRR-5PLs	QHERKTTPIHPLQYFGNAFYPTNKMSDDQNQLLDFEQDLKYNSCFPPNEI VEWFGHQSSG
RMja-LRR-5PLs	QHERKTTPIRPPQDFGSSFDPTNNMSSDQNQLLDFERDLKYNSCFPPNEI VEWFRHQSSG
7G065400-LRR-5PLs-iso1	QHERKTRPIRPPQDFGSSFDPTNNMSSDQNQLLDFERDLKYNSCFPPNEI VEWFRHQSSG
7G065400-LRR-5PLs-iso2	QHERKTRPIRPPQDFGSSFDPTNNMSSDQNQLLDFERDLKYNSCFPPNEI VEWFRHQSSG
	***** *. * * * * : * * * :*****.***** ***** *****
S2 nd p	
Ma-LRR-5PLs	PSVKIPLPSNLCEDTNWIGLALCAYFSVLDHSTIDLNLNPEISHNLTCLLETDESCLES
RMja-LRR-5PLs	PSVKIPLPSNLCEDTNWIGLALCAYFSVLDHSTIDLNLNPEISHNLTCLLETDESCLES
7G065400-LRR-5PLs-iso1	PSVKIPLPSNLCEDTNWIGLALCAYFSVLDHSTIDLNLNPEISHNLTCLLETDESCLES
7G065400-LRR-5PLs-iso2	PSVKIPLPSNLCEDTNWIGLALCAYFSVLDHSTIDLNLNPEISHNLTCLLETDESCLES
	*****.***** *****:*****.***** ***** *****
S2 nd p	
Ma-LRR-5PLs	LHGYSNTSQEFKWLRYRMGGFIWLSYI PRCWFSQDLKERGHLEASIGSDRGS LGVHRCGLR
RMja-LRR-5PLs	LHGYSNTSQEFEWLYRMGGFIWLSYI PRCWFSQDLKERGHLEASIGSDRGS LGVHRCGLR
7G065400-LRR-5PLs-iso1	LHGYSNTSQEFKWLRYRMGGFIWLSYI PRCWFSQDLKERGHLEASIGSDRGS LGVHRCGLR
7G065400-LRR-5PLs-iso2	LHGYSNTSQEFKWLRYRMGGFIWLSYI PRCWFSQDLKERGHLEASIGSDRGS LGVHRCGLR
	****.*****:*****.***** ***** ***** ***** *****
S2 nd p	
Ma-LRR-5PLs	LIYLEDEEGLKETIMHCMTSLSDINQGKDEQYQNC EAGSSITGSNI VNPVNPHERSEE
S2 nd p	
RMja-LRR-5PLs	LIYLEDEEGLKETIMHCMTSLSDINQGKDKQYQNC EAGSSITGSNV---VHPPLERSEK
7G065400-LRR-5PLs-iso1	LIYLEDEEGLKETIMHCMTSLSDINQGKDKQYQNC EAGSSITGSNV---VHPPLERSEK
7G065400-LRR-5PLs-iso2	LIYLEDEEGLKETIMHCMTSLSDINQGKDKQYQNC EAGSSITGSNV-----
	*****.*****:*****.***** ***** ***** ***** *****

S2ndP
 Ma-LRR-5PLs
 RMja-LRR-5PLs
 7G065400-LRR-5PLs-iso1
 7G065400-LRR-5PLs-iso2

PNDKKWNFGCHTMYNFCFPSSITLWFGDQSSGSSIRVPLPPLYRATNWIGLALCTSFS
 PNDKKWNFGCHLMYNFCFPSSITLWFGDQSSGSSIRVPLPPLYSATNWIGLALCASFS
 PNDKKWNFGCHSMYNFCFPSSITLWFGDQSSGSSIRVPLPPLYSATNWIGLALCASFS
 -----NFGCHSMYNFCFPSSITLWFGDQSSGSSIRVPLPPLYSATNWIGLALCASFS
 ***** **

S2ndP
 Ma-LRR-5PLs
 RMja-LRR-5PLs
 7G065400-LRR-5PLs-iso1
 7G065400-LRR-5PLs-iso2

IVDNPTADLDNLNPEISHHLICHLES DRGTIEPLHDYCTTNEEFQWL PFGGFIWVSYIPR
 IMENPTADLDNLNPEISHHLICHLES DRGTIEPLHDYCTTNEEFQWL PFGGFIWVSYIPR
 IMENPTADLDNLNPEISHHLICHLES DRGTIEPLHDYCTTNEEFQWL PFGGFIWVSYIPR
 IMENPTADLDNLNPEISHHLICHLES DRGTIEPLHDYCTTNEEFQWL PFGGFIWVSYIPR
 *.:*****

S2ndP
 Ma-LRR-5PLs
 RMja-LRR-5PLs
 7G065400-LRR-5PLs-iso1
 7G065400-LRR-5PLs-iso2

AWFSDQLNECDVLEASFASDHEAFIVHECGLRLVYQHDEEIKQTI LHYMTSLSDKKGKN
 VWFSDQLNECDVLEASFASDHEAFIVHECGLRLVYQHDEEIKQTI LHYMTSLSDKKVKN
 VWFSDQLNECDVLEASFASDHEAFIVHECGLRLVYQHDEEIKQTI LHYMTSLSDKKVKN
 VWFSDQLNECDVLEASFASDHEAFIVHECGLRLVYQHDEEIKQTI LHYMTSLSDKKVKN
 .*****

S2ndP
 Ma-LRR-5PLs
 RMja-LRR-5PLs
 7G065400-LRR-5PLs-iso1
 7G065400-LRR-5PLs-iso2

KQCPMGDTGSSSRPSSYIVKPHLERLGRPS**DEK**WDFDRHSIYNFCFPSSITLWFGHQSN
 KQCPTGEAVSSSRPSSYIVKPHLKR LGRPS---WDFDRHSIYNFCFPSSITLWFDQRSN
 KQCPTGEAVSSSRPSSYIVKPHLKR LGRPS---WDFDRHSIYNFCFPSSITLWFGQRSN
 KQCPTGEAVSSSRPSSYIVKPHLKR LGRPS---WDFDRHSIYNFCFPSSITLWFGQRSN
 **** *.:*****

S2ndP
 Ma-LRR-5PLs
 RMja-LRR-5PLs
 7G065400-LRR-5PLs-iso1
 7G065400-LRR-5PLs-iso2

DSSATISLPHNLNLDNSNWIGLAVCAYFSVLEHPTVDIDNLDIPAI SHHLICNLES DRDSL
 DSSATISLPHNLNLDNSNWIGLAVCAYFSVLEHPTVDIDNLDIPAI SHHLICNLES DRDSL
 DSSATISLPHNLNLDNSNWIGLAVCAYFSVLEHPTVDIDNLDIPAI SHHLICNLES DRDSL
 DSSATISLPHNLNLDNSNWIGLAVCAYFSVLEHPTVDIDNLDIPAI SHHLICNLES DRDSL

S2ndP
 Ma-LRR-5PLs
 RMja-LRR-5PLs
 7G065400-LRR-5PLs-iso1
 7G065400-LRR-5PLs-iso2

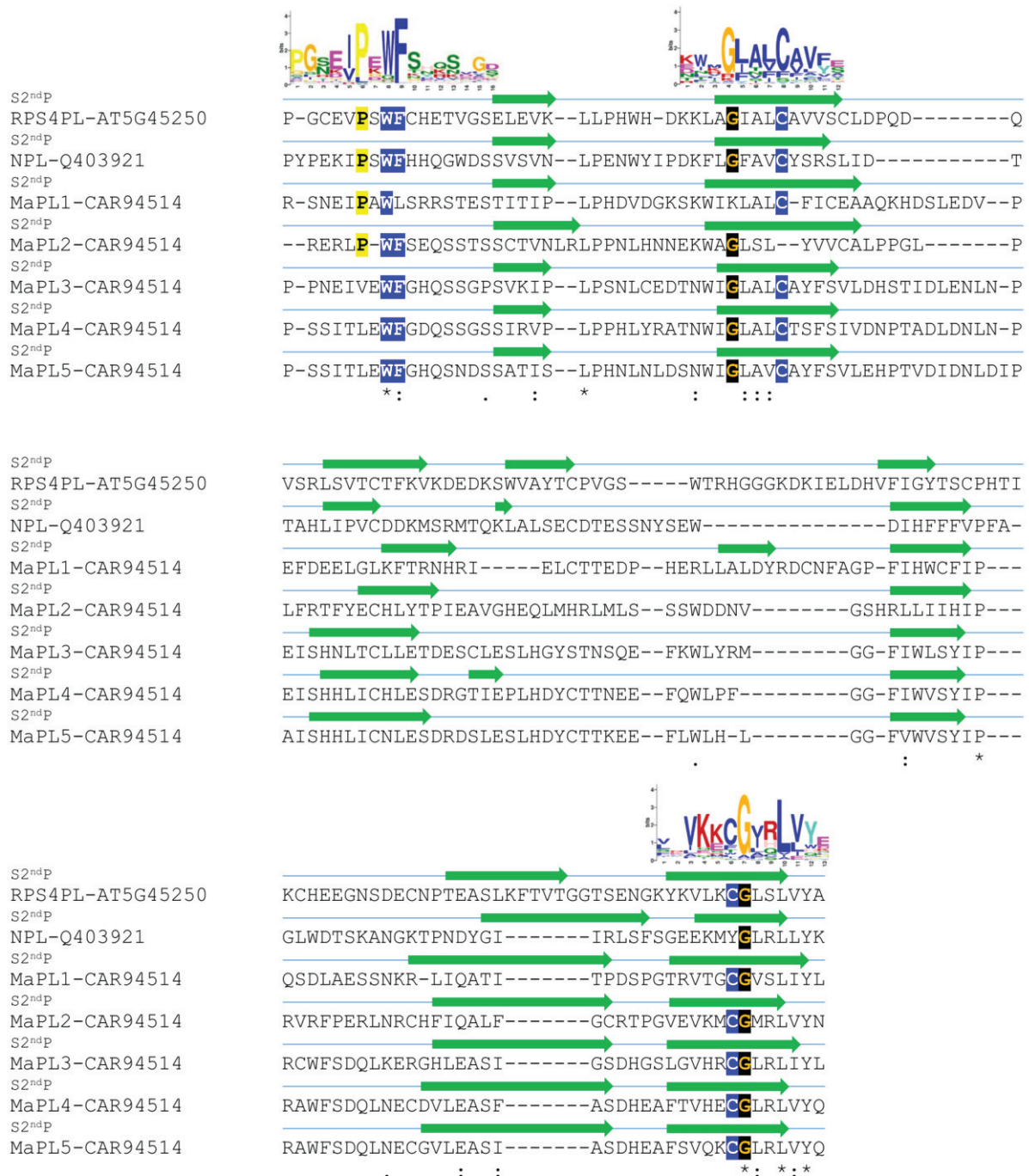
ESLHDYCTTKEEFLWLHLGGFVWVSYIPRAWFSDQLNECGVLEASIASDHEAFSVQKCGL
 ESLHDYCTTNEEFLWLHFGGFVWVSYIPRAWFSDQLNECGVLEASIASDHEAFSVQKCGL
 ESLHDYCTANEFLWLHFGGFVWVSYIPRAWFSDQLNECGVLEASIASDHEAFSVQKCGL
 ESLHDYCTANEFLWLHFGGFVWVSYIPRAWFSDQLNECGVLEASIASDHEAFSVQKCGL
 *****:*****

S2ndP
 Ma-LRR-5PLs
 RMja-LRR-5PLs
 7G065400-LRR-5PLs-iso1
 7G065400-LRR-5PLs-iso2

RLVYQHDEEEFKQTISRSQIRKKGISNTLTAR
 RLVYQHDEEEFKQTISRSQIRKKGISNTLTAR
 RLVYQHDEEEFKQTISRSQIRKKGISNTITAR
 RLVYQHDEEEFKQTISRSQIRKKGISNTITAR
 *****:***

ANNEXE 6

Alignement des domaines PL de RPS4, N et Ma avec leur structure secondaire prédite (S2ndP). Les brins β sont représentés avec une flèche verte et les résidus conservés dans les motifs 1, 2 et 3 sont surlignés.



ANNEXE 7

Alignement de quelques signatures PL (motifs 1, 2 et 3) chez *Amborella trichopoda*, *Ginkgo biloba* (Gb) et cinq TNLs de référence (N, Bs4, RPS4, Gro1-4 et Ma) chez les Dicotylédones.

Motif 1 **Motif 2**

```

      10      20      30      40      50      60      70      80      90      100
motifs  .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Gb_20740  P-GSEIPEWFSHQSVGD-----KWMGLALCAVFE-----
Gb_06500  SGSSVSVCNLNKWIREV-PSLQELNIIY-GSSVPLWFRHQMHGQLDGKE---HARDLCLDVPVIRVKPNVKIEAVIFCFVSHD--QGWSDYEWDHGR
A.trichopoda00001.586  SGGVLRKVDLFLKWKIGM-PLLRRELSFY-ASEIPEWFDHRMKTKEFVK---DFPRLCMDRVSMIFDEGRYAGTDVGYM--
A.trichopoda00109.130  -GCSPHTLRGSQSLAKETFTSLDKLSIP-GSKVDFWFMFQTP-----SFTVPRLSDVGLHIREVLLCFYSYS---TYDTSP---
A.trichopoda00056.119  -GCSPHVWRGQNLAKEMFTSLVQLSVP-GSKVFNWFAKCTPRCSLIQDGE---MLQQVSCALERPSPHAGRRRIKEIMVCLVIY---VHSSYTR--
A.trichopoda00085.83  -GCKPHVLRGQSLLEEMFTGLYLSVGP-GSKVFNWFAKCTPRCSLIQDGE---MLQQVSCALERPSPHAGRRRIKEIMVCLVIY---VHSSYTR--
A.trichopoda00085.87  -GCSYVSNVSPGLGKETFIQLRHFGVS-GSKVPSWFMFQTL-----SCVIVRVSDEGLKVRGVILCLVLS---VDSEMPIDDG
A.trichopoda00085.87  -GCSPRVRS--IAKETFNHLWCFCIS-GSKVDFWFMFQKL-----SCTVPRLSDDQLKIRGVILCFVIS---HNSIPIDFV
A.trichopoda00085.80  -GCT-HPSRS--IAKETFEHLRCFCVGS-GSNVDFWFMFQSL-----SYMISGMSDGHGVEIRGVILCLVLS---LDNEIPTIED
N(N.g.)Q40392.1  -QNISSMRHDISASDLSLTVETGPPYPEKIPSWFHHQGW-----DSSVSVNLPENWYIPDKFLGFAVCYS---RSLIDTT
Bs4(S.l.)AAR21295.1  -FLNISSFQHNISASDLSLTVETSL-GSSIPWFHHQGW-----TDSVSVNLPENWYIPDKFLGFAVCYS---RSLIDTT
RPS4(A.t.)CAB50708.1  -QLLSYARKRYNGGLVSESLFSTCFP-GCEVPSWFCHET-----VGSELEVKLLPHWHDK-KLAGIALCAVVS---CLDPQDOV
Gro14(S.t.)AAP44390.1  -TSMVDSLKQMLEALYMNVRFLVVP-GMEIPEWFTYKSW-----GTQSMSVALPTNWFPT-TERGFTVCVILDKMLFILGRFNTHKV
Ma(P.c.)CAR94514.1  ----DAIRR-----DERFEYGYR-SNEIPAWLSRRS-----TESTITIPLPHVDGKSKWKILALCFI-----CE

```

110 120 130 140 150 160 170 180 190 200

```

motifs  .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Gb_20740  YNSNTWMEATIIRGE--EEVHNTRILR---NRHADRRDQLHLSVYREG---HPFVMN--LQNGDRIRV-----WARS
Gb_06500  ----KATIVRGG--QEVFTEVIRI---REANRRDQLYLSMYLEN---DSFVKR--LHSGDRIRI-----RVCS
A.trichopoda00001.586  -GSPRSMVSIKRED--RGIREEQVDIQYFPLS--HQNVVFRRLRHGHMP-RLWFKW--FKDENQIEL---LTKF
A.trichopoda00109.130  -NRHVGLNIKRED--ETL--POIPLQIYPTVNLNQNIYFYFRREGQY--GNRKFW--LEDGDELEVEYDPIQCRCQAADVDEPTECEDQVEDIEEL
A.trichopoda00056.119  -QKHVNVLLIKRKG---REKRLPLGCIYPT--HRNQIYIYHFRDCIGF-EFQMNW--LKGDQIEF-----VSESTYEDQIEVINES
A.trichopoda00085.83  L-DIPHVQFKVMNNG--MQRMSPTLKLGN-PAK--NQDQVYLHFQED---DALVLN--LEDGDEVKI-----INNC
A.trichopoda00085.87  P-DIPDLVFMIRNG--TRRFSTMLRLGV-PAK--HQDQVYLHFQEG---GHILFM--LEGGDQIEV-----AKRS
A.trichopoda00085.80  DWGIPDVELETTRDG--IERFSSTLRLEMLPKA--RQDQVYLHFQEG---DDVLLM--LGGDQIEV-----LNRS
N(N.g.)Q40392.1  AHLIPVODDKMSRMT--QKLALSECDD---ESSNYSEWDIHFFVPEFA--GLWDTN--KANGKTPND-----YGII
Bs4(S.l.)AAR21295.1  NLTENTAEELIMSSAG---MPCITWKLKLLSNHSECTYIRIHFFVPEFA--GLWDTN--NANGKTPND-----YKHI
RPS4(A.t.)CAB50708.1  SRLSVTCTFKVKDED---KSWVAYTCPVGSWTRHGCKDKIELDHFVIGYTSCPHTIKC--HEEGNSDEC-----NPTBASLK
Gro14(S.t.)AAP44390.1  YLENMWLNLRKRYDGLRQIKISTSPGPIG--SEKPGGLGDTLITHIAFE--RSWK---LENLDYYR-----NNAF
Ma(P.c.)CAR94514.1  AAQKHDSLEDPVEFDEELGLKPTRNHRREL-CTTEDPHERLLALDYRDCNFA-GPFIHWCFFIPQSDLAES-----SNKR

```

Motif 3

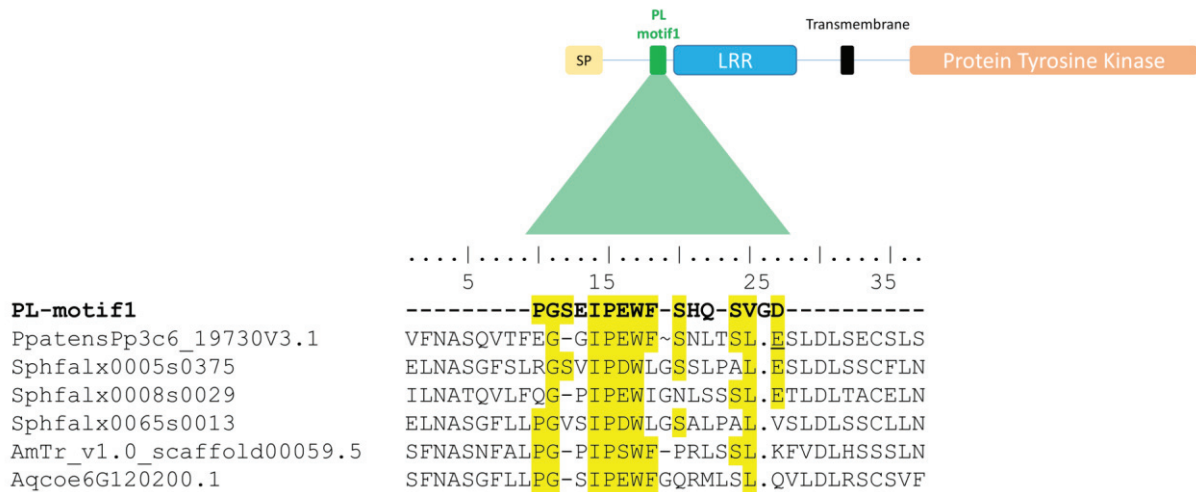
```

      210      220      230      240      250      260      270      280      290      300
motifs  .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Gb_20740  -----VKKCGVRLVYEQDVEELNQT-----
Gb_06500  IYA-----AWRIWVKEGGMHLLYRECEEKHEQDQLCTNIIYNHPSYTIHSNSAHKYAISL-----
A.trichopoda00001.586  QSI-----SHQICVKQGGMHLLYKTEETTSNNKLDNDTVFERLSKDLTWLHWNVWVDKHPVPIHQMDIKSFIAYDLQSCFIHQATLTKPSVMHASR
A.trichopoda00109.130  -----MVCL--KGMHLVYKPADTESIQERLANFFNLLSE-----
A.trichopoda00056.119  RAS---SLRFLPCVRLKVGIIHPIYSEEDTEDSTEELANFFNSL-----
A.trichopoda00085.83  GPSLTPSLGFPCCYHLKQAGIHLVYSEEDTEDSIEEKLADFFNSLHVQWQLTASLYVRYPFWFRIATPYFLCGIWIWK-----
A.trichopoda00085.87  PPL-----TPSIHKKVAIHLVYKPEDTKDSIQEKLAEFFNSMDAREDALEIVQFHSLQAHNLQTYHPERMRSGIILLVVLVLCFFYFWM-----
A.trichopoda00085.87  PPM-----FPSIQLKKGGIHLVYKPEQTDYIQEKLAEFNSMDCHEMLPPHYLRKRVQAGILLVWFVLLYFVWVRNSNENYSPFLPSLMTYSALYALK
A.trichopoda00085.80  PPE-----FPCVQMKRGGIHLVYKPEDTESIQEKLADFFNSLNCDDSS-----
N(N.g.)Q40392.1  RLS-----FPG--EEKMYGLRLLYKEGPEVNALL-QMRENSNEPTEHSTGIRRQYNNRTSFYELING-----
Bs4(S.l.)AAR21295.1  MLS-----FPG--ELKECGVRLVYEDESVLETTNDELTVGRRIRYDDDDSEHYEEAGCSSKKQRS-----
RPS4(A.t.)CAB50708.1  FVTVTGGTSENGKYVVKCGLSLVYAKDKDKNSALETKYDMLIKGSFQETSEGVDRVKKTKGKYVMPVEKNFQETTEGVDRVNNKKKTRMDNGRPKKKQ
Gro14(S.t.)AAP44390.1  QLEFSACDHYQKDVVKGLGVRLVYEN-----
Ma(P.c.)CAR94514.1  LIQATITPDSPTGTRVTCGGVSLIYLEDVPKFVRKLNKHSYCYHGYQPEEEDGMRSILSTSGIQTEQEEVQEQETTTSTRIVGQLRRNVESSLEKLFEE-

```


ANNEXE 8

Identification de signatures caractéristiques du motif 1 du domaine PL chez des protéines kinases des plantes archaïques *Physcomitrella patens*, *Sphagnum fallax*, *Amborella trichopoda* et *Aquilegia caerulea*.



The *Ma* resistance locus to root-knot nematodes in *Prunus*: Structural originality and evolution within the NBS-LRR gene family in plants

Root-knot nematodes (RKNs), *Meloidogyne* spp., are extremely polyphagous pests that severely challenge plants worldwide and especially perennials. The specific genetic resistance of plants mainly relies on NBS-LRR receptor genes (or NLRs grouping TNL, CNL and RNL subfamilies) that are pivotal factors for control of pests and pathogens. In *Prunus* spp., the *Ma* plum TNL gene confers resistance to all RKNs tested, whereas the *RMja* almond gene displays a more restricted spectrum of resistance (R). Moreover, the *Ma* predicted protein shows a peculiar TNL structure due to a C-terminal region made of five repeated domains, designated post-LRR domains (PLs). In this context, this thesis work has characterised the originality and the distribution of this uncommon structure among diverse plant proteomes and has revealed the genetic relationship between the *Ma* and *RMja* genes.

We first studied the frequency, distribution and structural characteristics of TNL genes and PL domains within the peach genome, the reference genome for Rosaceae. The finding of PL domains, which have been identified in two thirds of the 195 TNLs, allowed us to define specific motifs that improve the detection of this poorly known domain in Angiosperms. We found that the PL domain is specific of TNLs and is present in Angiosperm genomes in a proportion similar to the one established for peach. Besides, TNLs displaying multiple PL domains are rare in plants. The five-PL domain pattern is probably unique to *Ma* and its orthologues and was probably inherited from their common ancestor in the order Rosales.

We then investigated the NBS-LRR repertoire of the conifers (Gymnosperms), an ancient taxonomic group, for which the data related to this gene family are unclear. By analysing seven reference transcriptomes, we highlighted a large and diverse NBS-LRR arsenal in conifers but, surprisingly, no PL signatures have been detected. The examination of ancient plant proteomes revealed that only *Ginkgo biloba* displayed a few PL signatures. Our results suggest that a partial acquisition of the PL domain occurred early in seed plants and was followed by an adaptive expansion in Angiosperms. Additionally, we showed that conifers and Rosaceae have numerous RNLs and TNLs. By enlarging our study to other land plant genomes, we uncovered an average ratio of 1:10 between RNLs and TNLs numbers.

We finally carried out a high-resolution mapping of the *RMja* gene in almond. Using a BAC library, *RMja* was localised into the *Ma* resistance cluster and the *Ma* orthologue is by far the best candidate. The sequence comparison between three orthologous regions of the *Ma* locus, i.e. plum (complete R spectrum), almond (incomplete R spectrum) and peach (null R spectrum) highlighted a unique conserved structure of the *Ma* orthologues. Our results suggest that the polymorphism contained in the PL-domain repeats might underlie differential resistance interactions with RKNs and an original immune mechanism in woody perennials. In these immune processes for recognition or signalling, other components such as RNLs might be involved. This work paves the way for future comparative and functional approaches aiming to unravel the molecular determinants involved in the resistance to RKNs.

Keywords: Evolution, gene mapping, *Meloidogyne*, NBS-LRR, NLR, plant-nematode interaction, PL domain, *Prunus*, resistance genes, TNL