



HAL
open science

Diversité des bases génétiques de la résistance au virus de la panachure jaune du riz (RYMV) dans l'espèce de riz africain *Oryza glaberrima*

Hélène Pidon

► **To cite this version:**

Hélène Pidon. Diversité des bases génétiques de la résistance au virus de la panachure jaune du riz (RYMV) dans l'espèce de riz africain *Oryza glaberrima*. Amélioration des plantes. Université Montpellier, 2016. Français. NNT: 2016MONTT163 . tel-01686769

HAL Id: tel-01686769

<https://theses.hal.science/tel-01686769>

Submitted on 17 Jan 2018

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

Pour obtenir le grade de
Docteur

Délivré par l'Université de Montpellier

Préparée au sein de l'école doctorale **GAIA**
Et de l'unité de recherche **DiADE**

Spécialité : **BIDAP**
Biologie, Interactions, Diversité Adaptative des Plantes

Présentée par **Hélène Pidon**

**Diversité des bases génétiques
de la résistance au virus de la
panachure jaune du riz (RYMV)
dans l'espèce de riz africain
*Oryza glaberrima***

Soutenue le 01/12/2016 devant le jury composé de

Dr Valérie GEFROY	DR INRA Versailles	Rapportrice
Pr Laurent GENTZBITTEL	PR ENSAT	Rapporteur/Président
Dr Valérie SCHURDI-LEVRAUD	MC Univ. de Bordeaux	Examinatrice
Dr Benoît MOURY	DR INRA PACA	Examineur
Dr Alain GHESQUIÈRE	DR IRD	Directeur
Dr Laurence ALBAR	CR IRD	Co-encadrante



*The Road goes ever on and on
Down from the door where it began.
Now far ahead the Road has gone,
And I must follow, if I can,
Pursuing it with eager feet,
Until it joins some larger way
Where many paths and errands meet.
And whither then? I cannot say.*
J.R.R. Tolkien

Ce document est mis à disposition selon les termes de la licence [Creative Commons](#) “Attribution - Pas d’utilisation commerciale - Pas de modification 4.0 International”.



Remerciements

Il s'agit certainement de la partie la plus attendue. C'est la première chose que vous trouverez dans cette thèse mais c'est pourtant une des dernières à avoir été écrite. Occasion de faire le point sur trois ans de ma vie, et plus encore, beaucoup d'émotions se sont mêlées à son écriture. Elle aura donc été écrite avec un regard brouillé et aura nécessité trois paquets de mouchoirs et l'abus d'un grand nombre de répétitions de "merci" !

Tout d'abord, merci à mes rapporteurs, Valérie Geffroy et Laurent Gentzbittel, et à mes examinateurs, Valérie Schurdi-Levraud et Benoît Moury d'avoir accepté d'évaluer ma thèse.

Un immense merci à Laurence pour ton encadrement attentif. Tu as su me guider en douceur d'une position d'étudiante à celle de jeune chercheuse. Merci pour ta patience, ta rigueur, ton investissement et ta disponibilité. J'ai énormément appris auprès de toi et je ne t'en remercierais jamais assez.

Merci à Serge Hamon puis à Alain Ghesquière pour leur accueil dans l'unité DIADE. Alain, merci également d'avoir accepté d'être mon directeur de thèse et pour les discussions intéressantes que nous n'avons pu avoir malheureusement que trop rarement au vu de tes responsabilités dans l'unité.

Merci également aux membres de mon comité de thèse, Véronique Decroocq, Florence Vignols et Jacques David, de m'avoir aiguillé et prodigué vos conseils à des moments importants de ma thèse.

Et parce qu'une thèse ne commence pas par hasard, de nombreuses personnes m'ont guidée sur le chemin et ont donc leur place dans ces remerciements :

Mes premiers soutiens ont bien entendu été ma famille. Mes parents, qui m'ont toujours poussée à faire de mon mieux et à aller plus loin. Et qui ont accepté de jouer le

rôle de correcteurs orthographiques de dernière minute ! Mon frère, qui a toujours été là quand il le fallait. Mais aussi mes grand-parents, mes oncles et tantes et cousins-cousines. Même si je peux parfois avoir l'air réservée, votre présence et vos encouragements m'ont beaucoup aidée !

Ensuite, merci à "papa" Jacques David, qui a su conforter mon intérêt pour la génétique, m'éclairer dans mes périodes de doutes et m'a orientée vers Laurence. Et puisque j'en suis à SupAgro et à la spé APIMET, merci à "maman" Isabel, notre ingé pédagogique magique, qui nous a toujours supportés et nous supportera encore un moment !

Merci également à mes maîtres de stages précédents : Ingo Hein et Véronique Lefebvre qui m'ont montré le monde de la recherche, m'ont appris à l'aimer et auprès de qui j'ai beaucoup appris.

Merci à Anne-Marie Martinez, ma marraine du programme de mentorat Femmes & Sciences. Nos discussions m'ont permis d'y voir plus clair dans les passages obligés de la thèse et dans la façon d'appréhender l'après-thèse.

Merci à tous les membres de l'équipe GDR puis RICE : François et Christine pour les coups de mains en bioinfo et pour les données IRIGIN ! Sophie, Harold Élise et Pierre pour l'aide en serre. Andrès et Cécile, les co-bureaux pour les coups de main et de folie. Et merci à Pierre, Mathias, Hélène, Tom, Stéphane, Marie, Sophie et Nhung pour vos conseils et les discussions intéressantes tout au long de ma thèse.

Un grand merci à Mikaël, mon stagiaire, pour l'aide qu'il m'a apportée dans le phénotypage de la collection. Sans toi tout cela aurait été bien fastidieux (enfin, encore plus !). Merci pour ton enthousiasme et ta bonne humeur. J'aurais difficilement pu rêver d'un meilleur stagiaire que toi pour me froter pour la première fois à l'encadrement !

Merci à Eugénie pour les explications et discussions sur ce qui se passe côté virus. Merci à Denis pour les discussions sur l'origine évolutive des résistances et du virus. Dans la team virus, merci également à Agnès pour ta gentillesse et tes conseils.

Merci à Anthony Champion pour les transformations *RYMV3* (encore en cours au moment où j'écris ces lignes). Merci à Yves Vigouroux et Anne-Céline Thuillet pour les discussions et conseils de génétique des populations. Merci à François Anthony pour les éclairages sur la structure des NB-LRR. Merci à Ndomassi et Alexis pour les coups de mains bioinfo. Merci à Emira et Maryline de m'avoir aidée à aborder mon GWAS. Et merci à tous les IRDiens pour vos sourires, discussions au coin café et autres indulgences devant les bêtises des doctorants.

Et merci à la bande des thésards (et plus si affinité) pour tous les moments partagés, bien plus souvent bons que mauvais ! Merci pour les escapades, les soirées diverses et variées, les pauses partagées, le soutien dans les moments difficiles et pour nos délires fréquents !

Merci à Lucie pour ta trompette et ta folie.

Merci à Céline pour ton esprit aventureux, tes conseils toujours pertinents et tes gnoux.
Merci à Mathilde number one pour les soirées Couleur de Bières et pour m'avoir montré qu'on peut tout donner sur la fin de la rédaction.
Merci à Kim et à Adama pour les moments partagés lors de vos passages à Montpellier.
Merci à Sixtine pour notre amour partagé du chocolat.
Merci à Fabien pour le fil à linge et les discussions webseries.
Merci à Rémi pour le partage de ta passion pour la recherche depuis l'agro, et de pleins d'autres passions plus ou moins bizarres aussi !
Merci à Chloé, et à ton indissociable Lionel, pour ton amour des palmiers, tes soirées endiablées, ta redondance et tes belettes louches.
Merci à Jérémy, mon camarade geek, pour nos bavardages Skype en APIMET, ton suivi attentif des infos agros et surtout ton amitié et ton soutien indéfectible (c'est pas faux ?).
Merci à Mathieu, notre mouton frisé, pour les pourrissages de bureau, le kloug et le partage de tes voyages.
Merci à Maryline, même si tu n'es plus thésarde, pour ta passion du café et ton autodérision.
Merci à Hélène pour notre amitié en ligne puis en vrai et pour les corrections.
Merci à Mathildeux pour nos discussions chats, pour tes tentatives d'attaques de câlins, ta gentillesse et les corrections aussi.
Merci à Lucile, notre toctoc, pour ton rire omniprésent, tes tics rigolos et tes toctocqueries.
Merci à Emilie la chauve-souris pour tes idées foisonnantes de bêtises à faire, les ninjas, les barbecues dans ta savane, ton optimisme et plein d'autres choses encore.
Vous allez tous me manquer !

On me dit dans l'oreillette que j'oublie quelqu'un. Mais non, c'est juste que je ne peux pas me contenter d'une phrase pour ma co-bureau ! Merci Cécile ! Merci pour ton accueil dans l'équipe, pour les moments de délire partagés avec la tomate et Gaston le dragon. Merci pour le soutien et l'amitié toujours présents au cours de ces trois ans. Merci de m'avoir supportée au jour le jour (et ça peut être difficile !) pendant qu'on traversait ensemble les différentes étapes de la thèse. Merci d'avoir su me faire parler quand j'en avais besoin. Sans toi ces trois ans aurait été bien plus ternes !

Merci à mes plus vieux amis : Mélissa, ça fait maintenant quinze ans qu'on se supporte et s'encourage, j'espère que ça durera encore longtemps ! Merci à Lauriane, qui partage l'expérience de la thèse. J'espère que des jours heureux t'attendent nombreux ! Merci à Robin, avec qui c'est toujours comme si on s'était quittés hier. Garde ta bonne humeur et ta joie de vivre !

Merci à Bottine, ma chatte, qui m'a tenu compagnie depuis le tout premier week-end de ma thèse. Merci au chocolat, qui m'a donné du courage. Je ne sais pas où en seraient

les thésards sans toi. Merci à Chameyer, qui m'a permis d'avancer la rédaction tout en ayant un léger parfum de vacances.

Enfin, merci à toi lecteur ! (C'est potentiellement redondant d'ailleurs).

Et, comme on oublie forcément des gens que l'on aurait dû remercier, merci à toi, (écris ton nom ici), l'oublié à tort.

Sommaire

Remerciements	iii
Liste des figures	xi
Liste des tableaux	xiv
Introduction générale	1
Chapitre 1 Revue bibliographique	5
1 Résistance des plantes aux virus	7
1.1 Résistance non-spécifique	7
1.2 Les gènes majeurs	10
1.2.1 Les NB-LRR	11
1.2.2 Les facteurs d’initiation de la traduction	15
1.2.3 Les autres types de gènes majeurs	17
1.3 Résistance quantitative	18
1.4 Résistance aux vecteurs	19
1.5 Contournement et durabilité des résistances	20
1.6 Modèles évolutifs des gènes de résistance	22
2 Identification des locus de résistance	23
2.1 Cartographie génétique par analyse de liaison	23
2.1.1 Cartographie d’un gène majeur	23
2.1.2 Cartographie de caractères quantitatifs	24
2.2 Génétique d’association	25
2.2.1 Le principe	25
2.2.2 Le déséquilibre de liaison	25
2.2.3 Les études de génétique d’association sur génome entier	29
2.3 Utilisation de populations multi-parentales	30
2.4 Autres méthodes	32
3 Le pathosystème	32
3.1 Le riz	32
3.1.1 Le genre <i>Oryza</i>	32
3.1.2 Les espèces cultivées	34
3.1.3 Importance économique du riz en Afrique	38

3.1.4	Contraintes à la production en Afrique	39
3.2	La panachure jaune du riz	40
3.2.1	La maladie	40
3.2.2	Le <i>Rice yellow mottle virus</i>	43
3.3	Les résistances du riz au RYMV	46
3.3.1	<i>RYMV1</i>	46
3.3.2	<i>RYMV2</i>	48
3.3.3	QTL de résistance chez <i>O. sativa</i>	49
3.3.4	Le contournement de la résistance	50
Chapitre 2 Matériels et méthodes généraux		53
1	Matériel végétal	55
2	Matériel viral	57
3	Évaluation de la résistance au RYMV	58
3.1	Conditions d'inoculation	58
3.2	Observation des symptômes	59
3.3	Évaluation du contenu en virus	59
4	Ressources génomiques	60
5	Génotypage	61
Chapitre 3 Cartographie d'un gène majeur de résistance		65
1	Introduction	67
2	Fine mapping of <i>RYMV3</i> : a new resistance gene to <i>Rice yellow mottle virus</i> from <i>Oryza glaberrima</i>	68
2.1	Introduction	69
2.2	Materials and Methods	70
2.2.1	Plant material	70
2.2.2	Viral material	70
2.2.3	Phenotypic evaluation	71
2.2.4	Genomic resources	72
2.2.5	Development of molecular markers and PCR procedure	72
2.3	Results	73
2.3.1	Tog5307 resistance spectrum	73
2.3.2	Inheritance of Tog5307 resistance	74
2.3.3	Identification of a candidate locus for Tog5307 resistance on chromosome 11	75
2.3.4	Fine mapping of <i>RYMV3</i>	76
2.3.5	<i>RYMV3</i> candidate genes	76
2.3.6	Tog5672 second resistance gene is linked to <i>RYMV3</i>	78
2.4	Discussion	79
3	Conclusion générale et perspectives	83
3.1	<i>RYMV3</i> confère une résistance monogénique dominante	83

3.2	<i>RYMV3</i> est probablement un gène de type CC-NBS-LRR . . .	84
3.3	Contournement ou pénétrance incomplète de la résistance? . . .	84
Chapitre 4 Analyse de la résistance au RYMV dans l'espèce <i>O. glaberrima</i> et son apparenté sauvage <i>O. barthii</i>		87
1	Introduction	89
2	Résultats	90
2.1	Identification des individus résistants	90
2.2	Recherche d'allèles aux gènes de résistance connus	90
2.2.1	Allèles à <i>RYMV1</i>	91
2.2.2	Allèles à <i>RYMV2</i>	92
2.2.3	Polymorphismes à <i>RYMV3</i>	94
2.2.4	Conclusion sur les allèles de résistance identifiés	97
2.3	Répartition des résistances dans la diversité des riz africains	97
2.3.1	Diversité génétique	97
2.3.2	Répartition géographique	99
3	Discussion	100
3.1	Fréquence des sources de résistance élevée chez les riz africains	100
3.2	Gènes et allèles de résistance identifiés	102
3.2.1	<i>RYMV1</i>	102
3.2.2	<i>RYMV2</i>	103
3.2.3	<i>RYMV3</i>	105
3.2.4	Conclusion sur les sources de résistance de la collection IRD	105
3.3	Diversité des gènes et évolution	106
3.3.1	Profils de diversité aux gènes de résistance	106
3.3.2	Structuration de la diversité des allèles de résistance au RYMV	108
Chapitre 5 Génétique d'association dans l'espèce <i>O. glaberrima</i>		111
1	Introduction	113
2	Matériel et méthodes spécifiques	113
2.1	Obtention des données de phénotype	113
2.1.1	Matériel végétal	113
2.1.2	Matériel viral	114
2.1.3	Dispositif expérimental	114
2.1.4	Mise au point et réalisation de la mesure de l'accumulation virale	116
2.1.5	Traitement des données de phénotypage	116
2.2	Étude de génétique d'association	118
2.2.1	Traitement des données de séquences et analyse de la structure de population	118

2.2.2	Réalisation de l'étude de génétique d'association	119
3	Résultats	119
3.1	Évaluation de la résistance partielle	119
3.1.1	Choix des dates et des dilutions	119
3.1.2	Phénotypes obtenus	120
3.2	Étude d'association	124
3.2.1	Structure de la population	124
3.2.2	Choix du modèle	126
3.2.3	Détection des QTL	127
3.2.4	Étude préliminaire des QTL identifiés	128
4	Discussion et perspectives	130
4.1	Phénotypage de la résistance partielle	130
4.2	Génétique d'association	132
4.3	Conclusions et perspectives	135
Chapitre 6 Discussion générale et perspectives		137
1	Une grande diversité de résistances au RYMV	139
1.1	De la diversité des déterminants de résistance	139
1.2	De la diversité des mécanismes de résistance	141
1.3	Quelles raisons à la diversité de résistance chez <i>O. glaberrima</i> ?	144
2	Vers une résistance durable au RYMV au champ?	145
2.1	Exploitation des résistances de l'espèce <i>O. glaberrima</i>	145
2.2	Exploitation des QTL de résistance	146
2.3	Le contournement des résistances : quelle stratégie employer pour une résistance durable?	147
3	Conclusion générale	149
Références bibliographiques		151
Annexe A Tableaux annexes		I
Annexe B Figures annexes		XIII

Liste des figures

1.1	Mécanismes de <i>silencing</i> des ARN	8
1.2	Analogies au modèle de l'immunité virale du point de vue de l'infection virale	10
1.3	Comparaison des modèles de garde et <i>decoy</i>	12
1.4	Modèle " <i>bait and switch</i> " du fonctionnement des NB-LRR	14
1.5	Diagramme simplifié de l'initiation de la traduction	15
1.6	Étapes du contournement de la résistance des plantes par les virus . . .	20
1.7	Comparaison schématique de l'analyse de liaison réalisée avec une descendance de cartographie et d'une cartographie d'association réalisée avec des collections d'individus non apparentée	24
1.8	Shématisation de la relation triangulaire à la base de la génétique d'association	25
1.9	Relation entre l'étendue du DL et la résolution des études d'association	27
1.10	Différence de comportement des statistiques D' et r^2 de mesure du DL selon les scénarios évolutifs	28
1.11	Comparaison des $Q-Q$ <i>plot</i> obtenus pour deux modèles de GWAS exécutés sur les mêmes données	30
1.12	Exemple de populations multi-parentales	31
1.13	Classification phylogénétique des <i>Poaceae</i>	33
1.14	Distribution mondiale des espèces du génome AA du genre <i>Oryza</i> . . .	35
1.15	Évolution de la production de riz en Afrique.	39
1.16	Distribution de la panachure jaune du riz en Afrique	41
1.17	Évolution des symptômes de la panachure jaune du riz au champ . . .	42
1.18	Organisation génomique des <i>Sobemovirus</i>	43
1.19	Phylogéographie du RYMV	45
1.20	Modélisation de structure du domaine central du facteur eIF(iso)4G1 du riz, et localisation des différentes mutations de résistance contre le RYMV.	47
1.21	Fonctionnement de CPR5 chez <i>A. thaliana</i>	48
1.22	Position des QTL et gènes majeurs impliqués dans la résistance au RYMV	50
2.1	Distribution géographique des accessions de riz africains utilisées dans cette étude	56
3.1	Disease incidence on Tog5307 accessions four weeks after inoculation . .	74

3.2	Genomic regions in linkage disequilibrium with resistance in a population of 18 resistant BC1F1 plants from the cross (IR64 x Tog5307) x IR64	76
3.3	Physical map of the <i>RYMV3</i> locus	77
3.4	Collinearity between the CG14, IR64 and Nipponbare genomes in the genomic region containing <i>RYMV3</i>	77
3.5	Annotation of the <i>RYMV3</i> candidate region and polymorphisms detected between CG14 and Tog5307	78
4.1	Positions des polymorphismes non-synonymes dans la protéine RYMV1 présents dans la collection IRIGIN	91
4.2	Position des mutations non-synonymes sur le modèle protéique de RYMV2.	93
4.3	Position des mutations codantes dans le NB-LRR sur le modèle protéique de RYMV3	96
4.4	Accessions <i>O. glaberrima</i> résistantes dans la diversité de l'espèce	98
4.5	Accessions <i>O. barthii</i> résistantes dans la diversité de l'espèce	99
4.6	Répartition géographique des accessions résistantes	100
4.7	Diagramme des Fst entre les accessions porteuses d'un allèle de résistance à <i>RYMV1</i> et à <i>RYMV2</i>	109
5.1	Plan de l'expérience de phénotypage	115
5.2	Représentation de la pseudo-AUDPC	117
5.3	Exemple de choix de dilution pour le test ELISA	120
5.4	Accumulation virale mesurée dans certaines accessions du test de phénotypage aux trois dates et six répétitions	121
5.5	Boxplot de la concentration virale aux trois dates dans les six répétitions en fonction des accessions	121
5.6	Projection des répétitions selon les composantes 1 et 2 des ACP aux trois dates	122
5.7	Histogrammes de la distribution des moyennes de concentration virales des accessions aux trois dates	123
5.8	Régression linéaire des valeurs de concentrations virales entre les trois dates	123
5.9	Graphique de l'entropie-croisée calculé pour la structuration de la population en un à 25 groupes	124
5.10	Assignations des accessions <i>O. glaberrima</i> à deux, trois ou quatre groupes de structure	125
5.11	<i>Q-Q plots</i> des GWAS réalisé sur les données de mesure de l'accumulation virale à 17 jai	126
5.12	<i>Manhattan plot</i> du GWAS réalisé sur les données de mesure de l'accumulation virale à 17 jai	127
5.13	Effet des SNP significativement associés à la résistance sur le phénotype d'accumulation virale à 17 et 20 jai	128

5.14 DL et <i>Manhattan plot</i> autour des QTL identifiés	129
6.1 Mécanisme probable de la résistance conférée par <i>RYMV1</i>	142
6.2 Mécanisme probable de la résistance conférée par <i>RYMV3</i>	142
6.3 Hypothèses de mécanismes de la résistance conférée par <i>RYMV2</i>	142
B.1 Accumulation virale mesurée dans les accessions du test de phénotypage aux trois dates et six répétitions	XV
B.2 <i>Manhattan plots</i> et <i>Q-Q plots</i> obtenus pour les quatre variables de résistance partielle avec le modèle CMLM sans structure de population	XVI
B.3 <i>Manhattan plots</i> et <i>Q-Q plots</i> obtenus pour les quatre variables de résistance partielle calculés sans la répétition 4 avec le modèle CMLM sans structure de population	XVII

Liste des tableaux

2.1	Correspondance entre les numéros de la collection IRD et ceux de la collection IITA des accessions fréquemment utilisées.	57
2.2	Couples d’amorces utilisés pour le génotypage aux gènes de résistances.	63
3.1	Characteristics of the RYMV isolates used in this study	71
3.2	Characteristics of the markers developed for mapping and segregation analysis	73
3.3	Segregation of resistance in five BC3F3 families derived from the (IR64 x Tog5307) cross and subsequent backcrosses on IR64	75
4.1	Mutations rares non-synonymes dans le gène <i>RYMV2</i>	93
5.1	Corrélation entre les mesures de phénotype et le génotype aux marqueurs associés dans les différents modèles	128
A.1	Accessions de l’espèce <i>O. glaberrima</i> de la collection IRD.	III
A.2	Accessions de l’espèce <i>O. barthii</i> de la collection IRD.	VIII
A.3	Gènes annotés au QTL ₆	X
A.4	Gènes annotés au QTL ₁₁	X

Abréviations

- ACP : analyse en composantes principales
ADN : acide désoxyribonucléique
AFLP : *amplified fragment length polymorphism*
AGO : Argonautes
ARN : acide ribonucléique
ARNdb : ARN double brin
ARNm : ARN messenger
ARNsb : ARN simple brin
ATP : adénosine triphosphate
AUDPC : aire sous la courbe de progression de la maladie *Area under the disease progression curve*
Avr : protéine ou gène d'avirulence
CAPS : *Cleaved Amplified Polymorphic Sequence*
CC : *Coiled-Coil*
CITE : séquences améliorant la transduction en *cis* (*cis-acting translation enhancing sequences*)
CMLM : modèle mixte linéaire compressé
CP : protéine de capsid
CPR5 : *constitutive expressor of pathogenesis-related genes 5*
DAS-ELISA : *double antibody sandwich enzyme-linked immunosorbent assay*
dCAPS : *Derived Cleaved Amplified Polymorphic Sequence*
DCL : Dicer-like
DHL : lignée d'haploïdes-doublés (*doubled-haploid line*)
DL : déséquilibre de liaison
dNTP : désoxyribonucléotide
DO : densité optique

ABREVIATIONS

- eIF : facteur d'initiation eucaryote (*eukaryotic initiation factor*)
- ELISA : dosage d'immunoabsorption par enzyme liée (*enzyme-linked immunosorbent assay*)
- ER : résistance extrême
- ETI : Immunité déclenchée par effecteur (*effector-triggered immunity*)
- FDR : *false discovery rate*
- Fst : indice de fixation
- GBS : génotypage par séquençage (*genotyping by sequencing*)
- GLM : modèle linéaire global (*global linear model*)
- GWAS : étude d'association sur génome entier (*genome wide association studies*)
- HR : réponse hypersensible
- HSP : *heat-shock protein*
- INDEL : insertion-délétion
- IRES : sites internes d'entrée de ribosome (*internal ribosome entry site*)
- IRIGIN : *International RICE Genomic INitiative*
- JA : acide jasmonique
- jai : jours après inoculation
- LRR : *leucine rich repeat*
- MAGIC : *multi-parent advanced generation inter-cross*
- MAPK : *mitogen-activated protein kinase*
- MATH : *meprin and TRAF homology*
- MIF4G : domaine central du facteur d'initiation de la traduction 4G (*middle domain of eukaryotic initiation factor 4G*)
- MLM : modèle linéaire mixte (*mixt linear model*)
- NAM : *nested association mapping*
- NB-ARC : *nucleotide-binding adaptor shared by Apaf1, certain R genes and CED4*
- NERICA : *New rice for Africa*
- NGS : *next-generation sequencing*
- ORF : cadres de lecture ouverts (*open reading frames*)
- PABP : *poly-A binding protein*
- PAMP : *pathogen-associated molecular pattern*
- PAZ : Piwi/Argonaute/Zwille
- pAUDPC : pseudo-AUDPC
- PCR : réaction en chaîne par polymérase (*polymerase chain reaction*)
- PDI : *protein disulfide isomerase*
- PR : *pathogenesis related*
- PRR : *pattern recognition receptor*
- PTGS : *silencing de gène post-transcriptionnel (post-transcriptional gene silencing)*
- PTI : *PAMP-triggered immunity*
- QRL : *quantitative resistance loci*
- QTL : *quantitative trait loci*

- R : gène ou protéine de résistance
 RdDM : méthylation ADN dirigé par l'ARN (*RNA-directed DNA methylation*)
 RDRp : polymérase ARN dépendante de l'ARN (*RNA-dependent RNA polymerase*)
 RFLP : *restriction fragment length polymorphism*
 RIL : *recombinant-inbred line*
 RISC : complexe de *silencing* induit par l'ARN (*RNA-induced silencing complex*)
 RITS : complexe de *silencing* transcriptionnel induit par l'ARN (*RNA-induced transcriptional silencing complex*)
 RNAi : interférence ARN
 RNase : endoribonucléase
 ROS : espèces réactives de l'oxygène
 RTM : *restricted TEV movement*
 SA : acide salicylique
 SAR : résistance systémique acquise
 siRNA : petits ARNs interférant
 SNP : *single nucleotide polymorphism*
 TALEN : *transcription activator-like effector nucleases*
 TGS : *silencing* de gène transcriptionnel (*transcriptional gene silencing*)
 TILLING : *targeting induced local lesions in genomes*
 TIR : *Toll and Interleukin-1 Receptor*
 VPg : protéine liée au génome viral (*viral genome-linked protein*)
 VSR : supresseurs viraux d'interférence ARN

Virus

- BMV : *Brome mosaic virus*
 LMV : *Lettuce mosaic potyvirus*
 PIAMV : *Plantago asiatica mosaic virus*
 PPV : *Plum pox virus*
 PVX : *Potato virus X*
 RSV : *Rice stripe virus*
 RTSV : *Rice tungro spherical virus*
 RYMV : *Rice yellow mottle virus*
 SeMV : *Sesbania mosaic virus*
 TBSV : *Tomato bushy stunt virus*
 TEV : *Tobacco etch potyvirus*
 TMV : *Tobacco mosaic virus*
 ToMV : *Tomato mosaic virus*
 TYLCV : *Tomato yellow leaf curl begomovirus*
 WMV : *Watermelon mosaic potyvirus*

Introduction générale

Le riz est aujourd'hui l'une des principales sources alimentaires en Afrique. Sa production, principalement localisée en Afrique de l'Ouest, a augmenté de manière importante depuis 50 ans grâce aux avancées technologiques et à la mise en place de variétés plus productives. Toutefois, elle reste insuffisante pour combler les besoins grandissants de la population. L'augmentation nécessaire de la productivité est notamment freinée par la faible capacité d'irrigation, des pertes post-récolte importantes et des contraintes biotiques et abiotiques fortes s'exerçant sur les cultures.

Parmi les stress biotiques, la panachure jaune du riz, causée par le *Rice yellow mottle virus* (RYMV), est responsable de pertes sévères. Cette maladie, apparue suite à l'intensification de la culture de riz dans les années 1970, est maintenant une des principales maladies du riz en Afrique sub-saharienne et dans la plupart des pays producteurs du continent. Les moyens de lutte sont peu nombreux et, bien qu'efficaces, les mesures prophylactiques restent insuffisantes. L'utilisation de résistances génétiques apparaît comme la meilleure stratégie pour réduire les pertes associées au RYMV.

Deux espèces de riz cultivé existent : *Oryza sativa*, espèce domestiquée en Asie et cultivée dans le monde entier, et *Oryza glaberrima*, domestiquée en Afrique de l'Ouest et cultivée uniquement localement. Aujourd'hui, la culture d'*O. sativa* a supplanté celle d'*O. glaberrima*, notamment grâce à ses rendements bien supérieurs. Depuis plusieurs années, la recherche et la caractérisation de résistance au RYMV a fait l'objet d'études chez les deux espèces. Les variétés de la sous-espèce *O. sativa* spp *japonica* présentent fréquemment une résistance partielle au RYMV mais les sources de résistance élevée sont très rares dans l'espèce asiatique. Au contraire, malgré la faible diversité génétique de l'espèce *O. glaberrima*, de nombreuses accessions résistantes au RYMV y ont été identifiées et cette espèce semble posséder une richesse plus importante de déterminismes de résistance élevée.

Au démarrage de cette étude, deux gènes de résistance au RYMV avaient été identifiés : *RYMV1*, codant pour un facteur d'initiation de la traduction, et *RYMV2*, codant pour une protéine probablement impliquée dans la régulation des mécanismes de défense. Un allèle de résistance à *RYMV1* est présent dans deux variétés de l'espèce *O. sativa* alors que trois autres sont portés par des accessions de l'espèce *O. glaberrima*. Le seul allèle de résistance connu à *RYMV2*, correspondant à un allèle nul, n'a été identifié que dans l'espèce *O. glaberrima*. Toutefois, des isolats de RYMV contournant ces gènes et allèles de résistance ont été observés en conditions de laboratoire et la recherche d'autres déterminants de résistance est un enjeu critique afin d'obtenir une résistance durable.

Dans ce contexte, l'étude de la diversité de résistance de l'espèce *O. glaberrima* doit être poursuivie.

Les ressources génétiques et génomiques disponibles pour étudier cette espèce sont de plus en plus nombreuses. Un génome de référence, basé sur la variété CG14, a notamment été publié récemment (Wang et al., 2014c). De plus, l'IRD a réuni une collection de riz africains composée de 278 accessions d'*O. glaberrima* et de 103 accessions de son apparenté sauvage, *O. barthii*. Les données de séquences Illumina d'une partie de cette collection composée de 163 accessions d'*O. glaberrima* et 84 d'*O. barthii* ont également été acquises récemment dans le cadre d'un projet de séquençage coordonné par l'IRD.

L'objectif de cette thèse était donc d'utiliser ces ressources afin de poursuivre l'étude de la résistance au RYMV dans l'espèce *O. glaberrima* et d'en identifier les facteurs génétiques en utilisant plusieurs approches complémentaires.

Le premier chapitre de ce manuscrit est consacré à la description du pathosystème et des résistances des plantes aux agents pathogènes en général.

Un second chapitre, dédié aux matériels et méthodes employés, décrira les collections étudiées dans ce travail, les techniques d'évaluation de la résistance, ainsi que le génotypage et les ressources génomiques utilisées.

Le troisième chapitre présente la première approche mise en œuvre pour identifier un nouveau déterminant de résistance. Il s'agit d'une étude de cartographie fine d'un gène majeur, indépendant des gènes identifiés jusqu'alors. Cette étude a fait l'objet d'un article soumis dans la revue *Theoretical and applied genetics*.

Le chapitre suivant est consacré à l'étude de la diversité génétique aux trois gènes de résistance connus dans l'espèce *O. glaberrima* et chez son apparenté sauvage *O. barthii*. Nous y avons exploité les données de séquence afin d'identifier de nouveaux allèles. Cette étude nous a également permis d'identifier des schémas de structuration phylogénétique et géographique de la résistance au RYMV.

Le cinquième chapitre présente une étude de génétique d'association de la résistance partielle au RYMV dans l'espèce *O. glaberrima* réalisée sur 125 accessions non-hautement résistante. Cette étude a permis de compléter le panorama de la résistance au RYMV dans l'espèce *O. glaberrima* par les composantes génétiques de la résistance partielle.

Enfin, les résultats obtenus par les différentes approches seront discutés et mis en perspective dans le cadre de la compréhension des mécanismes d'interactions entre le riz et le RYMV et de l'optimisation du déploiement des résistances pour favoriser leur durabilité.

CHAPITRE

1

Revue bibliographique

Chapitre 1 : Revue bibliographique

1 Résistance des plantes aux virus

Le modèle actuel du système immunitaire des plantes définit deux strates de défenses contre les agents pathogènes microbiens (Jones and Dangl, 2006). Dans un premier temps, des récepteurs de la membrane plasmique, les *pattern recognition receptor* (PRR), reconnaissent des molécules conservées de l'agent pathogène, telles que la chitine ou la flagelline, appelées *pathogen-associated molecular pattern* (PAMP) et initient une réponse immunitaire appelée *PAMP-triggered immunity* (PTI). La plante répond notamment par la production d'éthylène, un choc oxydatif, un dépôt de callose ou l'induction de l'expression de gènes liés à la défense (Ingle et al., 2006). Ce mécanisme de défense est non-spécifique. L'agent pathogène peut contourner ces défenses lorsque les gènes codant les PAMP sont mutés ou par la sécrétion d'effecteurs, de petites molécules apoplastiques ou intracellulaires, qui suppriment ou modifient les défenses de la plante et permettent l'infection. La seconde state de défense, spécifique, met en jeu la reconnaissance directe ou indirecte de ces effecteurs par des protéines de résistances (protéines R) qui induit l'immunité déclenchée par effecteur (ETI : *effector-triggered immunity*).

Depuis peu, il a été montré que la sensibilité des plantes à un agent pathogène est rendue possible par l'utilisation de gènes de plantes favorisant l'attraction et l'attachement du pathogène, ou permettant le développement et la dispersion de l'agent pathogène (Lapin and Van den Ackerveken, 2013). Ce type de gène est appelé gène de sensibilité. On peut notamment citer l'exemple du gène OsSWEET14 du riz : l'activation du gène par un effecteur TAL (*Transcriptional activator like*) de *Xanthomonas oryzae* pv. *oryzae* est nécessaire à l'infection de la plante. La présence d'une mutation dans la séquence du promoteur reconnue par le TAL confère la résistance à la plante (Hutin et al., 2015).

La définition actuelle des PAMP et des effecteurs exclut les virus de ce modèle (Mandadi and Scholthof, 2013). Toutefois les mécanismes de défense contre les virus présentent de nombreuses analogies avec ceux qu'il décrit (Nicaise, 2014). On peut notamment distinguer les deux states de défense : non-spécifique et spécifique.

1.1 Résistance non-spécifique

La réponse normale d'une plante face à un agent pathogène, quelle que soit sa nature, est la résistance. Les plantes pour lesquelles cette résistance est effective contre un agent pathogène sont dites "non-hôtes" pour celui-ci.

Les premières barrières auxquelles fait face l'agent pathogène sont passives. Elles ne nécessitent pas la perception de l'agent pathogène. Ce sont des barrières structurales et biochimiques (Mysore and Ryu, 2004). L'épaisseur de la cuticule, la qualité des cires et de la cutine qui la composent, la structure des cellules épidermiques, l'épaisseur de

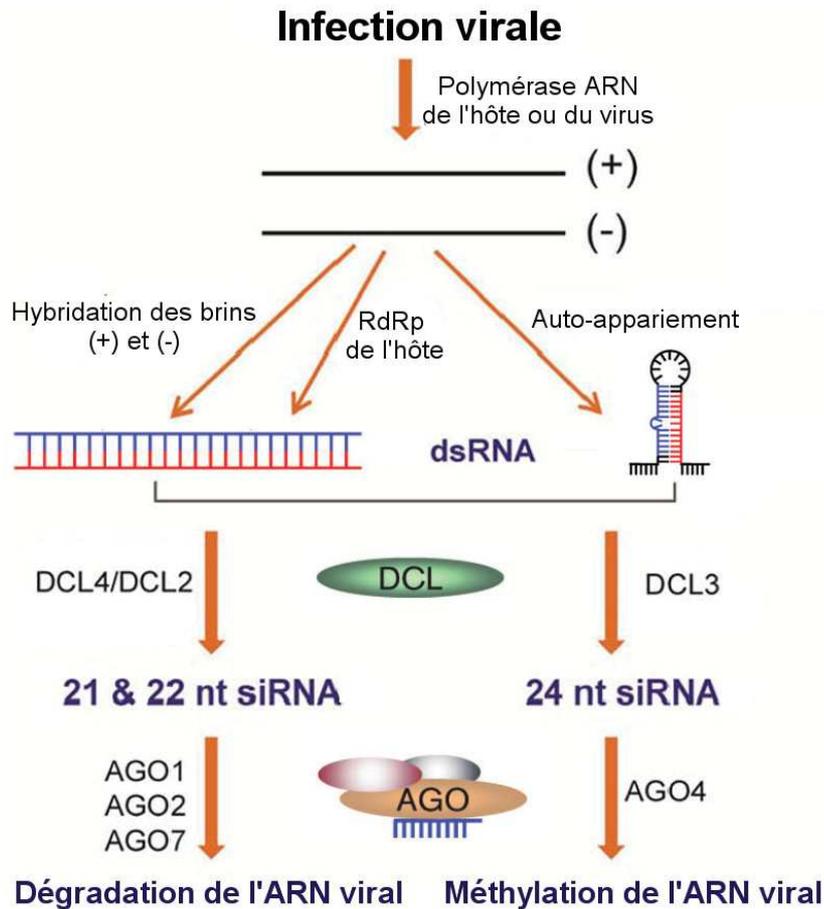


FIGURE 1.1 – Mécanismes de *silencing* des ARN (adapté de Wang et al. (2012)). Lors de l'infection virale, des ARN viraux à polarité positive et/ou négative sont produits. Des structures double-brin sont alors générées pendant la synthèse des brins positifs, par hybridation des brins à polarité positive et négative ou par formation de structure secondaire de l'ARN viral. Les ARNdb sont alors clivés en siRNA par différents DCL qui sont recrutés dans les complexes RISC qui permettent la dégradation de l'ARN viral ou sa méthylation.

la paroi pecto-cellulosique et la taille, forme et localisation des stomates sont autant d'obstacles à contourner par l'agent pathogène pour infecter la plante.

Les virus pénètrent les cellules par une blessure mécanique de la plante qui rompt ces barrières ou grâce à des vecteurs (insectes, nématodes ou champignons). Afin de permettre la pénétration des virus, leurs vecteurs doivent être adaptés à ces barrières physiques.

Une fois ces défenses passives traversées, un premier système de défenses inductibles entre en jeu. Dans le cas des virus, il s'agit du *silencing* des ARN (ou interférence ARN) (Figure 1.1) (pour revue Pumplin and Voinnet (2013); Sharma et al. (2013)). Ce mécanisme de défense induit une résistance souvent incomplète. Dans le cas des plantes on le subdivise en deux processus proches : le *silencing* de gène transcriptionnel (TGS) et le *silencing* de gène post-transcriptionnel (PTGS) (Hannon, 2002).

La première étape de ce processus est son initiation par les ARN double-brin (ARNdb). Il y a deux sources possibles d'ARNdb viraux. Il peut s'agir de fragments générés lors de la réplication des virus à ARN par les polymérase ARN dépendantes de l'ARN (RDRp) de l'hôte, lors d'une transcription bidirectionnelle des virus à ADN, ou encore par transcription convergente. Il peut également s'agir de structures secondaires de l'ARN viral. Que ce soit pour le PTGS ou le TGS, ces ARNdb sont reconnus par des endoribonucléases (RNase) appelées Dicer-like (DCL) qui génèrent des petits ARN interférants (siRNA). Les DCL ont un domaine Piwi/Argonaute/Zwille (PAZ) à une extrémité et des domaines RNase III à l'autre. Elles produisent les siRNA en identifiant les extrémités des ARNdb via leur domaine PAZ et en les clivant via leurs domaines RNase. *Arabidopsis thaliana* code pour quatre DCL dont les contributions relatives ont été récemment élucidées (Qu et al., 2008).

Dans le cas du PTGS, DCL4 et DCL2 sont les plus importants. Ils génèrent respectivement des siRNA de 21 et 22 nucléotides. Les siRNA sont ensuite adressés au complexe de *silencing* induit par l'ARN (RISC). Les composants actifs de ce complexe sont les protéines argonautes (AGO) qui clivent l'ARNm viral complémentaire du siRNA. *A. thaliana* a 10 protéines AGO. AGO1, 2 et 7 ont un rôle dans le PTGS, AGO1 étant le principal. Le PTGS est plus le mécanisme prédominant dans le *silencing* des virus à ARN (Pumplin and Voinnet, 2013).

Le deuxième mécanisme, le TGS, cible les virus à ADN. Les siRNA impliqués font 24 nucléotides et sont clivés par DCL3. Ils sont intégrés dans un complexe de *silencing* transcriptionnel induit par l'ARN (RITS) contenant AGO4. Ces siRNA servent de brin guide pour la formation d'hétérochromatine et la méthylation des cytosines de l'ADN viral cible. On parle de méthylation ADN dirigée par l'ARN (RdDM).

Pour les deux mécanismes, l'efficacité de l'interférence ARN dépend de la production de siRNA secondaires (Pumplin and Voinnet, 2013). Ces derniers se propagent de cellule en cellule par les plasmodesmes ou sur de longues distances via le réseau du phloème. Cette amplification du signal est dépendante des RDRp de la plante.

Tous les virus de plantes dans lesquels des protéines ayant une activité suppresseur

d'interférence ARN ont été recherchées codent pour au moins une d'entre elles (Burgýán and Havelda, 2011). On parle de suppresseurs viraux d'interférence ARN (VSR). Ils sont structurellement divers et affectent des niveaux différents de la machine de *silencing* de l'hôte en se liant à un de ces composants.

Le *silencing* des ARN est proche de la PTI du modèle de l'immunité végétale (Figure 1.2). Les ARNdb sont une structure commune aux virus, leur seule différence avec les PAMP est qu'ils sont intracellulaires contrairement à ces derniers. De même, les VSR jouent ici un rôle proche de celui des effecteurs pour contrer le *silencing*.

1.2 Les gènes majeurs

Si les défenses non-spécifiques de l'hôte sont contournées par l'agent pathogène, un autre niveau de résistance peut entrer en jeu. Il s'agit des gènes de résistances spécifiques à effet majeur. Ils peuvent être dominants ou récessifs.

Les résistances monogéniques dominantes représentent 51 % des résistances monogéniques aux virus décrites (Gómez et al., 2009). Elles sont généralement conférées par un gène de résistance (*R*) unique qui code une protéine interagissant directement

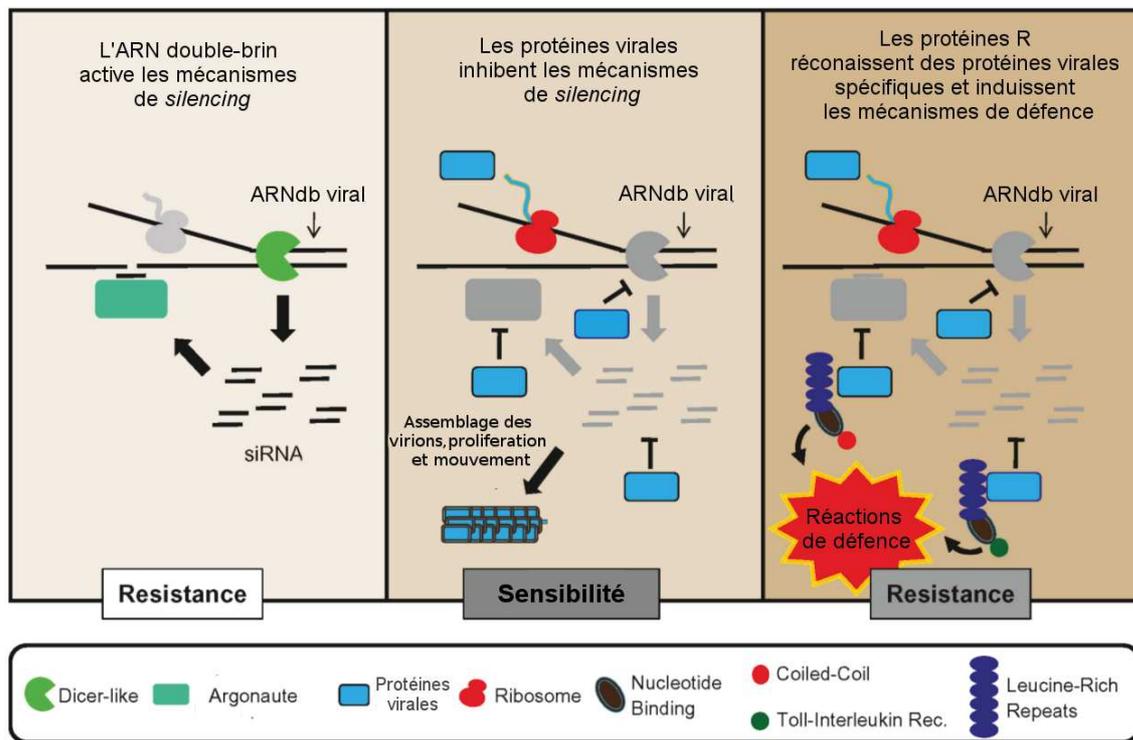


FIGURE 1.2 – Analogies au modèle de l'immunité virale du point de vue de l'infection virale (publié dans Moffett (2009)). À gauche : l'interférence ARN comme PTI. La reconnaissance d'ARNdb par des protéines DCL conduit à la formation de siRNA recruté dans un complexe RISC clive ou méthyle l'ARN viral, stoppant la réplication. Au milieu : Une protéine du virus, appelée VSR, inhibe l'interférence ARN. À droite : l'ETI contre les virus. Les protéines NB-LRR de la plante reconnaissent une protéine virale conduisant aux mécanismes de défenses aboutissant la plupart du temps à une HR.

ou indirectement avec la protéine de l'agent pathogène correspondante : la protéine d'avirulence (Avr) (de Ronde et al., 2014). Cette résistance est souvent isolat-spécifique et l'interaction entre les protéines de l'agent pathogène et de la plante se produit dans la cellule végétale. La reconnaissance, directe ou indirecte, de la protéine d'avirulence par la protéine de résistance conduit généralement à une mort cellulaire programmée rapide et localisée qui empêche la propagation du virus : c'est la réponse hypersensible (HR). Dans certains cas il n'y a pas de HR, on parle alors de résistance extrême (ER). Lorsque les protéines d'avirulence ou de résistance sont modifiées ou absentes, la reconnaissance n'a pas lieu et la plante est sensible. Le type de gène *R* utilisé, ainsi que la cascade de signalisation correspondante, sont les mêmes que celles d'agents pathogènes non-viraux. Ce type de résistance est donc très proche de l'ETI (Mandadi and Scholthof, 2013). Cependant, beaucoup des protéines d'avirulence virale ne sont pas des VSR et ne sont donc pas nécessairement analogues aux effecteurs des autres agents pathogènes.

Les résistances récessives ne sont la plupart du temps pas liées à des mécanismes de défense induits par la présence de l'agent pathogène. Il s'agit la plupart du temps de mutations de gènes de sensibilité. Si un facteur de la plante clé pour l'infection par l'agent pathogène est absent ou modifié, celle-ci devient impossible. On parle parfois de sensibilité altérée pour qualifier ce type de résistance (Maule et al., 2007), assez éloignée de la définition de l'ETI. De part la très faible taille de leur génome, les virus ne codent qu'un faible nombre de protéines et ont besoin d'un grand nombre de facteurs de l'hôte pour accomplir leur cycle. La nature des virus explique pourquoi on trouve beaucoup plus de gènes de résistance récessive aux virus qu'aux autres agents pathogènes.

Deux classes majeures de gènes de résistance correspondent à ces deux types de résistances spécifiques.

1.2.1 Les NB-LRR

Les NB-LRR font partie des familles les plus fréquentes de protéines impliquées dans des résistances dominantes. Ce sont des protéines cytoplasmiques formées d'un domaine *leucine-rich repeat* (LRR) lié à un domaine de liaison au noyau (*nucleotide-binding adaptor shared by Apaf1, certain R genes and CED4* (NB-ARC)) (Moffett, 2009; Takken and Govere, 2012). Le domaine LRR est impliqué dans la spécificité de reconnaissance de l'agent pathogène. Le domaine NB-ARC est formé de 3 sous-unités (NB, ARC1 et ARC2) comprenant une poche de liaison nucléotidique permettant la liaison et l'hydrolyse de l'adénosine triphosphate (ATP) (Tameling et al., 2002). Certains motifs très conservés du domaine NB-ARC pourraient également être impliqués dans la cascade de signalisation aval. Un troisième domaine en N-terminal permet de grouper les NB-LRR en deux classes :

- les TIR-NBS-LRR, possédant un domaine *Toll and Interleukin-1 Receptor* (TIR), absents chez les angiospermes monocotylédones. Le rôle du domaine TIR n'est pas encore clairement élucidé.
- les CC-NBS-LRR, possédant un domaine *coiled-coil* (CC) permettant des in-

teractions protéine-protéine. Les NB-LRR sans structure prédite en N-terminal sont classés avec les CC-NBS-LRR dans le groupe des "non-TIR".

Trois modèles d'interactions entre les protéines de résistance (R) et d'avirulence (Avr) ont été proposés. Ils expliquent le modèle de l'interaction gène pour gène au niveau moléculaire (Figure 1.3) (Moffett, 2009; van der Hoorn and Kamoun, 2008). Ils sont valables pour les NB-LRR mais également pour d'autres gènes de résistance majeurs comme les *receptor-like kinase* (LRR-RLK).

- Le modèle ligand-récepteur : dans ce modèle, la protéine R interagit de manière directe avec la protéine Avr. La transduction du signal est assurée par la protéine R. Bien que ce soit le premier modèle proposé, peu de couples R/Avr interagissant

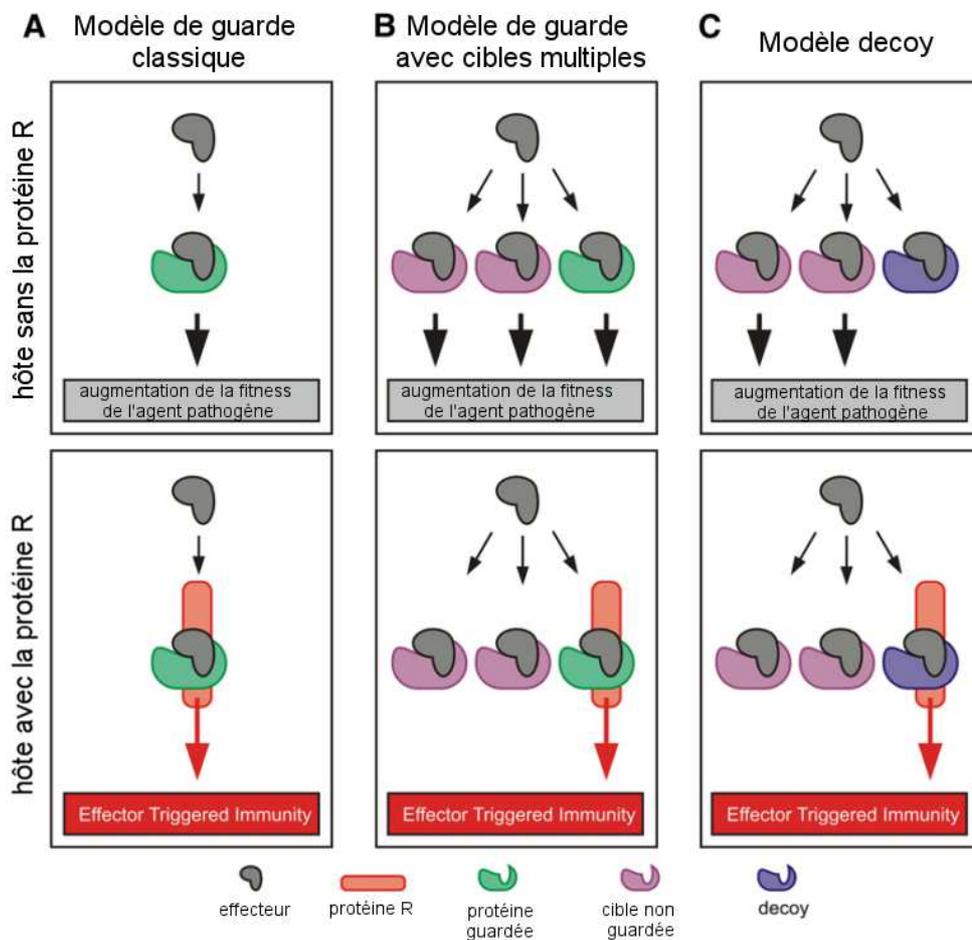


FIGURE 1.3 – Comparaison des modèles de garde et *decoy* (publié dans van der Hoorn and Kamoun (2008)). Dans le modèle de garde (A), la protéine de l'agent pathogène interagit avec sa cible de l'hôte (la protéine gardée) permettant l'infection. Dans les plantes résistantes, cette interaction est reconnue par une protéine R, permettant la résistance. On peut également imaginer un modèle de garde modifié (B) où la protéine de l'agent pathogène interagit avec plusieurs protéines de l'hôte, toutes améliorant la fitness de l'agent pathogène mais où une seule de ces interactions est reconnue par la protéine R. Dans le cas du modèle decoy (C), la protéine de l'agent pathogène interagit avec des cibles améliorant l'infection mais également avec une protéine de l'hôte sans effet sur sa fitness (le *decoy*). C'est cette dernière interaction qui est reconnue par la protéine R.

directement ont été identifiés.

- Le modèle de garde : c'est le modèle le plus communément accepté. La protéine R n'interagit pas directement avec la protéine d'avirulence mais avec un récepteur spécifique de cette dernière. C'est la liaison avec ce récepteur, la protéine gardée, lié à la protéine d'avirulence, qui va activer la protéine R et mener à la résistance. La protéine gardée a une fonction dans la sensibilité de l'hôte en l'absence de la protéine R. Dans une plante sensible, l'interaction de la protéine gardée avec la protéine Avr conduit à une meilleure fitness de l'agent pathogène. Cette protéine subit donc des forces évolutives différentes en présence ou en absence du gène de résistance. En absence du gène de résistance, elle va avoir tendance à accumuler des mutations conduisant à une suppression de l'interaction avec la protéine Avr alors qu'en présence du gène de résistance, ce sont des mutations améliorant cette interaction qui vont avoir tendance à se fixer.
- Le modèle *decoy* : ce modèle est dérivé du modèle de garde. La protéine permettant la reconnaissance du virus est un leurre (*decoy*) pour la protéine Avr. Cette protéine n'a pas de rôle dans l'établissement de la virulence de l'agent pathogène.

Les modèles de garde et *decoy* sont difficiles à différencier. Plusieurs protéines interagissant avec des protéines d'avirulence, et pouvant donc répondre aux critères d'un de ces deux modèles, ont été identifiées à ce jour. Dans le cas des pathosystèmes viraux il s'agit par exemple de la protéine NRIP1 qui interagit d'une part avec la protéine p50 du *Tobacco mosaic virus* (TMV) et d'autre part avec la protéine NB-LRR N du tabac (Caplan et al., 2008), ou encore de la protéine RanGAP2 qui interagit avec la protéine NB-LRR Rx de la pomme de terre et avec la protéine de capsid (CP) du *Potato virus X* (PVX) (Sacco et al., 2007).

Toutefois, malgré le rôle reconnu du domaine LRR dans la spécificité de reconnaissance, ces co-facteurs n'interagissent pas avec ce domaine du NB-LRR mais avec le domaine N-terminal (TIR dans le cas de la protéine N, CC dans le cas de Rx). Le modèle *bait and switch* a été proposé afin d'expliquer cette contradiction (Figure 1.4) (Collier and Moffett, 2009). Dans ce modèle, le domaine N-terminal des NB-LRR et le domaine LRR jouent un rôle conjoint dans la reconnaissance de l'agent pathogène. En l'absence de reconnaissance d'un agent pathogène, le NB-LRR est auto-inhibé. Il est replié de façon compacte, le domaine NB-ARC interagissant simultanément avec les domaines N-terminal et LRR. La protéine gardée, ou *decoy*, joue le rôle d'appât (*bait*). Elle est liée au domaine N-terminal de la protéine R et interagit avec l'Avr. Le domaine LRR interagit alors avec l'Avr. C'est cette interaction qui est la plus importante dans la spécificité de reconnaissance. Elle cause une perturbation du repliement de la protéine et provoque des changements de conformation, aboutissant à la transduction du signal de résistance.

Afin d'aboutir à une réponse de défense, plusieurs modules protéiques sont nécessaires aux gènes *R*. Le complexe SGT1/RAR1/HSP90 forme un complexe chaperonne permettant le repliement correct et la stabilité des NB-LRR (Takahashi et al., 2003).

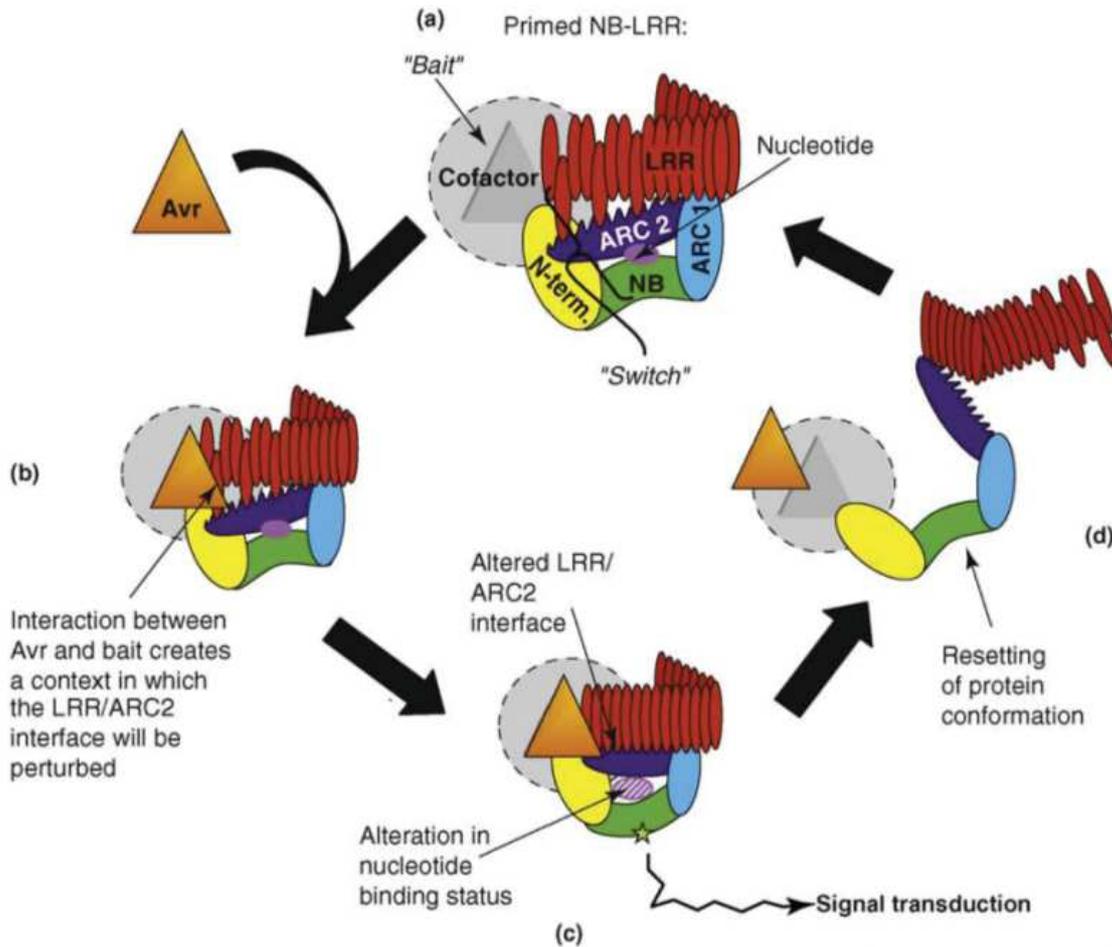


FIGURE 1.4 – Modèle "bait and switch" du fonctionnement des NB-LRR (publié dans Collier and Moffett (2009)). (a) La protéine NB-LRR est auto-inhibée. Le "switch" est sur off. Plusieurs interactions moléculaires, notamment entre le domaine LRR et le domaine ARC2, maintiennent la protéine dans cet état. (b) La protéine Avr de l'agent pathogène est introduite dans le système du NB-LRR via la protéine "bait" (appât). (c) Son contact avec le domaine LRR perturbe la liaison entre les domaines LRR et ARC2 change la conformation de la protéine et aboutit à la transduction du signal de défense. (d) La protéine est réinitialisée.

Le complexe EDS1/PAD4/SAG101 régule le niveau des réponses de défenses (Wiermer et al., 2005).

Ces dernières années, de nombreux cas de NB-LRR fonctionnant par paires ont été décrits (Eitas and Dangl, 2010). C'est notamment le cas du couple *N/NRG1* du tabac qui lui confère la résistance au TMV (Peart et al., 2005). Le fonctionnement des NB-LRR en homo- ou hétéro-dimer est de plus en plus admis.

La HR et la résistance sont des phénomènes étroitement liés, toutefois l'existence de la résistance extrême montre qu'il reste des phénomènes indépendants. Pour activer la HR, la plante active plusieurs réponses moléculaires, identiques quel que soit le type d'agent pathogène. Tout d'abord, des cascades de signalisation induisent de nombreuses protéines dont les *mitogen-activated protein kinase* (MAPK). En aval des ces cascades de signalisation, des hormones telles que l'acide salicylique (SA) et l'acide jasmonique sont synthétisées, il y a une accumulation d'espèces réactives de l'oxygène (ROS) et influx d'ions calcium, dépôts de callose au niveau des plasmodesmes et expression de gènes *pathogenesis related* (PR) (Carr et al., 2010; Loebenstein, 2009). La synthèse de SA active une réponse de résistance systémique acquise (SAR) qui permet l'amplification de réponses de défense dans l'ensemble de la plante, notamment en y activant des gènes PR. La SAR prévient ainsi l'infection par d'autres agents pathogènes.

1.2.2 Les facteurs d'initiation de la traduction

La principale classe de gènes de résistance récessive est celle de facteurs d'initiation de la traduction (eIF).

Ces protéines sont nécessaires à la synthèse de protéines dans la cellule végétale. La traduction des ARN messagers (ARNm) eucaryotes dépend de la liaison de eIF4E à leur coiffe en 5' (Robaglia and Caranta, 2006; Sanfaçon, 2015). La *poly-A binding protein* (PABP) se lie en 3' de l'ARN viral et la protéine eIF4G se lie à la fois à eIF4E et à PABP, permettant une circularisation de l'ARNm. La liaison entre eIF4E et eIF4G forme le complexe eIF4F. Une hélicase ARN à DEAD-box, eIF4A, interagit avec eIF4G et linéarise l'ARN, permettant une lecture plus efficace par le ribosome. eIF4F est lié à la sous-unité 40S du ribosome par le complexe protéique eIF3 (Figure 1.5).

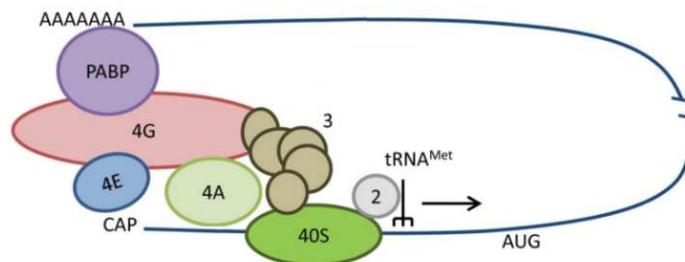


FIGURE 1.5 – Diagramme simplifié de l'initiation de la traduction chez une cellule eucaryote (adapté de Sanfaçon (2015)).

Le complexe eIF4F a deux isoformes : eIF4F, formé de eIF4E et eIF4G, et eIF(iso)4F, composé de eIF(iso)4E et eIF(iso)4G. Ces isoformes présentent de la redondance fonctionnelle mais également une spécificité distincte de cibles. Par ailleurs, eIF(iso)4G interagit également avec les microtubules, suggérant un rôle dans la dynamique des microtubules (Bokros et al., 1995). eIF4E a une localisation cytoplasmique mais également nucléaire (Bush et al., 2009). Par analogie avec le rôle des eIF4E des mammifères, son rôle dans l'export nucléaire est donc probable.

eIF4E, *eIF4G* et leurs isoformes sont des gènes de résistance récessive courants dans le cas des infections par virus à ARN positifs. Le premier décrit a été le gène *pvr2* qui code pour une protéine eIF4E et confère la résistance du piment à deux potyvirus (*Potato virus Y*, PVY et *Tobacco etch virus*, TEV) (Ruffel et al., 2002). Depuis, les eIF ont été décrits comme impliqués dans un grand nombre de résistances à des *potyvirus* mais également à d'autres familles plus éloignées. Un exemple d'eIF4G intervenant dans la résistance est celui du facteur codé par le gène *tsv1*, conférant au riz la résistance au *Rice tungro spherical virus* (RTSV), un des principaux virus du riz, appartenant à la famille des *Waikavirus* (Lee et al., 2010).

Les virus dépendent des facteurs de l'hôte pour synthétiser leurs protéines. Toutefois les ARN des virus à ARN positif n'ont pas de coiffe en 5', pas de queue polyA en 3' ou aucun des deux. Ils doivent donc déployer des stratégies originales afin de traduire leur ARN. L'une d'elle est le recrutement d'eIF ou d'ARN ribosomiaux par des séquences améliorant la transduction en *cis* (CITE) de l'ARN viral contenant des sites internes d'entrée de ribosome (IRES). Certains ARN viraux sont liés en 5' à la VPg, une petite protéine virale désordonnée, c'est-à-dire manquant de structure tridimensionnelle stable et ayant donc une plasticité importante. Les VPg de plusieurs virus sont capables d'interagir directement avec les facteurs d'initiation de la traduction eIF4E, eIF4G ou leurs isoformes (Sanfaçon, 2015).

Jusqu'à présent, le rôle précis de la VPg dans l'infection des virus chez les plantes n'a pas encore été élucidé. Plusieurs hypothèses, potentiellement non exclusives, ont cependant été proposées (Wang and Krishnaswamy, 2012). La VPg pourrait notamment fonctionner comme un analogue de la coiffe de l'ARNm de la cellule hôte. Elle augmenterait l'affinité du complexe eIF4F pour l'ARN viral et promouvrait donc sa traduction (Khan et al., 2008; Leonard et al., 2000; Plante et al., 2004). Cette augmentation de l'affinité pour l'ARN viral pourrait être liée à une diminution de la traduction des ARN de l'hôte. Par ailleurs, la colocalisation de facteurs eIF4G avec les microtubules de la cellule hôte suggère que cette interaction pourrait être impliquée dans le mouvement de cellule à cellule du virus dans son hôte. La localisation nucléaire de la VPg et d'eIF4E suggère un rôle dans la suppression de l'export des ARNm du noyau au cytoplasme et dans la perturbation du traitement des siRNA dans le noyau, contrant ainsi le *silencing* des ARN.

1.2.3 Les autres types de gènes majeurs

D'autres gènes de résistance dominante et récessive n'appartenant pas aux deux familles précédemment décrites ont été mis en évidence.

Parmi les gènes conférant une résistance dominante, au moins 5 gènes *restricted TEV movement (RTM)* sont impliqués dans la résistance d'*A. thaliana* aux *Tobacco etch potyvirus* (TEV), *Lettuce mosaic potyvirus* (LMV) et *Plum pox virus* (PPV) (Cosson et al., 2012). Seuls *RTM1*, *RTM2* et *RTM3* ont été clonés pour l'instant. *RTM1* code pour une protéine de la famille des jacalines, dont certains membres sont impliqués dans la défense contre des insectes ou des champignons (Chisholm et al., 2000). *RTM2* est une protéine avec un domaine N-terminal similaire aux petites protéines de type *heat-shock* (HSP) et un domaine C-terminal contenant de multiples répétitions (Whitham et al., 2000). Les HSP forment des complexes oligomériques et ont un rôle de chaperonne. *RTM3* code pour une protéine interagissant avec *RTM1*. Elle appartient à une famille non-décrite ayant un domaine *meprin and TRAF homology* (MATH) dans sa partie N-terminale et un domaine CC en C-terminal (Cosson et al., 2010a,b). Le domaine MATH pourrait être impliqué dans des interactions protéine-protéine. L'interaction de *RTM3* avec *RTM1* suggère que les protéines *RTM* seraient impliquées dans un complexe multi-protéines du phloème. La résistance est conférée par un blocage du mouvement longue distance du virus que la mutation d'un seul des gènes *RTM* suffit à rétablir. La protéine d'avirulence correspondante est la protéine de capsid (CP) du virus, qui n'interagit toutefois pas directement avec *RTM1*, *RTM2* ou *RTM3*. Il n'y a pas d'induction de HR ou de production d'acide salicylique dans le cas de ces résistances.

De manière intéressante, un autre gène dominant récemment identifié, *JAX1*, est également un gène de la famille des jacalines, proche de *RTM1*. Il confère à *A. thaliana* une résistance large spectre contre les *potyvirus*, notamment *Plantago asiatica mosaic virus* (PIAMV) (Yamaji et al., 2012). Dans le cas de cette résistance, l'accumulation des ARN viraux est altérée au niveau cellulaire.

De même, un ou plusieurs gènes contenant un domaine MATH sont candidats pour la résistance de l'abricot au PPV impliquant la limitation du mouvement longue distance du virus (Zuriaga et al., 2013).

Tm-1, identifié chez une espèce de tomate sauvage, *Solanum hirsutum*, confère une résistance semi-dominante contre le *Tomato mosaic virus* (ToMV). Il code pour une protéine contenant une structure de type *TIM-barrel* qui interagit directement avec la réplicase du virus, inhibant la formation du complexe de réplication précédant la synthèse des ARN négatifs et donc la réplication du virus. (Ishibashi and Ishikawa, 2013; Ishibashi et al., 2007; Kato et al., 2013).

Les gènes de résistance de la tomate contre le *Tomato yellow leaf curl begomovirus* (TYLCV), *Ty-1* et *Ty-3*, sont des formes alléliques d'un même gène et codent pour un RDRp de la tomate. Ils conféreraient la résistance en amplifiant le signal de *silencing*

des ARNs (Verlaan et al., 2013). À ce jour, ce sont les seuls gènes de résistance majeurs impliqués dans le *silencing* identifiés.

Un autre type de résistance original est celui conféré par *STV11*. Ce gène confère au riz la résistance au *Rice stripe virus* (RSV) (Wang, 2014). Il code pour une sulfo-transférase qui catalyse la conversion de SA en SA sulfoné, plus efficace que SA pour inhiber la réplication du virus.

Des gènes de résistance récessifs non-eIF ont également été décrits. C'est le cas de la *PDI5-1* chez l'orge (Yang et al., 2014) qui confère la résistance aux Bymovirus. Chez l'humain, il a été montré que les protéines de la famille des *protein disulfide isomerase* (PDI) jouent un rôle de chaperonnes cellulaires, permettant un repliement correct des protéines, leur stabilisation et facilitant leur transport. Par homologie, l'allèle de sensibilité de *PDI5-1* pourrait jouer un rôle de chaperonne pour les protéines virales et être donc essentiel à la réplication du virus.

Le gène *rum1* d'*Arabidopsis*, qui confère la résistance au *Watermelon mosaic potyvirus* (WMV), est lui un gène nucléaire codant pour une phosphoglycérate kinase chloroplastique (Ouibrahim et al., 2014). Il agit en limitant l'accumulation virale dans les tissus infectés à un stade précoce de l'infection.

Dans les mutants *tom1* et *tom2A* d'*Arabidopsis*, l'accumulation du TMV dans les cellules est supprimée. Les deux gènes codent pour des protéines transmembranaires localisées dans le tonoplaste et sont nécessaires à la réplication du tobamovirus (Ishibashi et al., 2012).

D'autres gènes de résistances non clonés ont été mappés dans des régions ne contenant pas de NB-LRR ou de facteur d'initiation de la traduction, permettant d'espérer la découverte de nouveaux mécanismes de résistance au virus.

1.3 Résistance quantitative

Plus de 80 % des résistances aux virus décrites sont monogéniques. Pourtant la plupart des traits agronomiques sont quantitatifs et polygéniques (Maule et al., 2007). L'étude plus facile des caractères qualitatifs et les résistances quantitatives ont donc été moins étudiées.

La résistance quantitative est apportée par plusieurs gènes qui contribuent chacun à une légère augmentation de la résistance (Kou and Wang, 2012). Les phénotypes de résistance ou de sensibilité observés dans une descendance ou une population non apparentée présentent ainsi une distribution continue. Les locus impliqués dans ce type de résistance sont appelés *quantitative trait loci* (QTL) de résistance ou *quantitative resistance locus* (QRL) (Kou and Wang, 2010). Ce type de résistance entraîne une pression de sélection moindre sur l'agent pathogène et est donc présumée plus durable. Toutefois, ce paradigme souvent invoqué pourrait également être remis en cause. En effet, si la pression de sélection exercée est moins forte, la multiplication résiduelle dans la plante peut permettre l'émergence de variants contournant la résistance partielle

non préexistante dans la population (Montarry et al., 2012). Associés ensemble ou avec un gène majeur dans un même germplasm, ils rendent cependant plus difficile le contournement complet de la résistance. La nature des QRL est encore mal connue et supposée multiple (Poland et al., 2009). En effet le gain de résistance peut être acquis à tous les niveaux de la résistance, de l'échappement, au gène de résistance classique atténué, en passant par la transduction du signal de défense.

Ainsi, il peut s'agir de gènes régulant la phénologie de la plante. Par exemple la capacité de régulation des stomates peut permettre de limiter l'humidité et donc le développement de certains agents pathogènes. De même, la date de floraison peut permettre de décaler le cycle de la plante par rapport à celui de l'agent pathogène.

Il peut également s'agir de variants alléliques de gènes impliqués dans la résistance non spécifique ou encore de formes plus faibles de gènes majeurs, soit de par une reconnaissance imparfaite de la protéine d'avirulence, soit à cause d'une réponse incomplète. C'est par exemple le cas de *Xa21D* qui code pour un récepteur kinase avec un domaine LRR extracellulaire et qui confère une résistance partielle à *Xanthomonas oryzae* pv *oryzae* avec le même spectre que *Xa21*, qui confère lui une résistance complète (Wang et al., 1998). De même, deux QRL clonés sont des NBS-LRR : *Pb1* qui est un gène de résistance à *Magnaporthe oryzae* chez *Oryza sativa* (Hayashi et al., 2010) et *Rcg1* qui confère au maïs une résistance à *Colletotrichum graminicola* (Broglie and Butler, 2008). Toutefois certains domaines conservés chez les NBS-LRR classiques sont absents ou modifiés dans ces deux gènes. Dans le cas des résistances aux virus, certains facteurs d'initiation de la traduction ont été décrits comme conférant une résistance partielle (Charron et al., 2008; Nicaise, 2003).

Il peut également s'agir de gènes impliqués dans la synthèse de molécules telles que SA ou JA. C'est notamment le cas du gène *StAOS2* de *Solanum tuberosum* qui confère une résistance à *Phytophthora infestans* et à *Erwinia carotovora* spp. *astrospetica* via son rôle dans la synthèse d'acide jasmonique (Pajerowska-Mukhtar et al., 2008).

Enfin, certains QRL ne colocalisent avec aucun gène connu comme étant impliqué dans les défenses de la plante et peuvent donc correspondre à des familles de gènes encore non-identifiés.

Plus nous en apprenons sur les résistances des plantes et la diversité des mécanismes qui y sont impliqués, et plus la pertinence de la catégorisation en défense basale, ETI, PTI et résistance quantitative est remise en question. En effet, tel qu'énoncé actuellement, le modèle de l'immunité des plantes ne permet pas d'expliquer toute la complexité des pathosystèmes.

1.4 Résistance aux vecteurs

Les virus dépendent de vecteurs pour leur transmission. Lutter contre les vecteurs peut donc être un moyen de contrôler les infections virales.

Deux NB-LRR contre des insectes vecteurs viraux ont été décrits à ce jour. Le gène

Mi-1 de la tomate confère la résistance au puceron de la pomme de terre (*Macrosiphum euphorbiae*), à l'aleurode (*Bemisia tabaci*) et aux nématodes à galle (*Meloidogyne* spp.) dont les deux premiers sont des vecteurs viraux (Vos et al., 1998). Le gène *Vat* contrôle la résistance au puceron *Aphis gossypii* et bloque le mode de transmission non-persistant des virus (Dogimont et al., 2014).

Par ailleurs des QTL et gènes majeurs de résistance de natures très diverses contre d'autres insectes vecteurs ont également été décrits. On peut par exemple citer le cas de la résistance du riz à différentes espèces de cicadelles (Fujita et al., 2013).

1.5 Contournement et durabilité des résistances

De nombreux gènes de résistance contre des agents pathogènes déployés au champ ont été contournés en quelques années. On parle de *boom-and-bust* des gènes de résistance (Mundt, 2014). Tout d'abord le gène est efficace et utilisé de façon massive (*boom*), puis il est contourné par l'agent pathogène et est abandonné (*bust*). Le nombre de gènes de résistance disponibles étant limité, les recherches s'orientent vers des moyens d'obtenir une résistance plus durable.

Le phénomène du contournement commence par l'apparition d'un variant virulent de l'agent pathogène dans une plante. Ce variant se multiplie alors dans la plante, qu'il infecte. Il se dissémine ensuite aux plantes hôtes voisines (Figure 1.6). Les virus étant des agents pathogènes pouvant évoluer très rapidement au sein même d'une plante, ils ont un haut potentiel de contournement des résistances.

La plupart des résistances impliquant des interactions protéine-protéine, les mécanismes de contournement impliquent souvent la rupture de cette interaction. Dans le

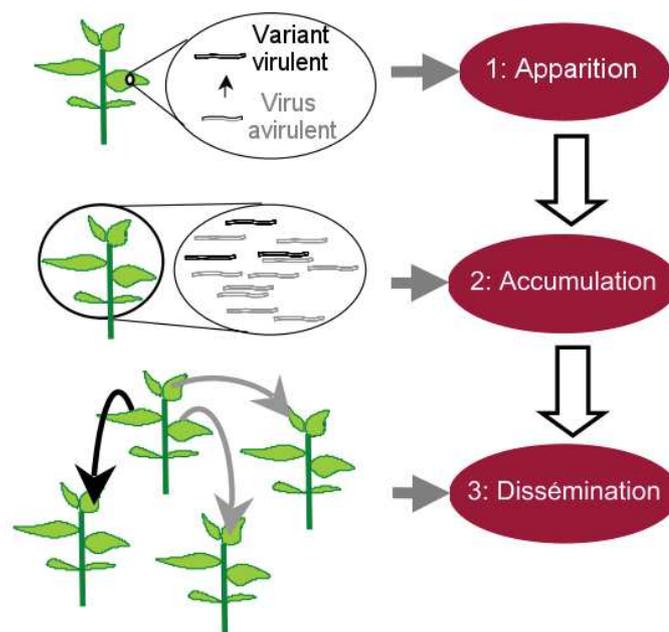


FIGURE 1.6 – Étapes du contournement de la résistance des plantes par les virus (adapté de (Gómez et al., 2009)).

cas des NB-LRR, la mutation de la protéine d'avirulence peut lui permettre d'échapper à sa reconnaissance cellulaire (Kobayashi et al., 2014). Cet échappement est parfois incomplet. La réponse de l'hôte et l'induction de la HR sont alors retardées. L'agent pathogène partiellement contournant peut alors circuler dans la plante, induisant une nécrose systémique pouvant occasionner des dégâts supérieurs à ceux de l'infection virale (Komatsu et al., 2010). Dans le cas des eIF, le rétablissement de l'interaction entre le facteur viral et l'eIF impliqué dans la résistance permet au virus d'infecter la plante (Moury et al., 2004).

Toutefois, les virus n'ont pas le même potentiel de redondance génétique que les autres agents pathogènes. Leurs protéines sont pléiotropes et leurs ORF sont parfois partiellement chevauchants. Les protéines mutées des variants virulents doivent conserver au moins une partie de leur fonction pour que la fitness du virus soit suffisante à sa dissémination. Les variants virulents ont donc souvent une fitness moindre dans un hôte sensible (García-Arenal and Fraile, 2013; Montarry et al., 2012).

Pour limiter le contournement, on peut agir sur l'apparition ou la dissémination du variant contournant. L'usage de plusieurs résistances aux déterminismes génétiques différents dans un même génotype permet de limiter leur apparition. Cette méthode, appelée *pyramiding*, repose sur le principe que plus le nombre de mutations à accumuler pour infecter la plante est grand, plus la probabilité que ces événements arrivent en même temps est faible (Quenouille et al., 2013). De plus, un virus ayant l'ensemble des mutations aurait certainement une fitness grandement réduite. Cette technique a également l'intérêt d'accroître le spectre d'efficacité de la résistance. Certains isolats viraux sont virulents vis-à-vis d'un gène de résistance donné, mais cette spécificité d'isolats est le plus souvent différente pour un autre gène de résistance au même virus. En cumulant des gènes aux spécificités différentes, la résistance totale obtenue est donc à plus large spectre. Cette technique ne s'applique pas qu'aux gènes majeurs. L'utilisation de plusieurs QRL ou d'une combinaison entre QRL et gènes majeurs est également souvent mise en avant pour améliorer la durabilité.

Le *pyramiding* a été appliqué avec succès dans plusieurs pathosystèmes viraux, que ce soit avec des gènes majeurs (Shi et al., 2009; Werner et al., 2005) ou par combinaison d'un gène majeur et de plusieurs QRL (Palloix et al., 2009; Quenouille et al., 2015).

D'autres méthodes reposant sur un déploiement rationalisé des gènes de résistance dans l'espace et le temps ont également été proposées (Mundt, 2014). L'une d'entre elles est la rotation des gènes de résistance. L'idée derrière cette méthode est de déployer un gène à la fois. La pression de sélection liée à un des gènes disparaîtrait alors avant que les mutations de contournement associées n'aient pu se fixer dans la population de l'agent pathogène. Cependant cette technique se heurte à plusieurs écueils : la surveillance de l'apparition d'isolats virulents est complexe, l'accord des agriculteurs pour changer brusquement de variété n'est pas toujours aisé à obtenir et, enfin, la régression des variants virulents dans la population virale n'est pas acquise si leur fitness est suffisamment bonne.

Une autre méthode est celle du mélange de gènes dans un champ. Plusieurs germoplasmes avec chacun un gène différent sont semés en mélange dans un champ. Le principe est une dissémination plus difficile des variants virulents. Toutefois l'efficacité de cette technique est controversée, elle pourrait également permettre une infection plus facile du champ et il est techniquement difficile d'obtenir des variétés avec des gènes de résistance différents qui soient suffisamment synchrones pour une culture intensive.

Enfin, des stratégies de déploiement régional des résistances sont discutées. Il s'agit de prendre en compte les aspects géographiques d'une maladie pour déployer les différents gènes disponibles. On peut ainsi créer une mosaïque de gènes à l'échelle régionale limitant la dissémination de variants contournants. Cette stratégie nécessite une bonne coordination avec les agriculteurs et leur efficacité n'a pas pu être démontrée.

Ces méthodes sont difficiles à tester et peu d'études ont pu les comparer. Djian-Caporalino et al. (2014) ont toutefois montré l'efficacité de l'alternance de gènes de résistance et de leur *pyramiding* dans un pathosystème piment/nématode. Devant les difficultés pratiques des autres méthodes, le *pyramiding* semble être la meilleure méthode.

1.6 Modèles évolutifs des gènes de résistance

La sélection, qu'elle soit naturelle ou anthropique, a plusieurs formes : elle peut être positive lorsqu'un allèle favorable augmente en fréquence, ou purificatrice lorsqu'elle agit sur l'élimination d'allèles défavorables. La sélection positive peut elle-même être directionnelle, auquel cas une unique mutation avantageuse est fixée, ou balancée lorsque plusieurs allèles avantageux sont maintenus à des fréquences proches dans la population.

La spécificité entre le gène de résistance et le gène d'avirulence engendre une sélection directionnelle forte sur l'agent pathogène qui favorise les variants porteurs de mutations conduisant à l'absence de reconnaissance de l'effecteur par la protéine R dans le gène d'avirulence (Jones and Dangl, 2006).

Deux scénarios coévolutifs entre l'hôte et l'agent pathogène ont été énoncés : la "guerre des tranchées" et la "course aux armements". Le scénario de la "course aux armements" (Dawkins and Krebs, 1979) décrit une forte pression de sélection qui assure un contournement et un renouvellement rapide des gènes de résistance par l'agent pathogène. Dans le scénario de "guerre des tranchées" (Stahl et al., 1999) les différents allèles à un locus de résistance sont maintenus par sélection balancée. Ils alternent entre un état de forte et de faible fréquence selon la fréquence du gène d'avirulence correspondant. Le scénario diffère selon le pathosystème étudié et est souvent intermédiaire aux scénarios extrêmes proposés ci-dessus. Par ailleurs, dans le cas des résistances au virus, les gènes de résistance sont souvent des gènes nécessaires au fonctionnement de la plante et donc très conservés comme les eIF4E (Charron et al., 2008). La sélection par le virus n'est donc pas la seule qui entre en jeu dans leur sélection ou contre-sélection.

2 Identification des locus de résistance

Plusieurs stratégies sont possibles pour identifier un gène ou un ensemble de gènes responsables d'un caractère phénotypique. Elles se divisent en deux groupes : la cartographie par analyse de liaison et la génétique d'association.

2.1 Cartographie génétique par analyse de liaison

Historiquement, la première stratégie utilisée afin d'identifier le déterminisme génétique d'un caractère a été la cartographie par analyse de liaison.

Que ce déterminisme soit mono ou polygénique, la première étape de ce type d'étude est la construction de la population de cartographie. Les populations utilisées sont issues de croisements, la plupart du temps biparentaux entre des parents polymorphes au niveau du caractère étudié, sur une ou plusieurs générations (génération F₂, populations de lignées recombinantes, lignées d'haploïdes doublés...). Une carte génétique est alors construite : la ségrégation des marqueurs génétiques est mesurée et le nombre d'événements de recombinaison entre deux marqueurs est transformé en distance génétique additive.

Afin de suivre la ségrégation des polymorphismes, plusieurs types de marqueurs ont été utilisés par le passé. Les premiers ont été les *restriction fragment length polymorphism* (RFLP) et les *amplified fragment length polymorphism* (AFLP), toutefois ce type de polymorphismes nécessitait des différences génétiques importantes entre les parents afin de disposer d'un nombre de marqueurs suffisant. Le développement des marqueurs microsatellites a permis de densifier les cartes génétiques. Avec l'avènement du séquençage et des méthodes NGS (*Next-generation sequencing*), les polymorphismes nucléotidiques (SNP) sont de plus en plus utilisés, notamment grâce à leur utilisation possible à très haute densité et de façon automatisée. De plus, le nombre de génomes disponibles augmentant de manière continue, la réalisation de la carte génétique n'est plus une obligation et les distances génétiques sont remplacées par des distances physiques.

La deuxième étape est le phénotypage, plus ou moins fin selon que le caractère est qualitatif ou quantitatif. Les méthodes employées par la suite diffèrent selon la nature du caractère étudié.

2.1.1 Cartographie d'un gène majeur

Afin de cartographier un gène majeur, les corrélations entre marqueurs et phénotype sont étudiées afin d'obtenir des régions du génome candidates. Pour affiner la cartographie du gène, les effectifs sont augmentés et les régions candidates densifiées en marqueurs.

La taille du fragment chromosomique identifié comme contenant le gène dépend du nombre de marqueurs disponibles et du nombre d'événements de recombinaison. La finesse de la cartographie est également liée au potentiel de recombinaison de la région

génique. Lorsqu'on ne trouve plus d'individus recombinants dans la région candidate, des approches gènes candidats sont alors utilisées pour identifier le gène sous-jacent au caractère. Il s'agit notamment de rechercher dans la région des gènes dont la fonction est connue pour être impliquée dans le caractère étudié et de chercher à valider les gènes soupçonnés d'être impliqués dans le caractère par des approches fonctionnelles telles que la transformation génétique.

2.1.2 Cartographie de caractères quantitatifs

Beaucoup de caractères d'intérêt agronomique ont une distribution continue (rendement, qualité, certaines résistances à des agents pathogènes...). La cartographie génétique permet également d'identifier les QTL sous-jacents à ces caractères. Afin d'obtenir une bonne résolution, la taille de la population doit être suffisamment importante. Différentes approches statistiques (régression simple, *interval mapping* ou *composite interval mapping*) peuvent ensuite être utilisées afin de détecter la localisation des QTL (Haley et al., 1992; Thoday, 1961; Zeng, 1993)

Les QTL à effet majeur identifiés peuvent ensuite être affinés par cartographie fine de la même manière qu'un gène majeur.

Bien que la cartographie génétique soit une stratégie intéressante pour identifier des gènes responsables de variations quantitatives, elle souffre de certains inconvénients. Ces études sont réalisées sur des populations biparentales, donc seul l'effet de deux allèles à chaque locus peut être quantifié et la génération des populations de cartographie peut prendre un temps long. Par ailleurs, le nombre d'événements de recombinaison dans ces populations est restreint et les QTL identifiés par cartographie sont donc définis sur

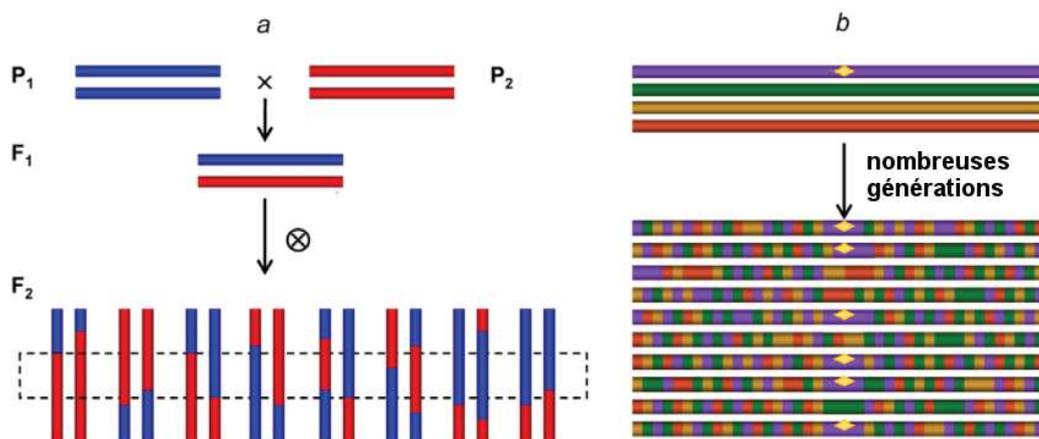


FIGURE 1.7 – Comparaison schématique de l'analyse de liaison réalisée avec une descendance de cartographie (a) et d'une cartographie d'association réalisée avec des collections d'individus non apparentés (b) (publié dans Zhu et al. (2008)). Les individus de la cartographie d'association présentent un nombre de recombinaisons beaucoup plus important que ceux du croisement de cartographie. La population présente une plus grande diversité allélique qu'un croisement biparental. L'analyse d'association permet de repérer plus précisément la région d'intérêt si le déséquilibre de liaison est faible.

des segments génomiques de 10 à 20 cM (Ingvarsson and Street, 2011).

2.2 Génétique d'association

La baisse des coûts de génotypage et le développement de nouvelles techniques telles que les puces SNP ou les NGS a vu l'essor de la génétique d'association. Cette méthode de cartographie permet l'étude de populations d'individus non apparentés. Les recombinaisons observées sont des recombinaisons historiques, bien plus nombreuses que dans une descendance de cartographie (Figure 1.7), ce qui permet de réduire l'intervalle des QTL.

2.2.1 Le principe

Les approches de génétique d'association ont été développées pour la recherche de déterminants génétiques de caractères chez l'humain, où l'étude de larges populations en ségrégation est impossible (Spielman et al., 1993). La méthode reposait à l'origine sur la constitution de deux cohortes au phénotype différent, par exemple sain et malade dans le cas d'une pathologie. Les individus des deux cohortes étaient ensuite génotypés à différents marqueurs et une association statistique entre le génotype et le phénotype était recherchée. Si la méthode a d'abord été appliquée à des caractères contrastés, elle l'est maintenant à des caractères complexes et continus.

En l'absence d'un reséquençage complet des individus, la probabilité de détecter le polymorphisme causal est extrêmement faible. La génétique d'association repose donc sur une relation triangulaire entre le phénotype d'intérêt, le locus causal de la variation observée et un marqueur en déséquilibre de liaison (DL) avec le locus causal (Balding, 2006) (Figure 1.8).

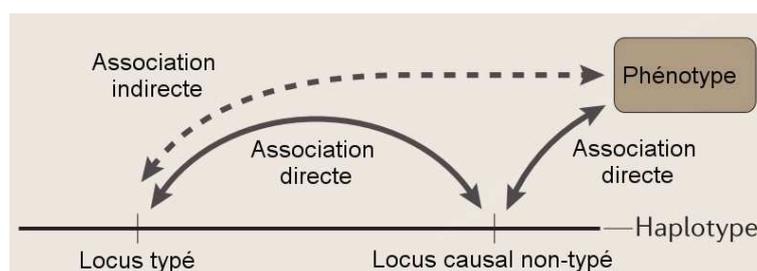


FIGURE 1.8 – Schématisation de la relation triangulaire à la base de la génétique d'association (d'après Balding (2006)). Le locus marqueur typé est directement lié par une relation de DL avec le locus causal du phénotype observé. Ce locus est donc indirectement lié au phénotype.

2.2.2 Le déséquilibre de liaison

Le déséquilibre de liaison (DL) est une propriété des génomes des populations naturelles comme domestiques. Il s'agit de l'association non aléatoire entre les allèles de

différents locus dans une population donnée. L'étendue de ce dernier le long du génome donne la résolution de la cartographie par génétique d'association.

Dans une population à l'équilibre d'Hardy-Weinberg (population panmictique sans migration ni mutation ou sélection), le DL ne dépend que de la distance physique entre deux locus et de la fréquence de recombinaison. Toutefois dans une population réelle plusieurs autres facteurs peuvent influencer sur le DL (Rafalski and Morgante, 2004). La mutation crée de nouveaux allèles à certains locus. La sélection rapide d'un allèle à effet positif crée un balayage sélectif : les allèles présents à d'autres locus dans les individus possédant l'allèle sélectionné augmentent alors en fréquence dans la population, augmentant le DL. La recombinaison est la seule force pouvant faire diminuer le DL, cassant ces associations au fur et à mesure des événements de reproduction. Les espèces à autogamie majoritaire expérimentent une hybridation entre génotypes moins intense que les allogames majoritaires. La recombinaison y est donc moins efficace à l'échelle de la population et le DL est plus élevé.

L'isolement génétique entre familles et la subdivision des populations crée du DL : dans chaque sous-population les fréquences alléliques augmentent de manière indépendante sous l'effet de la sélection et de la dérive génétique. Des blocs d'haplotypes différents apparaissent dans les différentes sous-populations et le DL de la population globale augmente sur l'ensemble du génome. La migration et l'hybridation de ces sous-populations fait diminuer le DL au cours du temps.

De même, dans des populations à faible taille efficace ou ayant subi un goulot d'étranglement récent, les fréquences alléliques peuvent changer rapidement sous l'effet de la dérive génétique. La disparition de certains allèles maintient le DL entre les haplotypes, en nombre limité dans la population.

Dans certaines populations, il peut y avoir du DL entre deux locus situés sur des chromosomes différents. C'est notamment le cas du riz asiatique *O. sativa*, où une très forte structuration existe entre les sous-espèces *indica* et *japonica* (Wang et al., 2014a). Ce patron peut être dû à la structuration de la population ou bien à un phénomène génétique réel (épistasie).

La structure chromosomique joue aussi un rôle sur le DL. Le taux de recombinaison n'est pas le même le long du génome. L'exemple le plus flagrant est la recombinaison très limitée à proximité des centromères.

L'ensemble de ces facteurs crée une variation de l'étendue du DL le long du génome spécifique à chaque population. La connaissance de l'étendue du DL dans le génome au sein de la population étudiée est primordiale pour la réalisation d'études de génétique d'association. Si le DL est très étendu dans la population, il est inutile de disposer d'un nombre très important de marqueurs car leur information sera redondante : plus le DL sera grand, plus la densité de marqueurs nécessaire à une bonne couverture sera faible, et inversement. Cependant, l'intervalle d'un QTL identifié sera également plus large dans une population à fort DL (Figure 1.9).

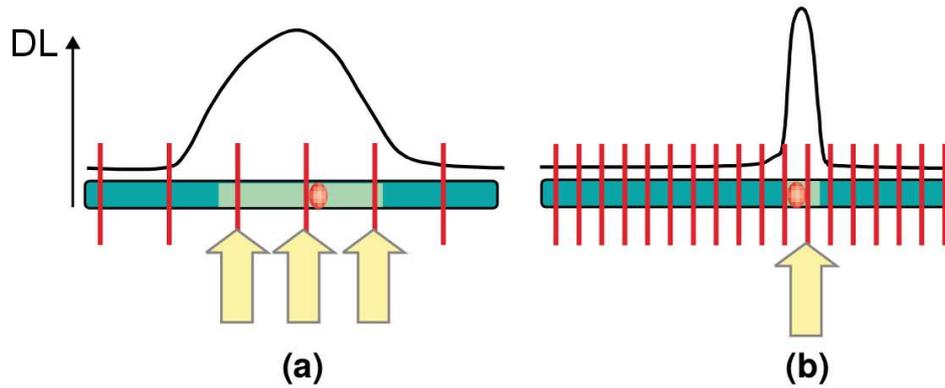


FIGURE 1.9 – Relation entre l'étendue du DL et la résolution des études d'association (publié dans Rafalski (2002)). En (a) le DL avec le gène responsable du phénotype (ovale rouge) décline doucement avec la distance. Une faible densité de marqueurs (barres rouges) est suffisante pour identifier des marqueurs associés au phénotype (flèches jaunes). En (b), le DL est très peu étendu, il faut donc une grande densité de marqueurs afin d'en identifier un associé.

Plusieurs indices permettent de calculer le DL entre une paire de marqueurs (Flint-Garcia et al., 2003; Gupta et al., 2005). Les plus fréquemment utilisés sont r^2 (Hill and Robertson, 1968) et D' (Lewontin, 1964). Ces deux statistiques n'expliquent pas les mêmes aspects du DL (Figure 1.10).

On considère deux locus bialléliques A/a et B/b, dont les fréquences alléliques sont notées π_A , π_a , π_B et π_b et les fréquences haplotypiques π_{AB} , π_{Ab} , π_{aB} et π_{ab} . À l'équilibre les fréquences haplotypiques attendues sont égales au produit des fréquences alléliques correspondantes. Le calcul des estimateurs D' et r^2 repose sur la différence entre les fréquences haplotypiques observées et attendues, notée D_{AB} :

$$D_{AB} = \pi_{AB} - \pi_A\pi_B.$$

D mesure l'écart absolu entre les fréquence et sa valeur dépend des fréquences alléliques. D' correspond à la mesure de D normalisé par rapport à sa valeur maximale théorique :

$$D' = \frac{|D_{AB}|}{\min(\pi_A\pi_b, \pi_a\pi_B)} \quad \text{si } D_{AB} < 0;$$

$$D' = \frac{|D_{AB}|}{\min(\pi_A\pi_B, \pi_a\pi_b)} \quad \text{si } D_{AB} > 0.$$

D' varie entre 0 et 1, permettant les comparaisons entre paires de locus. Il ne mesure cependant que l'histoire des recombinaisons.

r^2 mesure également celle des mutations. Il est calculé de la manière suivante :

$$r^2 = \frac{(D_{AB})^2}{\pi_A\pi_a\pi_B\pi_b}.$$

D' est fortement affecté par des fréquences alléliques faibles et r^2 est souvent préféré pour l'étude du DL en vue des études de génétique d'association.

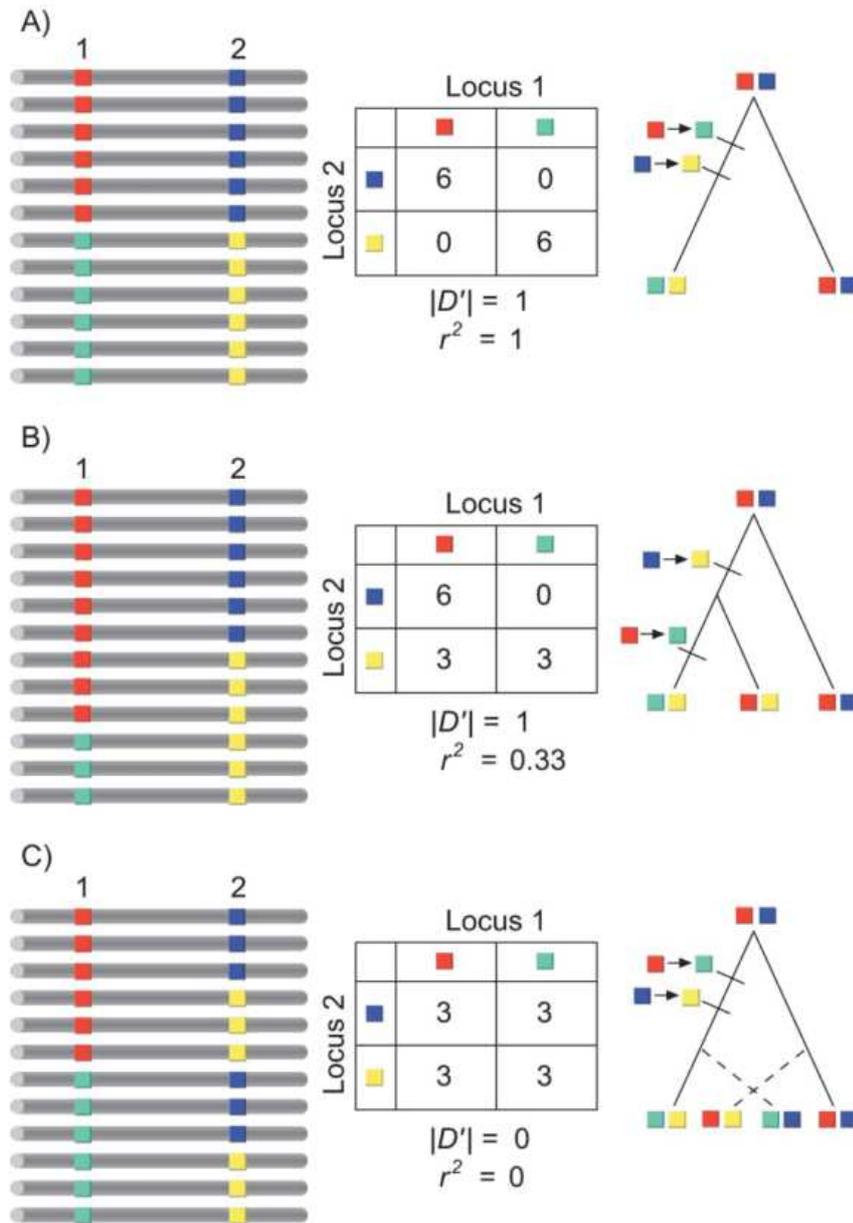


FIGURE 1.10 – Différence de comportement des statistiques D' et r^2 de mesure du DL selon les scénarios évolutifs (publié dans Flint-Garcia et al. (2003)). D' et r^2 sont calculés sur trois modèles hypothétiques de DL entre des polymorphismes liés causés par des événements de mutation et de recombinaison différents. Les images de gauche représentent les états alléliques aux locus dans la population. La colonne du milieu représente la matrice de corrélation des haplotypes et les calculs de D' et r^2 associés. La colonne de droite représente un arbre possible des événements évolutifs responsables du DL. **(A)** Les deux locus sont en DL complet. Ils partagent une même histoire évolutive. D' et r^2 sont égaux à 1. **(B)** DL résultant de mutations apparues dans différentes lignées sans événement de recombinaison. Les calculs de D' et r^2 montrent une grande différence. **(C)** Les deux locus ne sont pas en déséquilibre. Il y a eu des événements de recombinaison entre les locus, quel que soit l'historique des mutations. D' et r^2 sont égaux à 0.

2.2.3 Les études de génétique d'association sur génome entier

Les études de génétique d'association sur génome entier (GWAS) nécessitent :

- une population choisie pour maximiser la diversité génétique et phénotypique aux traits d'intérêts (souvent une *core collection*),
- le phénotype des individus la constituant,
- un nombre de marqueurs suffisamment grand pour couvrir le génome.

Plutôt que des marqueurs, il est de plus en plus fréquent d'utiliser des données de séquences des individus, souvent obtenues par des techniques de *next-generation sequencing* (NGS) à faible profondeur (génotypage par séquençage, GBS).

Un modèle statistique est ensuite appliqué pour tester l'association entre les locus et les valeurs phénotypiques. Le modèle le plus simple est le modèle linéaire global (GLM) pour lequel chaque SNP a un effet fixe linéaire sur le phénotype. Ce modèle très simple ne prend pas en compte l'effet de la structuration de la population sur la structure du DL, qui peut causer des associations entre le polymorphisme et des SNP non physiquement liés aux SNP causaux.

Des modèles plus complexes, tels que le modèle linéaire mixte (MLM), introduisent l'effet de la structure de la population comme effet fixe et celui de l'apparentement entre individus comme effet aléatoire dans le modèle (Yu et al., 2006). L'assignation des individus aux sous-populations peut se faire à l'aide de modèles cherchant à minimiser l'écart à l'équilibre d'Hardy-Weinberg et le DL entre locus au sein des sous-populations (STRUCTURE (Pritchard et al., 2000), sNMF (Frichot et al., 2014)...) ou bien en réalisant une analyse en composantes principales (ACP) (Price et al., 2010). L'intégration de la structure et de l'apparentement dans les modèles permet de diminuer le nombre de faux positifs. Toutefois cela peut engendrer une perte de puissance pour la détection des vrais positifs.

Afin de choisir le modèle le plus adapté aux données, la distribution des tests statistiques observés générés par le GWAS doit être vérifiée. Sous le modèle nul, la distribution des p -values est attendue uniforme. La déviation observée sur un quantile-quantile (Q-Q) *plot* doit ne pas présenter une trop grande inflation des p -values par rapport au modèle nul (Figure 1.11).

Les SNP associés se détachent des autres sur le *Q-Q plot* et forment un pic sur le *Manhattan plot* représentant la distribution des effets des SNP calculés par le modèle GWAS le long du génome. Plusieurs seuils de significativité peuvent être appliqués. Le plus classique est calculé par le critère de Bonferroni selon la formule :

$$S = \frac{\alpha}{N}$$

où α est le seuil de significativité global souhaité (la plupart du temps fixé à 5 %) et N le nombre de tests réalisés, fixé ici au nombre de SNP. Ce critère est toutefois jugé trop conservatif, notamment car il considère que tous les SNP sont indépendants alors

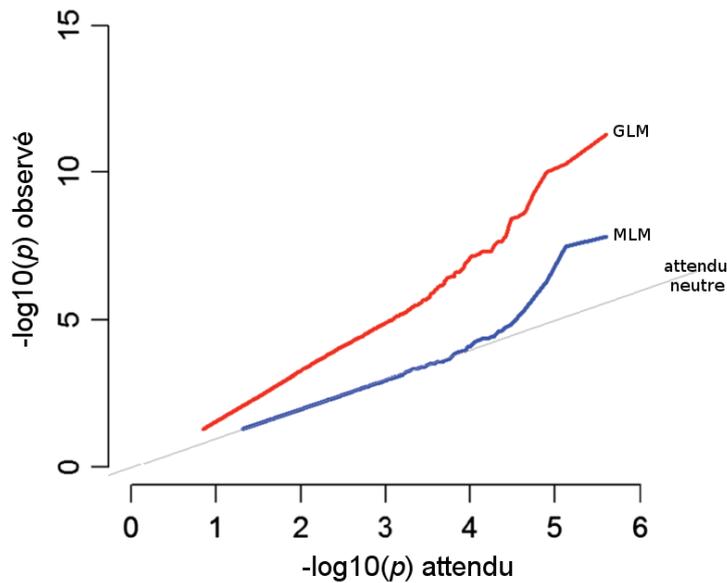


FIGURE 1.11 – Comparaison des $Q-Q$ plot obtenus pour deux modèles de GWAS exécutés sur les mêmes données (publié dans Korte and Farlow (2013)). Le modèle GLM montre une forte inflation des p -values (en rouge) comparé à l’attendu neutre (ligne grise). L’utilisation d’un modèle mixte permet de fortement réduire cette inflation (en bleu). Seuls les marqueurs les plus associés au caractère montrent une déviation par rapport à l’attendu neutre.

que nombre d’entre eux sont en DL. Plusieurs méthodes comme simpleM (Gao et al., 2008) ou K_{eff} (Moskvina and Schmidt, 2008) évaluent le nombre de tests effectivement indépendants pour corriger le critère de Bonferroni.

L’utilisation de permutations est également vue comme une méthode fiable d’évaluation de significativité (Churchill and Doerge, 1994). Les phénotypes sont permutés alors que la structure génétique demeure. Ces permutations permettent de simuler l’hypothèse nulle de non-association. Toutefois un grand nombre de permutations est nécessaire afin d’obtenir des résultats fiables et le temps de calcul est long.

Enfin, il est possible de calculer le *false discovery rate* (FDR) qui évalue la proportion de faux positifs parmi les tests déclarés positifs (Dolejsi et al., 2014; Yoav Benjamini, 1995). Le seuil obtenu est la plupart du temps bien moins conservatif que celui du critère de Bonferroni.

2.3 Utilisation de populations multi-parentales

Les populations multi-parentales combinent les aspects de la cartographie génétique et de la génétique d’association. Plusieurs types de populations sont utilisés. Les principales sont les populations de *nested association mapping* (NAM) et les *multi-parent advanced generation inter-cross* (MAGIC) (Figure 1.12). Les NAM, mis au point par Yu et al. (2008) sur maïs, sont créées en mélangeant en proportions égales les différentes *recombinant-inbred line* (RIL) impliquant un parent récurrent. Les MAGIC (Cavanagh et al., 2008) sont des populations issues d’un nombre de parents important

(huit au minimum) interfécondés afin que la contribution de chaque parent au génome des descendants soit la même. Les descendants obtenus sont interfécondés un grand nombre de fois afin d'obtenir des RIL.

Ce type de population permet de combiner les événements de recombinaisons anciens des populations naturelles et les événements récents des populations de cartographie. L'effet de la structure des populations est ainsi limitée tout en augmentant le nombre de recombinaisons dans le génome. Toutefois certains désavantages de la cartographie génétique subsistent (Mitchell-Olds, 2010) :

- le nombre de fondateurs reste limité et ne représente donc pas l'ensemble de la diversité allélique qu'on peut obtenir dans une population de génétique d'association classique ;
- les croisements nécessaires à l'obtention de ces populations nécessitent un nombre de générations important et donc un temps long.

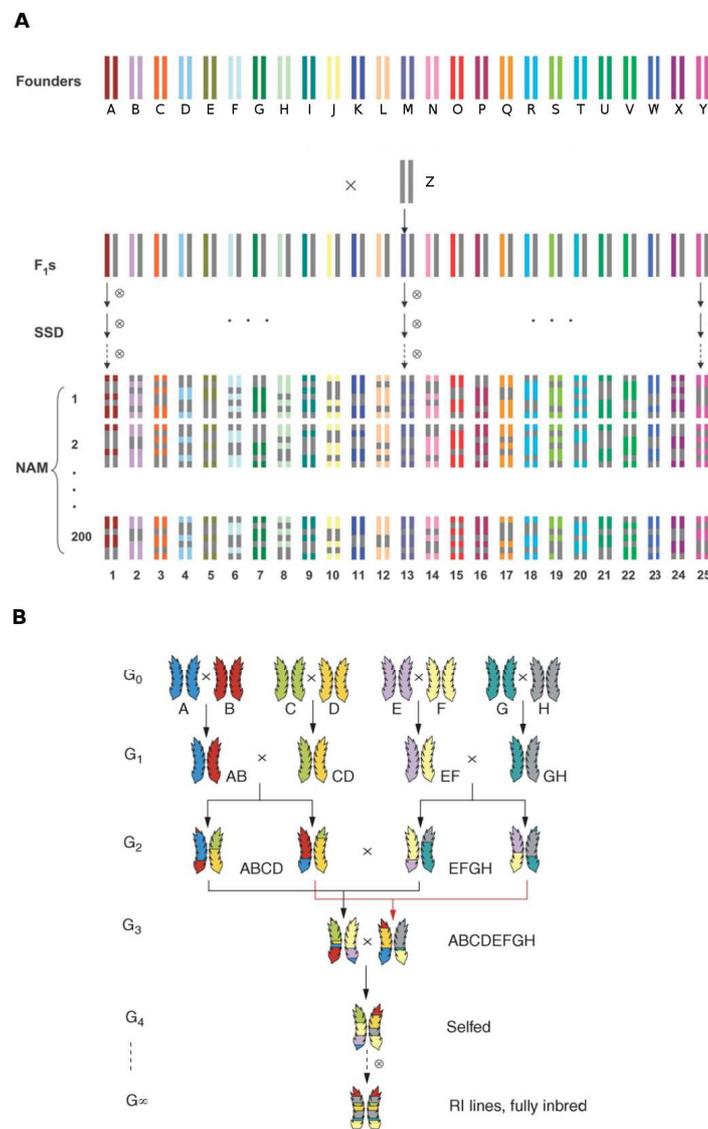


FIGURE 1.12 – Exemple de populations multi-parentales (adapté de Cavanagh et al. (2008) et Yu et al. (2008)). (A) : population NAM; (B) : population MAGIC

2.4 Autres méthodes

La cartographie par analyse de liaison et la génétique d'association sont des applications de génétique classique ou *forward genetics* : la recherche de la variation génétique est basée sur la variation phénotypique observée dans la variabilité existante. À l'ère de la génomique, les approches de génétique inverse ou *reverse genetics* se développent. Il s'agit de rechercher l'effet d'un gène en observant le phénotype de variants à ce gène. C'est notamment le cas des approches de type *targeting induced local lesions in genomes* (TILLING) (McCallum et al., 2000). Elles consistent en une étape de mutagenèse classique suivie du screening des mutations obtenues pour trouver celles présentes dans le gène d'intérêt. Cela peut également permettre de créer de nouveaux allèles à un gène et ainsi d'étudier le rôle des différents domaines d'un gène dans le phénotype.

Ces méthodes sont souvent combinées à la génétique classique. Une région candidate contenant plusieurs gènes est identifiée par génétique classique puis le gène causal du phénotype peut être validé par génétique inverse.

L'étude de systèmes hétérologues peut également permettre d'étudier efficacement la génétique gouvernant un caractère. La levure est notamment présentée comme un système intéressant pour étudier les interactions des plantes avec certains virus (Nagy et al., 2014). En effet, la levure possède un génome plus simple que les plantes avec un nombre de séquences intergéniques et d'introns bien inférieur. Les ressources génétiques et génomiques disponibles pour cet organisme simple sont également très importantes : un grand nombre de séquences, des banques de mutants... De plus, cet organisme a beaucoup de similarités fonctionnelles avec une cellule végétale et certains virus de plantes peuvent donc s'y répliquer. Cette répllication a été démontrée notamment pour le *Brome mosaic virus* (BMV) et le *Tomato bushy stunt virus* (TBSV). L'étude de la répllication en levure permet d'identifier des gènes impliqués. Leurs homologues dans la plante hôte sont donc candidats pour jouer un rôle dans leur répllication, et donc potentiellement permettre la résistance s'ils sont mutés.

3 Le pathosystème

3.1 Le riz

3.1.1 Le genre *Oryza*

Les riz sauvages et cultivés sont des monocotylédones de la famille de *Poaceae*. Les 12 000 espèces de la famille sont divisées en deux clades : PACMAD (*Panicoideae*, *Aristidoideae*, *Chloridoideae*, *Micrairoideae*, *Pooideae*) qui comprend notamment le sorgho et le maïs, et BOP (*Bambusoideae*, *Oryzoideae*, *Pooideae*) qui contient le riz et le blé (Figure 1.13). Le riz est hiérarchiquement classé dans la sous-famille des *Oryzoideae*, la tribu des *Oryzaceae* et la sous-tribu des *Oryzinae* (Soreng et al., 2015). Il appartient au genre *Oryza* qui comprend 23 espèces (Vaughan et al., 2003).

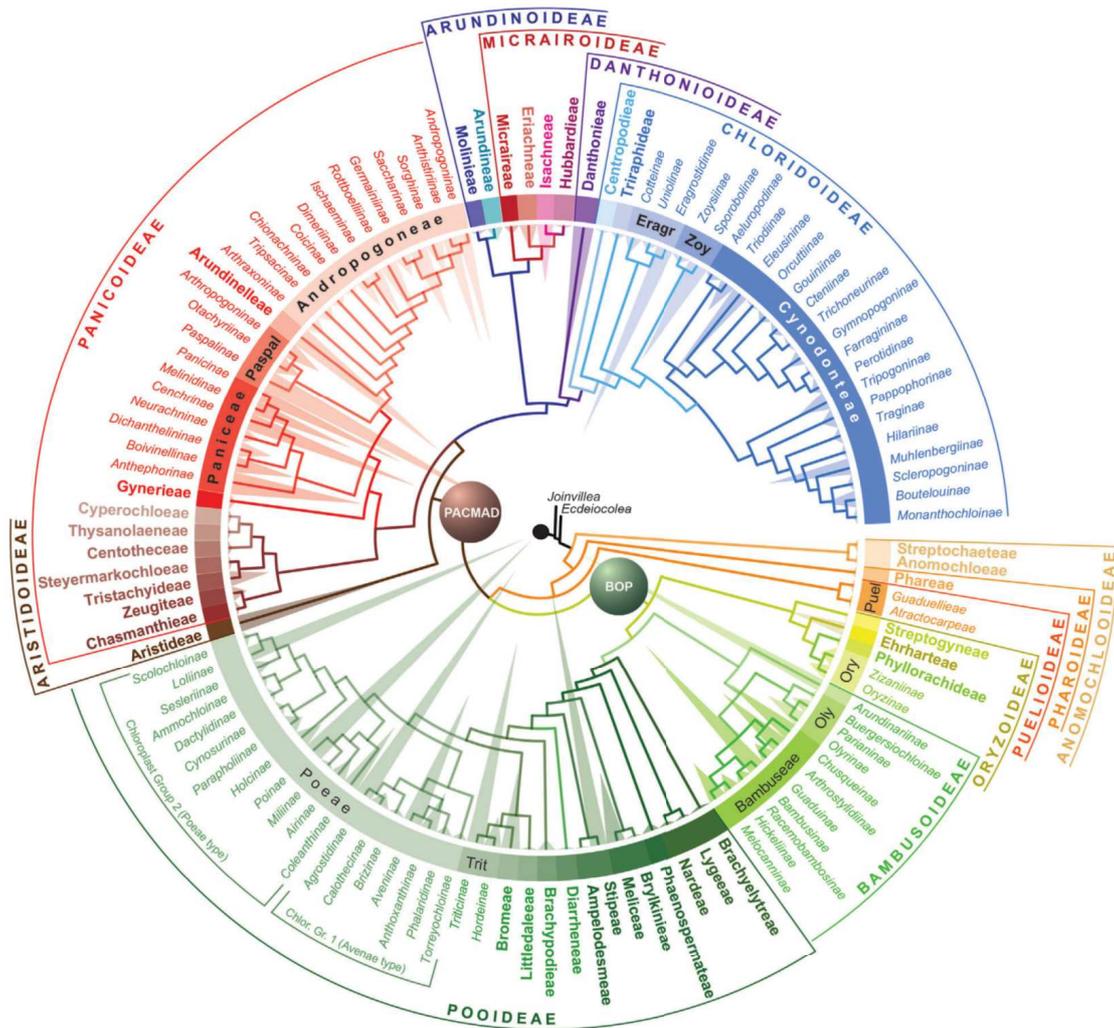


FIGURE 1.13 – Classification phylogénétique des *Poaceae* (publiée par Soreng et al. (2015)). Cette famille comprend 12 sous-familles, 51 tribus et 80 sous-tribus. La hauteur des triangles est proportionnelle à la taille du taxon. BOP : *Bambusoideae*, *Oryzoideae* et *Pooideae*. PACMAD : *Panicoideae*, *Aristidoideae*, *Chloridoideae*, *Micrairoideae*, *Pooideae*.

Parmi ces 23 espèces, on retrouve 9 types génomiques différents : 6 diploïdes (AA, BB, CC, DD, EE, FF et GG), et 3 allotétraploïdes (BBCC, CCDD, HHJJ). L'existence d'un quatrième type allotétraploïde HHKK est discutée (Vaughan et al., 2003). Le genre est divisé en quatre complexes d'espèces définis sur la base de critères morphologiques et génomiques (ploïdie et homologies) : *O. sativa*, *O. officinalis*, *O. ridelyi* et *O. granulata*, ayant tous $n=12$ chromosomes.

Le complexe *O. sativa* est composé de 8 espèces de génome AA réparties sur l'ensemble du globe (Figure 1.14). Il comprend les deux espèces cultivées (*O. sativa* et *O. glaberrima*) toutes deux annuelles, trois espèces sauvages annuelles (*O. barthii*, *O. nivara* et *O. meridionalis*) et trois espèces sauvages pérennes (*O. rufipogon*, *O. longistaminata* et *O. glumaepatula*) (Vaughan et al., 2008). Longtemps considérés comme deux espèces différentes, de plus en plus d'auteurs décrivent maintenant *O. rufipogon* et *O. nivara* comme deux écotypes d'une même espèce (Zhu and Ge, 2005).

3.1.2 Les espèces cultivées

Les deux espèces cultivées, *O. sativa* et *O. glaberrima* se distinguent par plusieurs caractères morphologiques (Linares, 2002) :

- un péricarpe souvent rouge chez *O. glaberrima*,
- une ligule courte et tronquée chez *O. glaberrima*, large chez *O. sativa*,
- l'absence de pilosité des glumelles chez *O. glaberrima*,
- une ramification secondaire des panicules plus faible chez *O. glaberrima*,
- un port plus retombant de la panicule chez *O. sativa*,
- les grains d'*O. glaberrima* sont plus petits et ont une plus forte déhiscence.

Les deux espèces sont séparées par une importante barrière de stérilité (Bougerol and Pham, 1989; Sano, 1985, 1990). Plusieurs locus sont impliqués dans cette barrière de stérilité (Garavito et al., 2010) dont le plus important est le locus *S1*. (Lorieux et al., 2000)

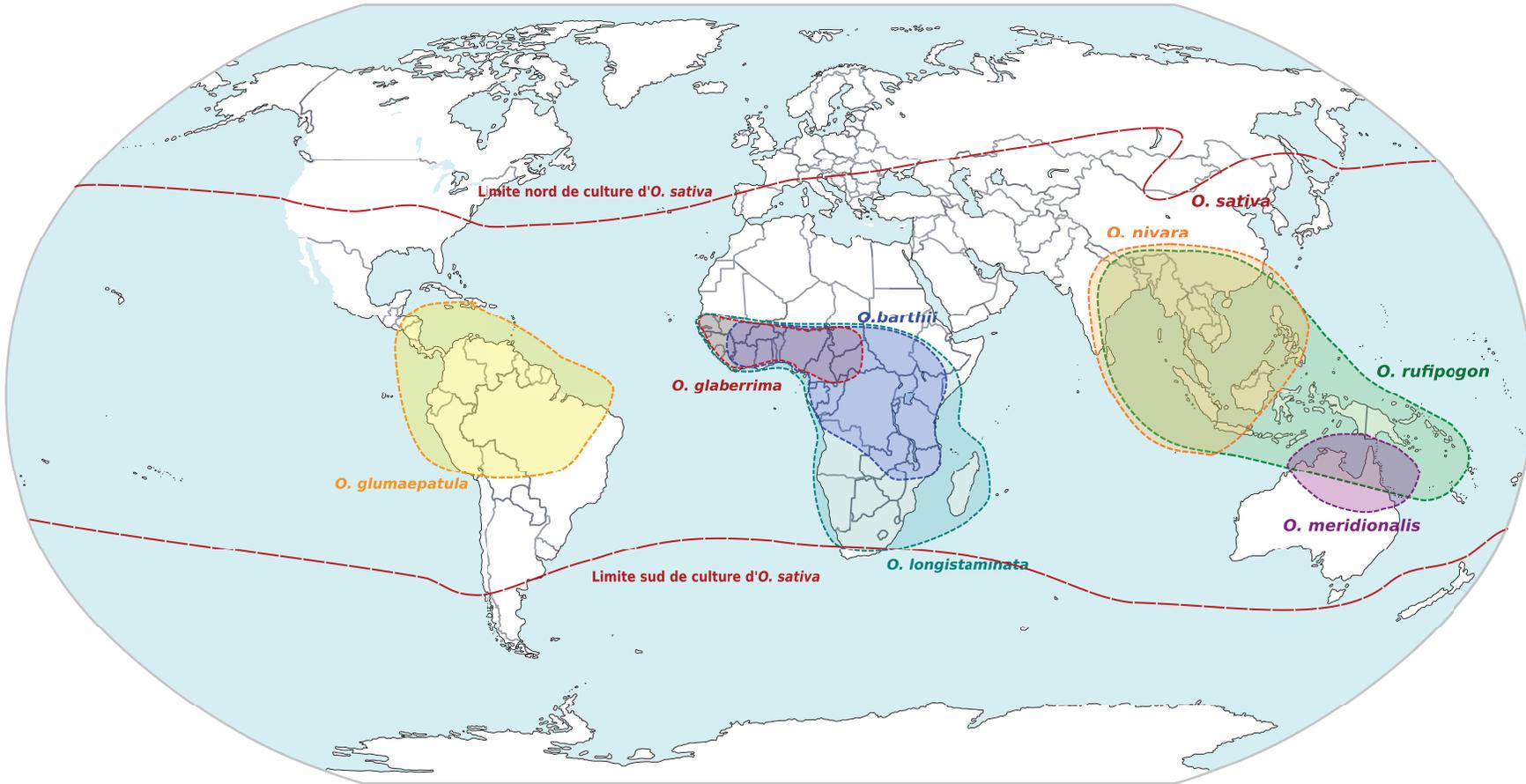


FIGURE 1.14 – Distribution mondiale des espèces du génome AA du genre *Oryza* (adapté Ashikari and Matsuoka (2006)).

3.1.2.1 *O. sativa*

L'espèce *O. sativa* a été domestiquée en Asie. Elle est apparentée à l'espèce sauvage *O. rufipogon*. Sa domestication est datée d'il y a 5 à 10 000 ans. C'est l'espèce de riz la plus cultivée mondialement en raison de rendements importants, de sa plasticité écologique et de ses qualités gustatives. L'espèce présente une très grande diversité de formes et comprend plus de 10 000 cultivars.

O. sativa est divisée en deux sous-espèces et six sous-groupes (Garris, 2005; Glaszmann, 1987; Wang et al., 2014a) :

- la sous-espèce *indica*, originaire d'Inde, est adaptée aux régions tropicales de basse altitude (inférieure à 1 200 m). Elle est caractérisée par un fort tallage, des feuilles étroites, des racines fines et un grain court et effilé. Le groupe *aus* est rattaché aux *indica*. Il est caractérisé par sa précocité et une résistance à la sécheresse.
- la sous-espèce *japonica* se divise en deux sous-groupes. Les *japonica* tempérés sont adaptés aux régions tempérées et ont un tallage moyen, des feuilles et racines fines et un grain court et arrondi. Les *japonica* tropicaux sont adaptés aux régions subtropicales d'altitude. Leur tallage est réduit, leurs feuilles larges, et leurs racines épaisses et profondes. Le grain est long et large. À cette sous-espèce se rattachent deux groupes mineurs : les *aromatiques*, ayant un arôme plus développé, et les *rayada*, riz d'hiver au développement particulièrement lent.

L'origine simple ou multiple de la domestication de l'espèce a longtemps été discutée (Sweeney and McCouch, 2007). Au vu des dernières études il semblerait que l'espèce proviennent de trois domestications différentes (Civáň et al., 2015). Les sous-groupes *indica* et *aus* et la sous-espèce *japonica* proviendraient ainsi d'événements de domestication indépendants. Par ailleurs, il semblerait que le sous-groupe *aromatique* soit un hybride entre *aus* et *japonica*.

O. sativa est une espèce modèle pour les monocotylédones. Plusieurs génomes ont été séquencés et assemblés. Celui de la variété *japonica* Nipponbare est le premier génome de référence de l'espèce (Goff et al., 2002; International Rice Genome Sequencing Project, 2005; Kawahara et al., 2013). Il fait environ 430 Mb, est diploïde et contient peu de séquences répétées. D'autres génomes sont disponibles, dit *draft* car de qualité inférieure à celui de Nipponbare. Il s'agit de ceux des variétés *indica* 93-11 (Yu et al., 2002), IR64 (Schatz et al., 2014), HR12 (Mahesh et al., 2016) et Kasalath (Sakai et al., 2014), de la variété *aus* DJ123 (Schatz et al., 2014) et de la variété *japonica* Koshihikari (Kawahara et al., 2013). Depuis peu, les 3 000 génomes d'*O. sativa* sont disponibles (Alexandrov et al., 2015; Li et al., 2014).

3.1.2.2 *O. glaberrima*

L'espèce *Oryza glaberrima*, apparentée au riz sauvage *O. barthii*, a été domestiquée de manière indépendante d'*O. sativa* il y a 3 500 ans dans le delta du Niger (Murray,

2004; Portères, 1962; Zhu and Ge, 2005). Sa culture est limitée à l’Afrique de l’Ouest, notamment à cause de ses faibles rendements. *O. glaberrima* présente cependant des avantages agronomiques dans un contexte de culture africain. Cette espèce est plus rustique et présente des résistances à des stress biotiques et abiotiques. Ainsi, il présente des résistances à divers agents pathogènes tels que la panachure jaune du riz (RYMV), la pyriculariose (*Magnaporthe grisea*), la bactériose vasculaire (*Xanthomonas oryzae*) ou encore certains nématodes endoparasitaires (*Heterodera sacchari*, *Meloidogyne graminicola*, *M. incognita*). Il est également plus tolérant à la sécheresse, la salinité, la toxicité ferrique ou encore la submersion (Linares, 2002).

La diversité génétique d’*O. glaberrima* est particulièrement faible (Nabholz et al., 2014). En effet l’espèce a subi un double goulot d’étranglement lors de son histoire : le premier associé à la divergence des ancêtres des riz Asiatiques en direction de l’Afrique, et le deuxième associé à la domestication Africaine (Vaughan et al., 2008). Une étude de diversité sur la collection d’*O. glaberrima* disponible à l’IRD a révélé deux groupes distincts au niveau génétique (Orjuela et al., 2014). Ces deux groupes présentent une divergence faible et ne répondent pas à une répartition géographique ou à un phénotype particulier. Malgré la faible diversité génétique, on trouve dans l’espèce une diversité de forme adaptée à diverses conditions, notamment des variétés flottantes et des variétés pluviales. Cette diversité morphologique reste cependant inférieure à celle d’*O. sativa*.

Un génome de référence, réalisé sur l’accession CG14, a été publié en 2014 (Wang et al., 2014c). Afin d’assembler le génome, les extrémités des clones BAC ont été *mappées* sur le génome de Nipponbare, occasionnant la perte de certaines régions spécifiques d’*O. glaberrima*. Un assemblage *de novo* des accessions CG14 et Tog5681 a été réalisé à l’IRD mais n’est pas encore publié. Le projet IRIGIN (France génomique) a également permis d’obtenir les données de séquences de 163 accessions d’*O. glaberrima* et 84 accessions d’*O. barthii*.

Ces accessions séquencées proviennent de la collection IRD composée de 278 accessions d’*O. glaberrima* et de 103 accessions *O. barthii*. Les accessions constituant la collection *O. glaberrima* ont été choisies conjointement par l’IRD et l’AfricaRice. Cette collection rassemble notamment des accessions de populations références, et des accessions ayant des caractères intéressants de résistance à des stress biotiques et abiotiques. Les accessions choisies représentent au mieux la distribution géographique de l’espèce en Afrique de l’Ouest avant 1950. Par ailleurs, un effort particulier a été fait pour rassembler des accessions *O. barthii* réellement sauvages, représentant la diversité actuelle.

3.1.2.3 Importance des deux espèces en Afrique

Plusieurs phases d’introduction d’*O. sativa* ont eu lieu en Afrique de l’Est, tout d’abord en Tanzanie entre le VI^e et le X^e siècle, puis à Madagascar entre le VI^e et le XVI^e siècle (Linares, 2002; Portères, 1962; Vaughan et al., 2008). Dès le XVI^e siècle, les navigateurs portugais, hollandais et français ont introduit le riz asiatique en Afrique de

l'Ouest. La large gamme variétale a rapidement été adoptée et a permis d'intensifier la culture de riz dans cette région.

En 200 ans, la culture d'*O. sativa* s'est étendue à l'ensemble des pays africains producteurs de riz où il a peu à peu remplacé *O. glaberrima*, moins productif. Toutefois *O. glaberrima* reste cultivé localement en Afrique de l'Ouest pour des raisons culturelles, culinaires et agronomiques (Linares, 2002).

Depuis les années 90, un programme de l'AfricaRice a visé à combiner les avantages des deux espèces en créant des variétés hybrides, appelées NERICA (*NEw RICEs for Africa*) (Jones et al., 1997; Ndjiondjop et al., 2008; Semagn et al., 2007). Les deux espèces sont séparées par une barrière de stérilité importante (Ghesquiere et al., 1997) ayant rendu nécessaires des approches telles que le sauvetage d'embryons, les rétrocroisements ou la culture d'anthères afin d'obtenir ces lignées. Une vingtaine de lignées NERICA a ainsi été développée et est maintenant diffusée en Afrique subsaharienne. Les variétés de NERICA les plus populaires combinent les meilleurs traits des deux parents : un rendement élevé et une capacité à se développer dans des environnements difficiles, elles sont donc mieux adaptées aux systèmes de production de la région (Ndjiondjop et al., 2008). Les NERICA possèdent un plus grand nombre de tiges, des panicules plus grandes et surtout des tiges plus fortes capables de supporter plus de panicules (Sarla and Swamy, 2005). Les grains arrivent à maturité plus rapidement ce qui permet non seulement une double récolte mais également de faire une rotation avec une autre culture et ainsi d'enrichir le sol en azote (Sarla and Swamy, 2005). Le développement de variétés NERICA spécifiquement adaptées aux types de cultures locales a conduit à une augmentation du rendement allant jusqu'à 50 % (Sarla and Swamy, 2005).

3.1.3 Importance économique du riz en Afrique

Le riz est une céréale cultivée dans les régions tropicales, subtropicales et tempérées chaudes. Avec une production mondiale de 740 millions de tonnes en 2014, le riz occupe le deuxième rang de la production céréalière mondiale, derrière le maïs (1 milliard de tonnes) (FaoStats; <http://faostat3.fao.org/>). Sa culture est d'une importance socio-économique majeure, en particulier en Asie et en Afrique.

La majorité de la production se situe en Asie. Le continent produit à lui seul 90 % du riz. Le principal pays producteur est la Chine, à l'origine de 28 % de la production mondiale en 2014, suivi par l'Inde (21 %). La production de riz en Afrique représente à peine 4 % de la production mondiale (FaoStats; <http://faostat3.fao.org/>).

Suite aux changements d'habitudes alimentaires et à l'évolution démographique, en particulier la forte augmentation de la population urbaine, le riz est devenu en Afrique l'une des principales sources alimentaires devant des cultures plus traditionnelles comme le sorgho et le mil. Pour répondre à la demande croissante, la riziculture s'est développée depuis 50 ans, permettant d'atteindre une production six fois plus importante (Figure 1.15). L'essentiel de la production africaine se situe en Afrique de l'Ouest, où la

production continue à progresser.

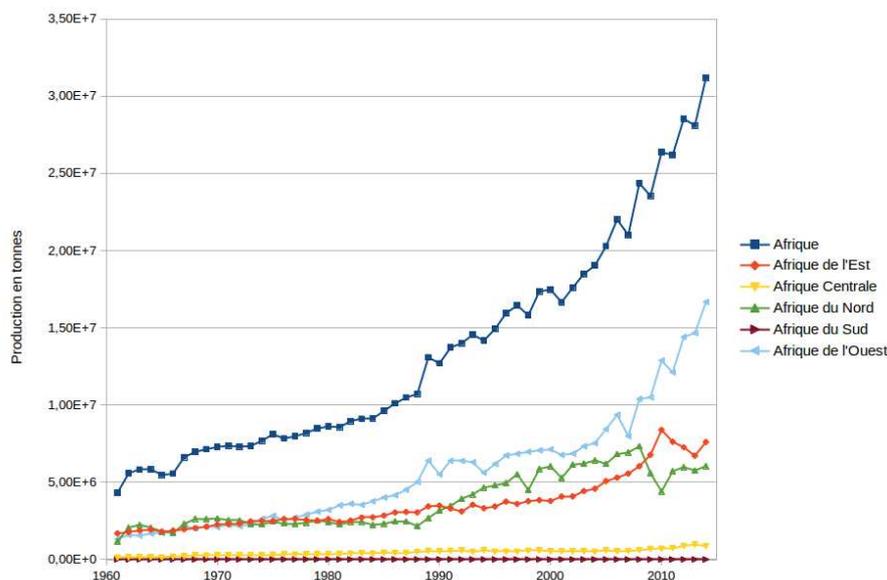


FIGURE 1.15 – Évolution de la production de riz en Afrique.
Source : FaoStats (<http://faostat3.fao.org/>)

La production locale reste cependant insuffisante pour combler les besoins grandissants de la population. Les rendements sont les plus faibles du monde à cause de la faible capacité d'irrigation, des pertes post-récolte et des contraintes biotiques et abiotiques fortes s'exerçant sur les cultures.

L'importation compense partiellement cette insuffisance de production. En 2014, 32 % du riz consommé en Afrique provenait de l'importation. Dans certains pays comme le Burkina Faso, le Sénégal, le Cameroun et le Mozambique, plus de la moitié du riz consommé est importée. Pour certains pays, comme le Tchad, le taux très faible d'importation ne reflète pas une production suffisante mais une insuffisance de financements pour l'importation de riz, malgré le très net besoin des populations. L'augmentation de la production est donc nécessaire afin de répondre aux besoins futurs.

3.1.4 Contraintes à la production en Afrique

Depuis 50 ans, les avancées technologiques et la mise en place de variétés plus productives a permis une augmentation importante de la production en Afrique. Cependant l'augmentation de la productivité du riz est freinée par des obstacles difficiles à surmonter tels que la salinisation, l'érosion, l'épuisement en éléments nutritifs (azote, phosphate, potassium, magnésium), la toxicité ferrique ou la sécheresse. Ces stress abiotiques conduisent à des pertes de rendement et peuvent rendre un sol non-cultivable, entraînant la réduction des surfaces cultivables.

S'ajoutent à cela la pression des déprédateurs (oiseaux, rongeurs...) et de nombreux stress biotiques :

- les insectes parasitaires comme *Chilo zacconius*, *Maliarpha separatella* et *Orseolia oryzivora*,
- des nématodes endoparasitaires dont *Heterodera sacchari*, *Meloidogyne graminicola* et *M. incognita*,
- la pression des plantes adventices et de plantes parasitaires telles que le *Striga*,
- les maladies fongiques telle que la pyriculariose causée par *Magnaporthe oryzae*,
- les maladies bactériennes dont la bactériose à stries foliaires et la bactériose vasculaire causées par *Xanthomonas oryzae*,
- les maladies virales dont la plus importante en Afrique est la panachure jaune du riz causée par le *Rice yellow mottle virus* (RYMV), décrit dans la partie 4.2. D'autres virus du riz sont présents en Afrique comme le *Rice stripe necrosis virus* (RSNV).

Les pertes post-récolte sont également très importantes et menacent la sécurité alimentaire.

3.2 La panachure jaune du riz

On dénombre une quinzaine de maladies virales du riz dans le monde. Elles appartiennent à cinq familles distinctes et ont une répartition majoritairement limitée à un continent (Siré and Brugidou, 2002). La plupart d'entre elles sont présentes en Asie.

Parmi les maladies virales les plus importantes, on peut citer le tungro en Asie, causé par deux virus distincts : le *Rice tungro bacilliform virus* (RTBV) et le RTSV appartenant respectivement à la famille des *Caulimoviridae* et à celle des *Sequiviridae*. La vection est assurée par des insectes de la famille des *Cicadellidae* et les symptômes principaux sont une décoloration jaune-orange des feuilles et une réduction du tallage. Le principal virus du riz en Amérique latine est le *Rice hoja blanca virus* (RHBV) de la famille des *Sequiviridae*. Il est également transmis par un insecte vecteur, *Sogatodes orizicola*, et ses symptômes sont une réduction de croissance, des panachures chlorotiques et des nécroses.

En Afrique, le principal virus du riz est le RYMV, responsable de la panachure jaune du riz.

3.2.1 La maladie

3.2.1.1 Apparition et extension

La panachure jaune du riz est une maladie endémique à l'Afrique et à Madagascar (Kouassi et al., 2005). Elle a été décrite pour la première fois en 1966 au Kenya, près du lac Victoria (Bakker, 1974). Elle a ensuite été mise en évidence dans la plupart des pays africains producteurs de riz (Figure 1.16).

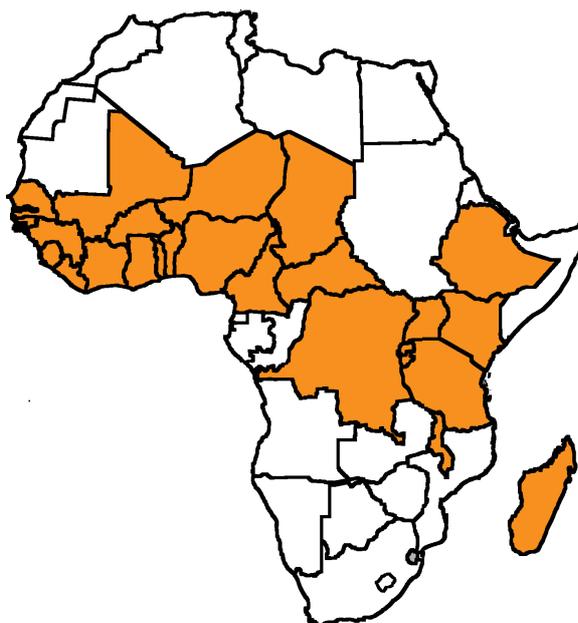


FIGURE 1.16 – Distribution de la panachure jaune du riz en Afrique
Les pays où le virus a été isolé apparaissent en orange.

3.2.1.2 Biologie de la maladie

La panachure jaune du riz tient son nom du jaunissement des feuilles qu'elle provoque. Les premiers symptômes sont des points chlorotiques qui apparaissent sur les feuilles, puis se rejoignent pour former des stries. À un stade avancé, la maladie provoque d'importantes nécroses et le dessèchement des feuilles, pouvant aller jusqu'à la mort de la plante chez les génotypes particulièrement sensibles (Kouassi et al., 2005).

Dans la plupart des cas, la maladie provoque un rabougrissement des plantes et une diminution du nombre de thalles. Elle altère également la formation des fleurs et des gaines, provoquant l'émergence partielle des panicules, la stérilité des épillets et le mauvais remplissage des grains, ce qui affecte fortement la productivité des plants de riz (Fauquet and Thouvenel, 1977).

Au champ, l'infection commence en bordure de parcelle ou en spots dispersés qui s'étendent en îlots jaunes circulaires. Avec l'augmentation de la sévérité de la maladie, les plages fusionnent et la coloration jaune s'étend au champ entier et peut virer au grisâtre, lui donnant un aspect brûlé.

3.2.1.3 Epidémiologie

Le virus de la panachure jaune du riz ou *Rice yellow mottle virus* (RYMV) a une gamme d'hôtes étroite. Il infecte principalement les deux espèces de riz cultivé *O. sativa* et *O. glaberrima*, ainsi que les espèces de riz sauvages présentes en Afrique *O. longistaminata* et *O. barthii*. D'autres Poacées sauvages comme certaines *Chlorideae* (*Chloris prieri*, *Shoenefeldia gracilis*) et des *Eragrostideae* (*Dactyloctenium aegyptium*, *Eragrostis ciliaris*, *E. cilianensis*, *E. tenella*, et *E. tremula*) peuvent également être

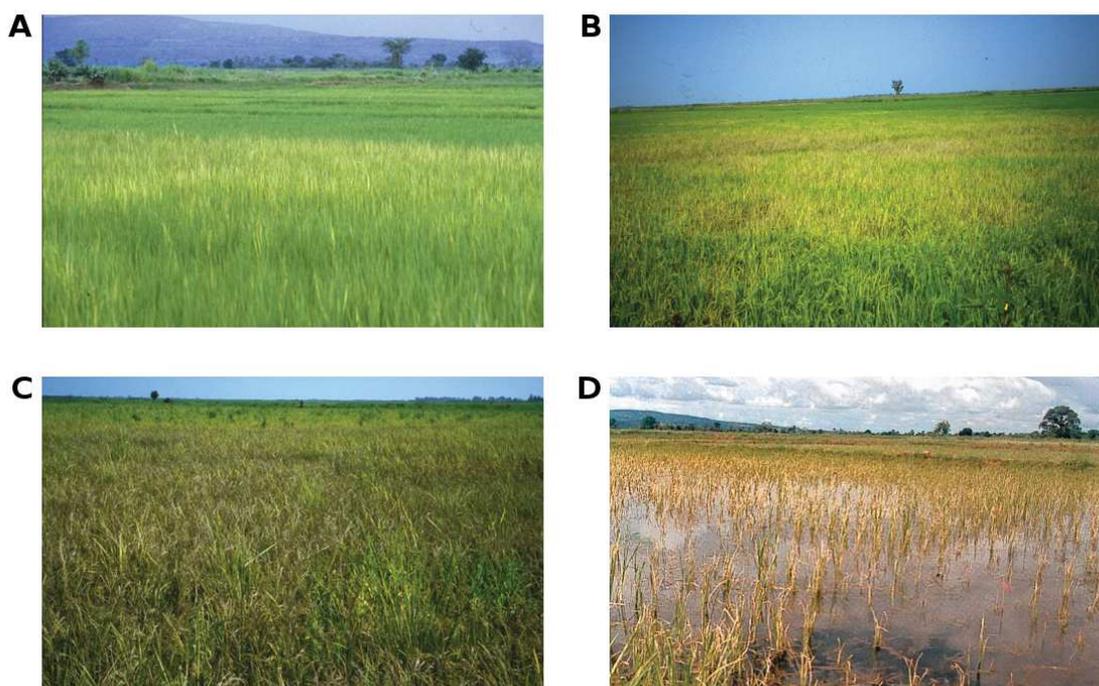


FIGURE 1.17 – Évolution des symptômes de la panachure jaune du riz au champ. Une tâche initiale de panachure jaune apparaît (A), puis s'étend à l'ensemble de la parcelle (B, J.-L. Notteghem). La maladie provoque ensuite un dessèchement généralisé (C, Office du Niger), pouvant aller jusqu'à la destruction complète de la culture (D, Warda).

infectées et pourraient constituer des réservoirs permanents de virus dans les écosystèmes (Allarangaye et al., 2006; Traoré et al., 2005).

Le virus peut être transmis de différentes manières. Une vingtaine d'espèces de coléoptères dont plusieurs du genre *Chrysomelidae* mais également dix espèces de sauterelles et six de cicadelles ont été décrites comme vecteur du RYMV. Elles le transmettent selon un mode semi-persistant non circulant (Abo et al., 2000; Bakker, 1974; Koudamilo et al., 2015). La transmission mécanique est extrêmement aisée en condition de laboratoire et certainement aisée en condition de culture *via* les contacts entre les feuilles, le pâturage et les pratiques agricoles comme la germination en pépinière, le repiquage, l'enfouissement des plants après récolte ou encore l'irrigation (Sarra et al., 2004; Traoré et al., 2008, 2009). Par ailleurs, bien que le virus ait été détecté dans les grains de riz avant dessiccation, la transmission par graine n'a jamais été observée (Allarangaye et al., 2006).

3.2.1.4 Importance économique

La panachure jaune du riz touche aujourd'hui la plupart de pays africains producteurs de riz. Dans les années 1980, de grandes phases épidémiques se sont matérialisées par des surfaces de cultures touchées par la maladie très importantes. Elles ont été estimées à 75 % dans le Sahel, 40 % au Soudan, 18 % en Guinée et 7,5 % dans les zones forestières tropicales humides (Kouassi et al., 2005). Plusieurs études menées dans différents pays

d’Afrique ont montré l’importance de l’incidence du RYMV et sa reconnaissance par les producteurs comme l’une des principales maladies du riz (Hubert et al., 2016; Issaka et al., 2012; Kam et al., 2013; Ochola and Tusiime, 2011; Sow et al., 2015). Par ailleurs, il a été classé par Rybicki (2015) comme l’un des dix principaux virus d’importance économique dans le monde.

3.2.2 Le *Rice yellow mottle virus*

3.2.2.1 Génome

Le RYMV appartient au genre des *Sobemovirus* (Hull and Fargette, 2005). Le genre n’a longtemps appartenu à aucune famille virale. Il a récemment été proposé de créer la famille des *Sobemoviridae*, incluant également le genre des *Polemovirus* (Sõmera et al., 2015). Actuellement 19 espèces de *Sobemovirus* ont été reconnues par l’*International Committee on Taxonomy of Viruses*.

La particule virale du RYMV est de symétrie icosaédrique et de 28 nm de diamètre. La capside est formée de 180 sous-unités organisées en 60 triangulations de 3 sous-unités (Opalka et al., 2000; Qu et al., 2000). Le génome est composé d’un ARN monocaténaire de polarité positive d’environ 4,5 kb lié en 5’ à une protéine virale associée au génome (VPg) (Hull and Fargette, 2005). L’acide nucléique comprend 5 cadres de lecture ouverts (ORF) et ne comprend pas de queue poly-A (Figure 1.18) (Ling et al., 2013; Sõmera et al., 2015).

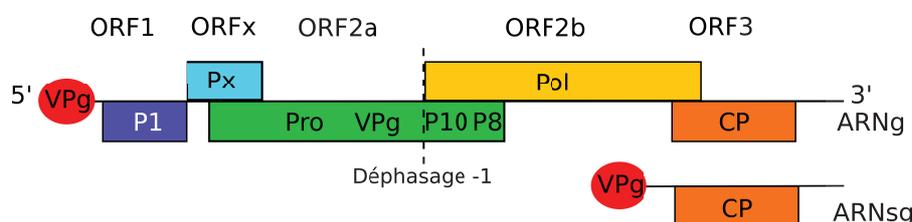


FIGURE 1.18 – Organisation génomique des *Sobemovirus*.

Les différentes protéines codées par les ORFs sont indiquées sous l’organisation du génome. Pro : protéase, Pol : polymérase, CP : protéine de capsid, VPg : protéine liée au génome viral, ARNg : ARN génomique, ARNsg : ARN subgénomique.

Les ORF en 5’ sont exprimés à partir de l’ARN génomique et l’ORF3, en 3’, est exprimé à partir d’un petit ARN subgénomique. L’ORF1 code la protéine P1, impliquée dans le mouvement de cellule à cellule et a un rôle dual dans l’activation et la suppression de l’interférence ARN (Bonneau et al., 1998; Lacombe et al., 2010; Siré et al., 2008). Cette protéine a également des points communs avec les protéines doigt-de-zinc (Gillet et al., 2013). L’ORF2a est traduit en une polyprotéine 2a dont la maturation produit une protéase à sérine, la protéine liée au génome viral (VPg) et les protéines P8 et 10, récemment décrites chez le *Sesbania mosaic virus* (SeMV) (Nair and Savithri, 2010; Satheshkumar et al., 2004). Après la traduction des 2/3 de l’ORF2a, un événement de déphasage ribosomique peut également survenir, conduisant à la traduction d’une

polyprotéine commune aux ORF2a et 2b. La maturation de cette dernière produit la VPg, la protéase à sérine et une polymérase RDRp. L'ORF3 code pour la protéine de capsid (CP : *Coat Protein*) qui est traduite à partir d'un ARN sub-génomique. La CP est nécessaire au mouvements de cellule à cellule et à longue distance pour lesquels l'encapsidation est nécessaire. Le cinquième ORF a été décrit récemment (Ling et al., 2013). Appelé ORFx, il est chevauchant avec l'extrémité 5' de l'ORF2a avec un cadre de lecture +2. Il code pour la protéine Px dont la fonction n'est pas connue mais dont l'inactivation dans le *Turnip rosette virus* bloque l'infection.

Par ailleurs, un ARN satellite de 220 nucléotides a été décrit comme étant associé à 75 % des isolats du RYMV (Pinel et al., 2003). Sa structure en bâtonnet rappelle celle d'un viroïde. Il n'a pas de fonction messagère et ne semble pas doté de propriétés pathogéniques particulières.

3.2.2.2 Phylogénie

La diversité du RYMV a été évaluée à l'échelle du continent africain. Six souches majeures ont pu être identifiées par sérotypage et par séquençage (Figure 1.19). Elles se divisent en 3 lignées, S4-S1, S5 et S6 qui ont une structuration spatiale très importante. La lignée S5 est présente uniquement dans la région Kilombero, près de l'Océan indien. La souche S6 est distribuée en Afrique de l'Est et les souches de la lignée S4-S1 sont présentes dans toute l'Afrique (Konaté et al., 1997; N'Guessan et al., 2001; Pinel et al., 2000; Pinel-Galzi et al., 2015; Trovão et al., 2015).

La diversité nucléotidique moyenne entre isolats est de 7 % et peut aller jusqu'à 15 % (Fargette et al., 2004). Par ailleurs, la vitesse d'évolution de ce virus est estimée entre 4 et 8×10^{-4} nucléotides par site et par an, ce qui correspond à une vitesse similaire à celle des virus animaux (Fargette et al., 2008a).

La phylogénie du RYMV a été reconstruite à partir de 366 isolats représentatifs de la diversité géographique et génétique du virus (Abubakar, 2003; Fargette et al., 2006, 2004, 2002a, 2008b; Pinel-Galzi et al., 2009, 2015) (Figure 1.19). Il y a une forte corrélation entre les typages moléculaires, sérologiques et l'origine géographique des isolats. La diversité génétique est maximale à l'est de la Tanzanie et décroît progressivement de l'Est vers l'Ouest. Ces analyses suggèrent une origine Tanzanienne du virus.

La datation moléculaire du RYMV suggère une émergence de ce virus il y a environ 180 ans en Afrique de l'Est (Fargette et al., 2008a; Pinel-Galzi et al., 2015; Trovão et al., 2015). L'introduction de l'espèce *O. sativa* en Afrique de l'Est au XIX^e siècle et l'intensification de la culture du riz sur le continent dès la moitié du XX^e siècle pourraient expliquer l'émergence de la panachure jaune du riz en Afrique de l'Est. L'apparition du virus en Afrique de l'Ouest aurait eu lieu il y a 100 ans dans le delta intérieur du Niger.

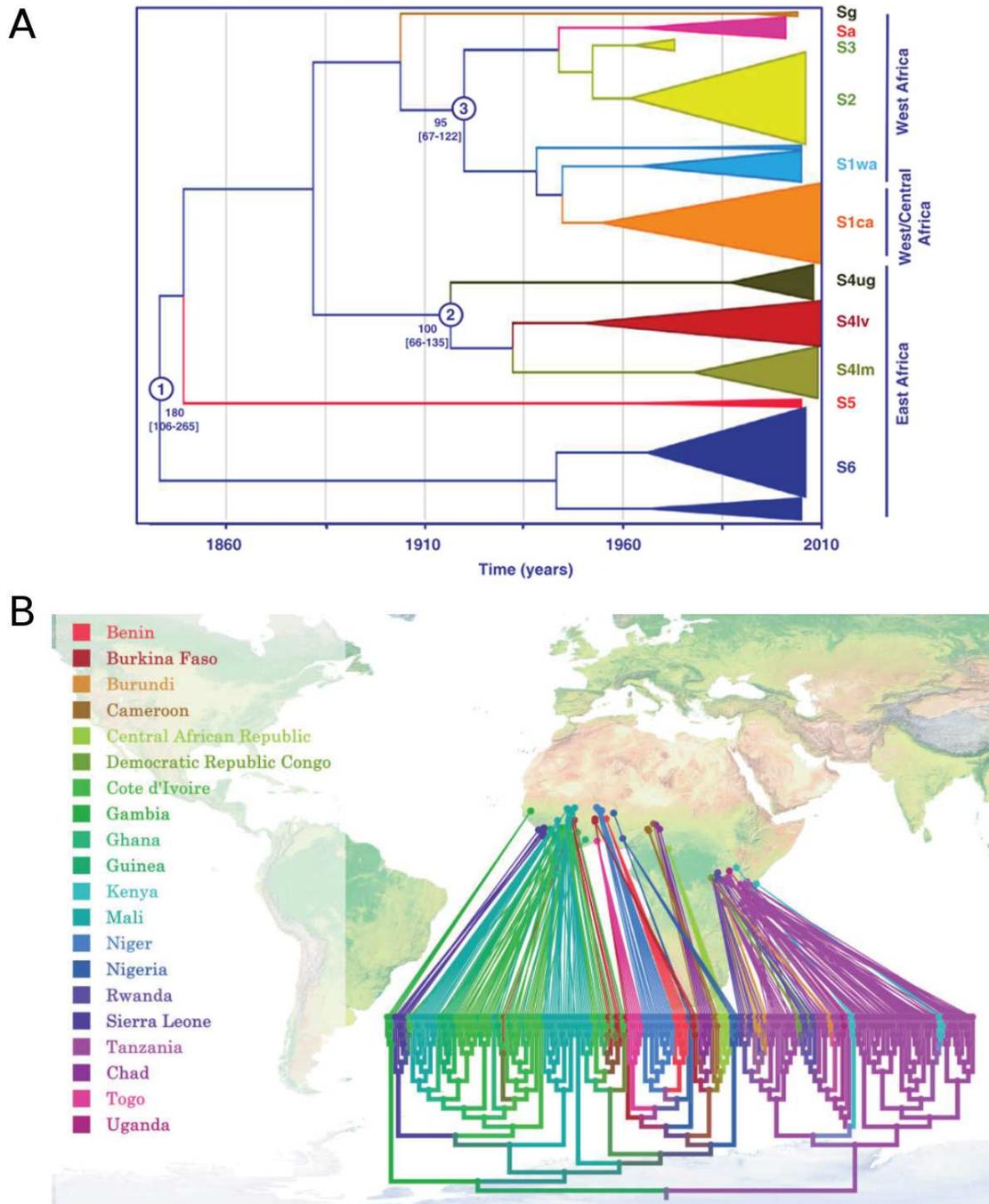


FIGURE 1.19 – Phylogéographie du RYMV. **A** : Arbre phylogénétique construit selon la méthode *maximum clade credibility* (MCC) et annoté avec les temps de divergence estimés (publié dans Pinel-Galzi et al. (2015)). L'arbre MCC a été construit à partir des séquences de l'ORF3 de 366 isolats originaires de 20 pays d'Afrique et collectés entre 1966 et 2012. **B** : Arbre MCC relié à la localisation de collecte des isolats (publié dans Trovão et al. (2015)). Les branches de l'arbre sont colorées en fonction des pays d'origine des isolats.

3.3 Les résistances du riz au RYMV

Afin de lutter contre la maladie, la présence de résistances a été recherchée dans les deux espèces cultivées de riz. Ces études ont permis d'identifier plusieurs sources de résistance majeures : deux variétés *O. sativa* et plusieurs accessions *O. glaberrima* (Ndjiondjop et al., 1999; Thiémélé et al., 2010). Deux des gènes expliquant ces résistances ont été identifiés : *RYMV1* et *RYMV2*. D'autres gènes sont encore à découvrir dans l'espèce *O. glaberrima*.

L'étude de la résistance partielle d'*O. sativa* a montré que les *indica* étaient globalement moins résistants que les *japonica*. Au sein des *japonica*, les variétés dites tropicales ont une plus forte résistance partielle. Des études QTL ont été réalisées sur des descendances Azucena (*O. sativa* spp *japonica* tropical partiellement résistant) x IR64 (*O. sativa* spp *indica* sensible) et ont permis d'identifier plusieurs QTL (Albar et al., 1998; Boisnard et al., 2007).

L'espèce *O. glaberrima* contient également des sources de résistances partielles (Thiémélé et al., 2010) mais leur déterminisme n'a pas été étudié.

3.3.1 *RYMV1*

L'étude du déterminisme génétique de la résistance de Tog5681 (*O. glaberrima*) et de Gigante (*O. sativa*) a permis de mettre en évidence un gène, nommé *RESISTANCE TO RICE YELLOW MOTTLE VIRUS-1* (*RYMV1*) présent sur le chromosome 4. Le virus se multiplie dans les protoplastes des accessions résistantes mais n'est pas détecté dans les plantes, suggérant que la résistance se traduirait par le blocage du mouvement de cellule à cellule du virus chez Tog5681 et Gigante (Ndjiondjop et al., 2001).

RYMV1 code pour un facteur d'initiation de la traduction eIF(iso)4G1 (Albar et al., 2006). Cinq allèles différents ont été identifiés dans deux accessions d'*O. sativa* et 15 accessions d'*O. glaberrima* (Albar et al., 2006; Thiémélé et al., 2010). L'allèle *rymv1-2* est spécifique aux accessions d'*O. sativa* résistantes (Gigante et Bekarosaka), l'allèle *rymv1-3* a été détecté dans neuf accessions d'*O. glaberrima*, parmi lesquelles Tog5681, l'allèle *rymv1-4* est présent dans cinq accessions d'*O. glaberrima* et l'allèle *rymv1-5* dans une seule. Les quatre allèles de résistance se caractérisent par des substitutions (E309K pour *rymv1-2* et E321K pour l'allèle *rymv1-4*) ou de courtes délétions (RRD322-324 pour l'allèle *rymv1-3* et LTG313-315 associé à une substitution K312N pour l'allèle *rymv1-5*) dans la même région centrale du domaine MIF4G de la protéine. La modélisation du domaine (Figure 1.20) met en évidence que ces mutations sont très proches et situées en surface de la protéine (Albar et al., 2006).

Des études de copurification de la VPg du virus et de eIF(iso)4G1 ainsi que de double-hybride en levure ont mis en évidence une interaction directe entre la VPg du RYMV et le domaine MIF4G de *RYMV1* (Hébrard et al., 2010). Cette interaction est rompue par la mutation *rymv1-2*. Au vu des mécanismes des facteurs d'initiation de la traduction, il est probable que l'interaction de la VPG du virus avec *RYMV1* soit

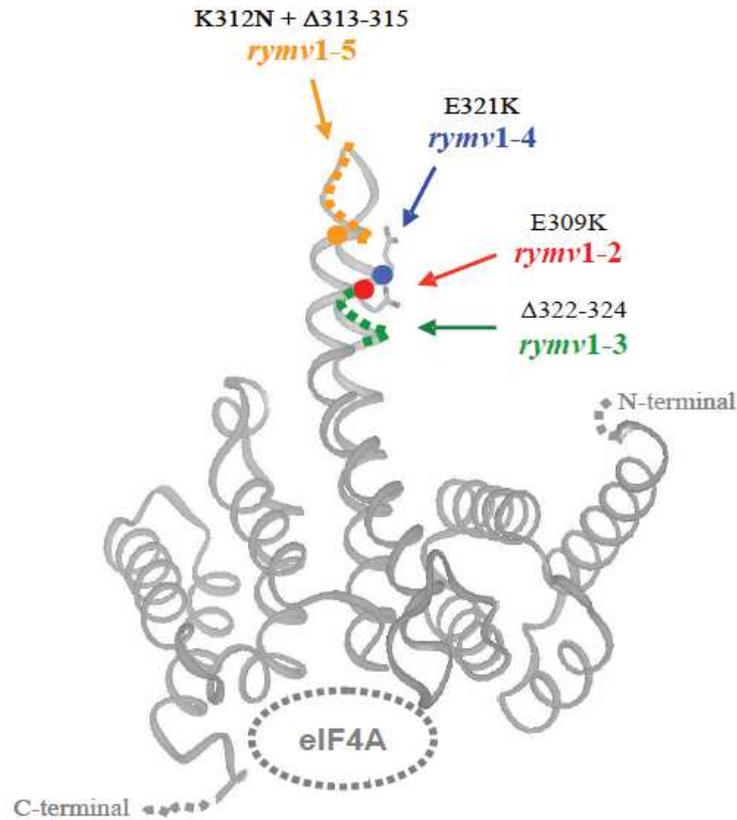


FIGURE 1.20 – Modélisation de structure du domaine central du facteur eIF(iso)4G1 du riz, et localisation des différentes mutations de résistance contre le RYMV (d'après Albar et al. (2006) ; Thiémélé et al. (2010)).

Le nom des allèles de résistance et les mutations associées sont mentionnés. L'allèle *rymv1-2* correspond à une substitution d'un acide glutamique (E) en lysine (K) en position 309 du facteur eIF(iso)4G1. Les allèles *rymv1-3*, *rymv1-4* et *rymv1-5* correspondent respectivement à une délétion de trois acides aminés chargés (arginine-arginine-acide aspartique ; RRD) en positions 322-324, à une substitution de E en K en position 321, et à la combinaison d'une substitution d'une lysine en asparagine (N) et d'une délétion de trois acides aminés (leucine-thréonine-glycine ; LTG) respectivement en positions 312 et 313-315. Les extrémités C- et N-terminal, ainsi que le domaine de fixation avec le facteur eIF4A sont indiquées

nécessaire au virus pour détourner le complexe d'initiation de la traduction à son profit.

3.3.2 RYMV2

L'accèsion *O. glaberrima* Tog7291 présente également une résistance élevée au RYMV, indépendante de *RYMV1*. Cette résistance implique également un gène majeur récessif, appelé *RESISTANCE TO RICE YELLOW MOTTLE VIRUS-2* (*RYMV2*) (Thiémiélé et al., 2010).

Le gène de résistance a été identifié à l'extrémité distale du bras long du chromosome 1 dans un intervalle d'environ 35 kb (Orjuela et al., 2013). Cet intervalle comprend notamment le gène *CPR5-1* (LOC_Os01g68970), dont l'allèle chez Tog7291 est un allèle nul, engendrant un décalage du cadre de lecture après le dix-huitième codon. La protéine est tronquée et supposée non fonctionnelle. Sept accèsions résistantes d'*O. glaberrima* de la collection, toutes résistantes, présentent ce polymorphisme. L'étude de deux événements d'insertion T-DNA indépendants dans des introns du gène *CPR5-1* dans la variété *O. sativa* spp *japonica* Hwayoung (Jeon et al., 2000; Jeong et al., 2006) a permis de confirmer qu'un allèle nul à *CPR5-1* confère une résistance élevée au RYMV (Albar, communication personnelle).

Le gène *CPR5-1* est un homologue du gène *constitutive expressor of pathogenesis-related genes-5* (*CPR5*) d'*Arabidopsis thaliana*. Les mutants *cpr5* d'*A. thaliana* présentent une résistance constitutive à des agents pathogènes bactériens et à des oomycètes (Bowling et al., 1997). Cette résistance est notamment associée à une augmentation du niveau d'acide salicylique et de l'expression constitutive du gène *pathogenesis-related 1*.

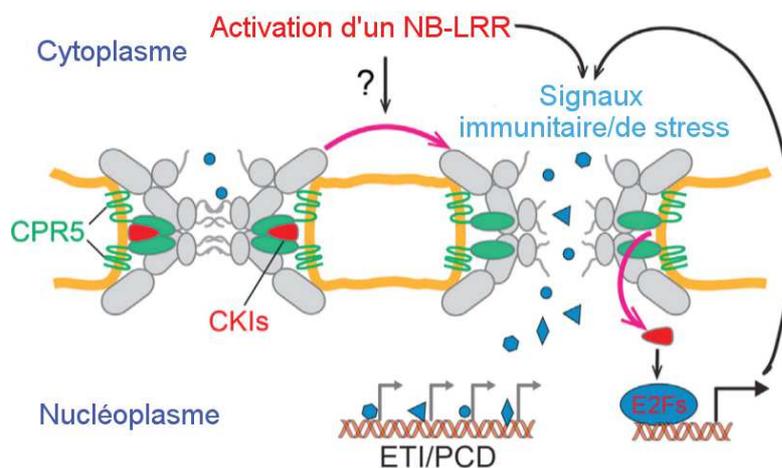


FIGURE 1.21 – Fonctionnement de CPR5 chez *A. thaliana* (publié par Gu et al. (2016)). Après l'activation d'un NB-LRR, un signal intracellulaire inconnu est généré et transduit au pore nucléaire, conduisant à la rupture des oligomères de CPR5. Ce changement du pore nucléaire permet la libération des CKI pour la signalisation de l'ETI et change la perméabilité du pore nucléaire, permettant un influx important de signaux de stress dans le noyau au travers du pore nucléaire. Ces effets combinés activent différentes voies de signalisation nucléaire contribuant à l'ETI et la HR.

L'effet de *CPR5* sur la résistance aux virus est contrasté. Les mutants *cpr5* d'*A. thaliana* ont une résistance accrue aux *Cauliflower mosaic virus* (Love et al., 2007), *Pepper mild mottle virus* et *Tobacco mild green mosaic virus* (Fujisaki et al., 2009). En revanche, la multiplication du *Spring beauty latent virus* n'est pas affectée par la mutation *cpr5* et le *Brome mosaic virus* a même un taux de multiplication supérieur dans les mutants, bien qu'*A. thaliana* soit considérée non-hôte pour ce virus. *CPR5* a aussi été rapporté avoir un effet sur la prolifération des cellules et le contrôle de la mort cellulaire (Kirik et al., 2001), sur la synthèse de la paroi cellulaire et le développement des trichomes (Brininstool et al., 2008), dans la réponse aux sucres (Yoshida et al., 2002), et dans la régulation de la sénescence (Jing et al., 2002). À partir de l'analyse de doubles mutants, Jing et al. (2007) ont proposé que *CPR5* pourrait contrôler plusieurs mécanismes cellulaires de manière indépendante et pourrait donc être un super régulateur de l'état et de la signalisation des dérivés réactifs de l'oxygène. Récemment, il a été décrit que *CPR5* est une nucléoporine faisant partie du pore nucléaire (Gu et al., 2016) (Figure 1.21). Cette protéine formerait des homo-oligomères et interagirait notamment avec un inhibiteur des kinases dépendant des cyclines (CKI) (Wang et al., 2014d). L'activation de la signalisation avale des NB-LRR romprait les oligomères, libérant les CKI et rendant le pore nucléaire plus perméable aux signaux de stress. Ces signaux de stress seraient ensuite impliqués dans l'activation de l'expression de gènes de défense.

Les mutants *cpr5* d'*A. thaliana* ayant une résistance à plusieurs agents pathogènes différents, 4 agents pathogènes disponibles dans les unités montpelliéraines ont été testés sur un des deux mutants d'insertion dans *CPR5-1* chez le riz. Il s'agit de deux bactéries (*Xanthomonas oryzae* pv. *oryzae* et *Xanthomonas oryzae* pv. *oryzicola*), un champignon (*Magnaporthe oryzae*) et un nématode (*Meloidogyne graminicola*). Combiné à la présence d'un deuxième homologue de *CPR5* chez le riz pouvant compléter fonctionnellement la perte de fonction de *CPR5-1*, cela suggère que le mécanisme en jeu chez le riz est plus complexe que celui observé chez *A. thaliana*.

De plus, l'introggression de *CPR5-1* dans un fond IR64 (*O. sativa* spp *indica*) ne suffit pas à procurer la résistance. Un ou plusieurs autres gènes doivent interagir avec *CPR5-1* au sein de Tog7291 pour procurer la résistance. De premiers résultats pointent vers un effet du deuxième homologue de *CPR5* : *CPR5-2*.

3.3.3 QTL de résistance chez *O. sativa*

Parmi les accessions d'*O. sativa* spp *japonica* tropicaux, plusieurs sont partiellement résistantes. Deux études QTL ont été réalisées sur des descendance issues du croisement Azucena (*O. sativa* spp *japonica* tropical partiellement résistant) x IR64 (*O. sativa indica* sensible) (Albar et al., 1998; Boisnard et al., 2007). Ces études, réalisées sur des RIL et sur des lignées d'haploïdes-doublés (DHL) ont permis de mettre en évidence deux QTL principaux à l'extrémité de chromosome 1, et sur le chromosome 12, ainsi que des QTL mineurs sur les chromosomes 2, 4, 6, 7 et 8 (Figure 1.22).

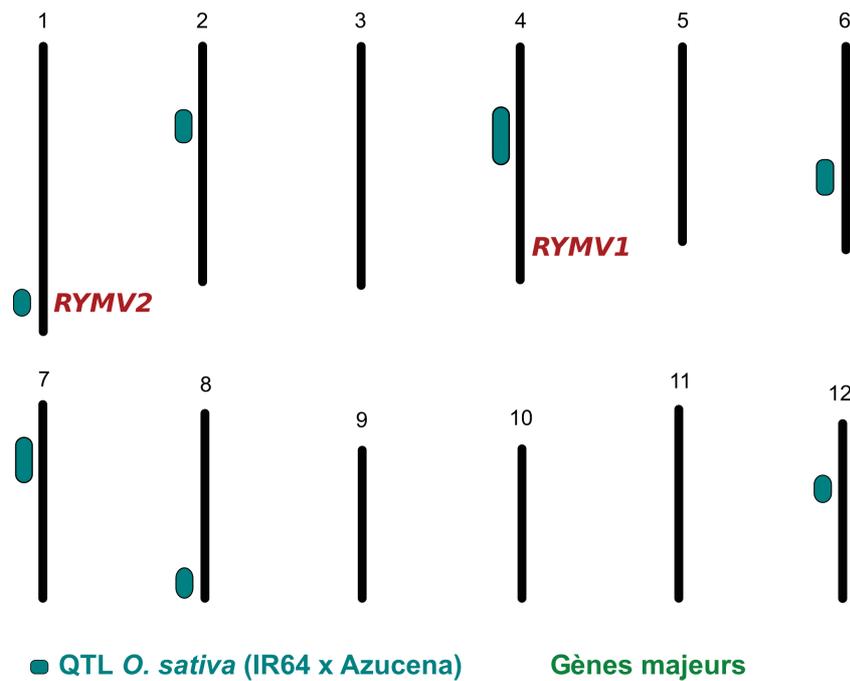


FIGURE 1.22 – Position des QTL et gènes majeurs impliqués dans la résistance au RYMV. Les QTL ont été identifiés dans des populations DH et RIL du croisement (IR64 x Azucena) (d’après (Boisnard et al., 2007)).

Une corrélation entre le niveau de résistance et les caractères morphologiques et développementaux de la plante a été relevée (Albar et al., 1998). Elle pourrait expliquer la résistance partielle généralement observée chez les *japonica* tropicaux.

Le QTL₁₂, indépendant de critères morphologiques, est un candidat particulièrement intéressant pour l’introgession dans des variétés *indica*. Il a été cartographié dans un intervalle de 1,5 Mb mais la faible recombinaison dans la région est un frein à une cartographie plus fine (Boisnard, 2007).

La cartographie du QTL₁ de résistance d’*O. sativa* spp *japonica* dans des populations de back-cross IR64 x Azucena a permis de réduire ce QTL a un intervalle de 150kb qui contient le gène *CPR5-1*, suggérant que le même gène pourrait être impliqué dans la résistance partielle d’Azucena et dans la résistance complète de Tog7291. Toutefois à ce jour aucune mutation candidate n’a été identifiée.

3.3.4 Le contournement de la résistance

Le contournement des sources de résistances du RYMV a fait l’objet de plusieurs études. Elles permettent de mieux comprendre l’interaction entre la plante et le virus, mais également d’avoir un aperçu des interactions en jeu dans l’établissement de la résistance.

Plusieurs mutations, apparaissant *de novo* dans la plante, sont impliquées dans le contournement des résistances de Gigante, Bekarosaka et Tog5681. Toutes sont situées dans le domaine central de la VPg entre les codons 37 et 52. Le contournement de l’allèle *rymv1-2* chez Gigante et Bekarosaka est permis par la mutation du codon 48, ou,

dans des cas plus rares, du codon 52 (Pinel-Galzi et al., 2007). Le contournement de la résistance de Tog5681 est, lui, rendu possible par des mutations aux positions 41 et 52 (Traoré et al., 2010). Ces mutations permettent le rétablissement de l'interaction entre la VPg et *RYMV1* dans les plantes résistantes (Hébrard et al., 2008), avec toutefois une perte de fitness de ces variants virulents (Poulicard et al., 2010).

La résistance de Gigante implique également des mutations dans l'extrémité 3' de l'ORF2a, chevauchante de l'ORF2b (Poulicard et al., 2014). Cette région correspond à l'homologue de la protéine P8 décrite chez le SeMV (Nair and Savithri, 2010; Satheskumar et al., 2004). Le mécanisme expliquant ce contournement n'est toutefois pas établi. Ces mutations, impliquant une baisse de fitness importante, sont réversées dans un hôte sensible.

Le contournement de la résistance de Tog7291 (*RYMV2*) implique des mutations au début de l'ORF2a, dans le domaine d'ancrage à la membrane de la polyprotéine 2a (Pinel-Galzi et al., sub). Ce domaine est supposé permettre l'ancrage de la protéine dans les membranes cellulaires et semble impliqué dans la synthèse du brin négatif du virus (Sõmera et al., 2015).

Par ailleurs, le résidu 49 de la VPg, très conservé, est une position clé dans le contournement des différentes résistances. Le polymorphisme E/T à cette position détermine notamment le contournement de *RYMV1* (Traoré et al., 2010). Ainsi les isolats possédant un acide glutamique à cette position (E49) ne contournent pas *rymv1-3* chez Tog5681 alors que ceux qui ont une thréonine (T49) contournent difficilement *rymv1-2* chez Gigante. Par ailleurs, les isolats T49 ont un avantage sélectif dans les accessions *O. glaberrima* sensibles (Poulicard et al., 2012). Ce polymorphisme est également impliqué dans le contournement de la *RYMV2* : les isolats T49 contournent plus facilement la résistance de Tog7291 (Pinel-Galzi et al., sub). Le polymorphisme T49 pourrait donc être une adaptation du virus à l'espèce *O. glaberrima*.

Ce polymorphisme est en lien avec la distribution géographique des isolats. Les isolats E49 sont majoritaires en Afrique de l'Est et dans la partie savane de l'Afrique de l'Ouest alors que ceux qui ont l'acide aminé T49 le sont dans la partie boisée de l'Afrique de l'Ouest. Cette différence peut notamment permettre de mettre en place un plan de déploiement régional des différents allèles de résistance afin d'obtenir une meilleure durabilité de la résistance.

CHAPITRE

2

Matériels et méthodes généraux

Chapitre 2 : Matériels et méthodes généraux

Nous avons décidé de présenter ici les éléments de matériels et méthodes communs à l'ensemble des chapitres de cette thèse et qui constituent le socle de l'ensemble des expérimentations réalisées. Les éléments particuliers à un chapitre seront spécifiquement développés dans les chapitres suivants. Ainsi, le chapitre 3 est présenté sous forme d'article et possède sa propre section matériel et méthodes. De même, les conditions expérimentales et les analyses réalisées pour l'étude de la résistance partielle sont développées au chapitre 5.

1 Matériel végétal

Les riz africains ont des caractères agronomiques d'intérêt vis-à-vis des stress biotiques et abiotiques auxquels font face les cultures en Afrique (Linares, 2002) tels que la sécheresse, la toxicité ferrique, ou encore la résistance à certains agents pathogènes. Afin d'étudier ces caractères, l'IRD a réuni une collection d'accessions de riz cultivé africain *O. glaberrima* et de son apparenté sauvage *O. barthii*. Nous nommerons par la suite cette collection "collection IRD". Elle est composée de 278 accessions d'*O. glaberrima* et de 103 accessions *O. barthii* (Table A.1 en annexe). Le pays d'origine de chacune de ces accessions a été référencé (Figure 2.1).

Les accessions constituant la collection *O. glaberrima* ont été choisies conjointement par l'IRD et l'Africa Rice. Cette collection rassemble notamment les accessions parentes des lignées NERICA, des accessions de populations références, et des accessions ayant des caractères intéressants de résistance à des stress biotiques et abiotiques. Les accessions choisies représentent au mieux la distribution géographique de l'espèce en Afrique de l'Ouest avant 1950, date à partir de laquelle son remplacement progressif par des variétés de l'espèce *O. sativa* s'est accentué (Portères, 1962).

Lors de l'établissement de la collection *O. barthii*, un effort particulier a été fait pour rassembler des accessions réellement sauvages, représentant la diversité actuelle. Ces accessions proviennent de populations collectées dans des régions éloignées des champs. Elles sont issues de l'ancienne collection ORSTOM (ex-IRD), construite entre 1974 et 1982 au cours de missions prospectives.

L'ensemble des accessions a ensuite été vérifié afin d'en éliminer les *O. sativa* hors-types, les hybrides entre sauvage et cultivé ainsi que les riz adventices. Ces tris ont été effectués sur la base de la morphologie et de la phénologie des plantes ainsi que de leur répartition dans la diversité, obtenue à l'aide d'une puce SNP Illumina spécifiquement définie pour étudier la diversité des riz africains (Orjuela et al., 2014). En outre, il a été vérifié que ces accessions possédaient l'allèle spécifique des riz africains au locus de stérilité interspécifique *S1* (Orjuela, communication personnelle).

Afin de stabiliser les accessions, chacune d'entre elles a été autofécondée au moins une fois. Des graines de l'ensemble des accessions de la collection ont été envoyées au

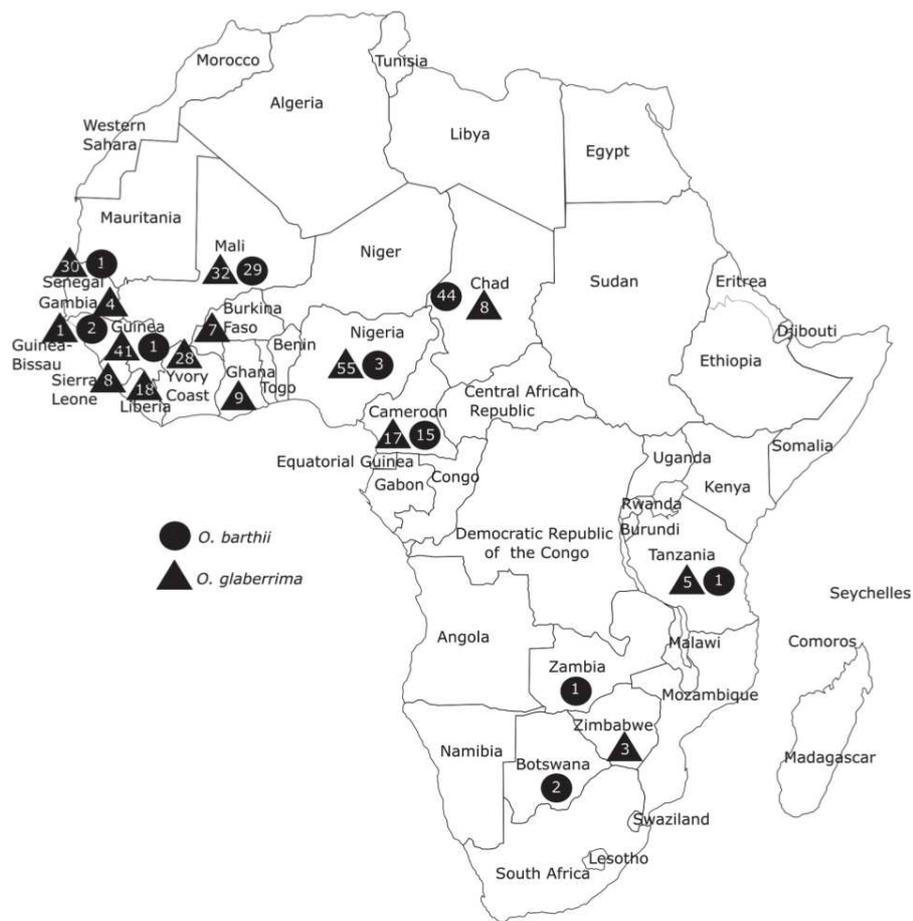


FIGURE 2.1 – Distribution géographique des accessions de riz africains utilisés dans cette étude (publié dans Orjuela et al. (2014)). Le nombre d'accessions *O. glaberrima* (triangle) et *O. barthii* (cercle) est indiqué pour chaque pays.

Burkina Faso où elles ont été multipliées. Lorsque le stock local d'une accession était insuffisant nous avons parfois utilisé les graines issues de cette multiplication. Plusieurs sources d'erreurs existent entre ces deux lots, de l'échantillonnage dans la collection IRD à l'ensachage au Burkina-Faso. Nous avons en effet pu observer que la conformité de cette multiplication avec la collection IRD n'était pas parfaite.

Parmi les accessions de la collection IRD, un sous-ensemble de 163 accessions *O. glaberrima* et 84 *O. barthii* a été sélectionné pour représenter au mieux la diversité d'écotypes, d'origine géographique et la diversité génétique de l'espèce. Le choix a également pris en compte la disponibilité en graines et les accessions possédant un phénotype intéressant. Ces accessions forment la collection ici appelée "collection IRIGIN" du nom du projet de séquençage dont elles ont fait partie et qui est détaillé dans la suite de ce chapitre.

Plusieurs accessions de la collection IRD ont été identifiées comme résistantes. Parmi elles, Tog5681 est l'accession de référence pour la résistance conférée par l'allèle de résistance *rymv1-3* et Tog7291 pour celle conférée par l'allèle de résistance à *RYMV2*. Le déterminisme de la résistance de Tog5307 est étudié au chapitre 3, et Tog5672 possède à la fois l'allèle de résistance *rymv1-4* et un second déterminisme de résistance inconnu. (Thiémélé et al., 2010).

Pour des raisons historiques et de cohérence entre publications, nous utilisons pour les principales accessions résistantes les noms donnés dans la collection de l'IITA ("Tog"). En revanche, les analyses sur l'ensemble de la collection sont faites avec les numéros donnés dans la collection IRD. Le tableau 2.1 recense la correspondance pour les accessions concernées dans cette thèse ainsi que leurs déterminismes de résistance.

TABLE 2.1 – Correspondance entre les numéros de la collection IRD et ceux de la collection IITA des accessions fréquemment utilisées.

Numéro IITA	Numéro IRD	allèle ou gène de résistance
Tog5681	Og49	<i>rymv1-3</i> (Ndjiondjop et al., 1999)
Tog5672	Og44	<i>rymv1-4</i> + autre déterminisme (Chapitre 3)
Tog7291	Og172	<i>RYMV2</i> (Orjuela et al., 2013)
Tog5307	Og6	voir chapitre 3

Deux accessions *O. sativa* de référence sont utilisées comme témoins sensibles au RYMV au laboratoire : IR64 (*O. sativa* spp *indica*) et Nipponbare (*O. sativa* spp *japonica*). La variété *O. sativa* spp *japonica* Azucena est un témoin de résistance partielle.

2 Matériel viral

Le RYMV a une diversité d'isolats bien caractérisée. Depuis les 30 dernières années, plus de 500 isolats provenant de la plupart des pays africains producteurs de riz ont été

collectés.

L'étude de la diversité de cette collection a permis d'identifier six souches regroupées en trois lignées, sur la base de la séquence de l'ORF3 (Konaté et al., 1997; N'Guessan et al., 2001; Pinel et al., 2000; Pinel-Galzi et al., 2015). Les isolats se différencient également en deux pathotypes établis sur la base de leur profil d'interaction avec des accessions possédant un allèle de résistance à *RYMV1* et caractérisés et nommés par leur acide aminé en position 49 de la VPg.

Deux isolats principaux ont été utilisés pour cette étude. L'isolat BF1, de souche S2 et originaire du Burkina Faso est un isolat T49 très agressif utilisé comme isolat de référence pour l'évaluation de la résistance au laboratoire (Albar et al., 2006; Orjuela et al., 2013). L'isolat CI4, de souche S1 et provenant de Côte d'Ivoire, est un isolat E49, supposé moins contourner les résistance d'*O. glaberrima* et moins agressif (Pinel et al., 2000). Il permet donc des réponses plus quantitatives et a été utilisé afin d'évaluer la résistance partielle au RYMV dans la collection *O. glaberrima*.

Les isolats ont été multipliés sur la variété *O. sativa* sensible IR64. Les feuilles infectées sont récoltées deux semaines après inoculation et stockées à -80°C . Les virus à ARN évoluant rapidement, deux lots de multiplication du même isolat peuvent être différents au niveau génétique. Afin de limiter cet effet, une même expérience a été réalisée, autant que possible, avec un inoculum provenant du même lot de multiplication.

3 Évaluation de la résistance au RYMV

Les plantes ont été cultivées en serres tropicalisées. Leur résistance au RYMV a été évaluée en les inoculant avec le virus puis en observant les symptômes ou en mesurant le contenu en virus dans les feuilles avec un test ELISA.

3.1 Conditions d'inoculation

L'inoculum est composé de ces feuilles broyées dans du tampon phosphate (0,1 M KH_2PO_4 , 0,1 M Na_2HPO_4 , pH 7,2) à raison d'un gramme pour 10 mL. Nous avons utilisé deux techniques d'inoculation différentes selon le type d'expérience.

Classiquement, l'inoculation est réalisée de façon mécanique, par frottement de la feuille entre deux doigts trempés dans la solution d'inoculation contenant un abrasif.

Dans le cas de l'évaluation de la résistance partielle, l'inoculation a été faite par infiltration à la seringue sans pointe en deux points à la base de la dernière feuille développée, trois semaines après semis. Une date d'inoculation plus tardive permet une réponse plus quantitative. L'inoculation par infiltration permet de repérer le point d'inoculation. On peut ainsi distinguer les feuilles inoculées des feuilles infectées par voie systémique et vérifier que la plante a bien été inoculée. Nous supposons également que cela permettrait une meilleure répétabilité de la quantité de virus inoculée.

3.2 Observation des symptômes

La plupart du temps, la résistance élevée a été évaluée sur la base de la présence ou de l'absence de symptômes deux à trois semaines après inoculation. Le développement de la maladie débute par l'apparition de tâches jaunes sur la dernière feuille. Puis ces tâches se multiplient jusqu'à former une panachure de l'ensemble de la feuille. Lorsque les symptômes sont très avancés, les feuilles deviennent complètement jaunes puis sèchent. Cependant, pour la plupart des accessions *O. glaberrima*, la panachure n'est pas nette et les symptômes se caractérisent le plus souvent par un jaunissement progressif des feuilles.

3.3 Évaluation du contenu en virus

Les symptômes sur *O. glaberrima* étant souvent peu clairs dans nos conditions de culture, nous avons évalué l'accumulation virale dans la feuille systématiquement infectée par le virus par un test de dosage immunoenzymatique DAS-ELISA (Double Antibody Sandwich-Enzyme Linked ImmunoSorbent Assay). Ce test se base sur la reconnaissance spécifique d'une protéine du virus par des anticorps polyclonaux dirigés contre elle et la révélation de ces liaisons par colorimétrie. Selon l'objectif de chaque expérience, ce test était qualitatif ou semi-quantitatif.

Le test DAS-ELISA a été réalisé avec un anticorps polyclonal dirigé contre un isolat de RYMV de Madagascar (N'guessan et al., 2000). Les échantillons de feuilles ont tout d'abord été broyés au broyeur à billes puis suspendus dans une solution saline tamponnée au phosphate contenant 0,5 % de Tween-20 (PBS-T 1X : 1,36 M NaCl, 14 mM KH_2PO_4 , 80 mM Na_2HPO_4 anhydre, 26 mM KCl, 5 mL/L Tween20). À partir de cette suspension initiale, plusieurs dilutions ont été réalisées selon les besoins.

L'anticorps, dilué au 1:1000 dans du tampon carbonate (15 mM Na_2CO_3 , 34 mM NaHCO_3 , pH 9,6), a été fixé sur les plaques ELISA (NUNC-Immuno Plate Maxi Sorp) à raison de 100 μL par puit. Après deux heures d'incubation à 37 °C, les plaques ont été lavées trois fois avec du tampon PBS-T 1X. Les puits ont ensuite été saturés avec 200 μL de lait écrémé en poudre dilué à 3g pour 100 mL de tampon PBS-T 1X. Après 1h à 37 °C, les plaques ont été lavées deux fois. Les échantillons, dilués dans du tampon PBS-T 1X, ont ensuite été déposés à raison de 100 μL par puit. Après incubation de 2h à 37 °C et réalisation de quatre lavages des plaques, l'anticorps secondaire, conjugué à la phosphatase-alcaline et dilué au 1:1000 dans du tampon PBS-T 1X, a été déposé à raison de 100 μL par puit. Les plaques ont à nouveau été incubées 2h à 37 °C et lavées quatre fois. Cent microlitres d'une solution à 1 mg/mL de para-nitrophényl phosphate dans un tampon diéthanolamine (98 mL/L ; pH 9,8) ont ensuite été ajoutés à tous les puits. Les plaques ont été incubées à l'obscurité à température ambiante puis leur densité optique (DO) à 405 nm a été lue. Cette expérience durant plus d'une journée, une étape d'incubation de 2h à 37 °C était remplacée par une nuit à 4 °C.

Dans les tests qualitatifs, les plantes étaient considérées résistantes lorsque leur

valeur de DO était inférieure à trois fois la valeur du témoin négatif.

Afin d'éviter les effets bords pouvant induire une variation de la mesure, seuls les 60 puits au centre de la plaque ELISA ont été utilisés dans les tests semi-quantitatifs. De même, pour assurer une meilleure répétabilité, l'étape d'une nuit à 4 °C a toujours été celle de la fixation des anticorps à la plaque.

4 Ressources génomiques

Plusieurs ressources génomiques ont été utilisées au cours de ce travail.

Nous avons notamment utilisé les génomes de référence des variétés *O. sativa* Nipponbare (Os-Nipponbare-Reference-IRGSP-1.0) (Kawahara et al., 2013) et IR64 (Os-IR64-Draft-CSHL-1.0) (Schatz et al., 2014) ainsi que de la variété *O. glaberrima* CG14 (Oryza_glaberrima_AGI1.1)(Wang et al., 2014c), toutes trois sensibles au RYMV . Nous avons également exploité les annotations des génomes de Nipponbare et CG14 (2011-05-AGI, Munich Information Center for Protein Sequences).

Les domaines des différents gènes d'intérêt ont été obtenus selon les cas par Interpro (<https://www.ebi.ac.uk/interpro/>) ou NCBI Conserved Domains (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

Nous avons bénéficié des données de séquences de riz africains issues du projet IRIGIN (International Rice Genomic INitiative). Ce projet est coordonné par l'IRD, en collaboration avec le CIRAD, le CIAT (Colombie) et l'AfricaRice (organisme pan-Africain du CGIAR), le tout au sein du Mégaprogramme GRiSP (*Global Rice Science Partnership*). Il a pour objet le séquençage d'environ 7000 lignées de riz d'intérêt agronomique et évolutif qui seront séquencées, avec des couvertures de 1 à 35x. Cela inclut les riz africains, les NAM (*Nested Association Mapping populations*, créées pour optimiser la découverte de gènes), ou les populations de sélection génomique.

Les génomes de l'ensemble des accessions de la collection IRIGIN ont été acquis au travers du projet IRIGIN (France génomique). L'ADN des différentes accessions a été extrait à partir de plantes issues de multiplications IRD, en utilisant le kit QIAGEN DNAeasy (Allemagne). La librairie de séquençage a été construite avec le kit TruSeq V3 (Illumina). Le séquençage a été réalisé par le Génomscope (Évry, France) par Illumina (HiSeq 2000/2500/4000) avec des *reads paired-ends* de 100 paires de bases. La profondeur de séquençage moyenne était de 35x, avec un maximum de 52x et un minimum de 21x.

L'analyse bioinformatique des données, décrite ci-après, a été réalisée à l'IRD par Cécile Monat, Christine Tranchant-Dubreuil et François Sabot. Les *reads* ont été traités en utilisant "fastX-trimmer" pour éliminer les bases de mauvaise qualité aux extrémités (qualité inférieure à 20) et retirer les adaptateurs. Les *reads* ont ensuite été mappés au travers du *pipeline* "TOOGLE" (Monat et al., 2015) sur le génome de référence de Nipponbare et sur celui de CG14 pour 120 accessions *O. glaberrima*. Le *pipeline* était composé de plusieurs étapes : contrôle de la qualité des séquences ("FASTQC", <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), élimination

des séquences de taille inférieure à 30 pb avec "cutadapt" (Martin, 2011), *mapping* avec "BWA aln" et "sampe" en autorisant cinq mésappariements et des *gaps* de deux nucléotides (Li and Durbin, 2009), conversion du fichier de *mapping* au format BAM avec "picard-tools sortSam" (<http://broadinstitute.github.io/picard/>), sélection des *reads* alignés correctement en paire avec "samtools view" (Li et al., 2009), réalignement local avec "GATK realignerTargetCreator" et "indelRealigner" (McKenna et al., 2010), puis détection des *reads* correspondant à des duplicats techniques avec "picard-tools markduplicates".

Les fichiers BAM obtenus à l'issue du *pipeline* ont été séparés par chromosomes (logiciel "BamUtil splitChromosome") puis l'appel de SNP a été réalisé avec "GATK UnifiedGenotyper" (options par défaut). Les polymorphismes obtenus ont été filtrés avec "GATK variantFiltration". Les polymorphismes conservés avaient pour caractéristiques une profondeur comprise entre 10 et 2000 et une qualité supérieure à 200. De plus, les polymorphismes regroupés en *cluster* ont été éliminés (moins de trois polymorphismes dans une fenêtre de 10 pb) ainsi que ceux détectés sur des *reads* s'étant alignés à plusieurs positions sur le génome. À l'issue de ces filtres, les individus présentaient environ cinq millions de SNP par rapport à la séquence de Nipponbare.

J'ai bénéficié des fichiers vcf résultant de cette analyse sur lesquels j'ai pu appliquer des filtres supplémentaires selon les besoins des différentes analyses.

5 Génotypage

Afin d'établir les génotypes des différentes accessions dans les régions d'intérêt, nous avons utilisé plusieurs marqueurs moléculaires. Ces derniers ont notamment servis aux typages des individus de la population de cartographie de la résistance de Tog5307, mais également au criblage d'accessions de la collection pour la présence d'allèles bien caractérisés.

Trois types de marqueurs moléculaires différents ont été utilisés :

- des marqueurs INDEL basés sur des polymorphismes d'insertion/délétion de 20 à 50 paires de bases,
- des marqueurs CAPS (*cleaved amplified polymorphic sequence*) pour lesquels le polymorphisme est révélé par une digestion enzymatique spécifique d'un génotype,
- des marqueurs dCAPS (*derived cleaved amplified polymorphic sequence*), similaires aux marqueurs CAPS, mais dont le site de restriction permettant de repérer le polymorphisme est créé avec une amorce présentant un mésappariement avec le brin matrice.

La majorité des amorces a été dessinée à l'aide du logiciel Primer3 (Untergasser et al., 2007) puis leur appariement unique a été vérifié sur le génome de Nipponbare avec l'outil Primer Blaster (Droc et al., 2009). Les amorces des marqueurs dCAPS ont été dessinées avec l'outil dCAPS Finder 2.0 (Neff et al., 2002).

Un protocole de PCR directe, réalisé directement à partir de fragments de feuilles, a été utilisé sur certaines plantes (Terra PCR Direct Polymerase Mix de Clontech). La PCR a été réalisée dans un volume final de 10 μ L contenant 1X de tampon Terra PCR Direct, 0,3 mM de chaque amorce, et 1,25 U de Terra DNA polymerase. Les conditions PCR ont été programmées pour une dénaturation initiale de 2 minutes à 98 °C, suivie par 30 cycles à 98 °C 10 secondes, 60 °C 15 secondes et 68 °C une minute par kilobase, puis une extension finale de deux minutes à 68 °C. Dans le cas des régions riches en GC, l'étape à 60 °C était supprimée.

L'ADN des plantes a également pu être extrait selon le protocole décrit dans Edwards et al. (1991) ou via le kit d'extraction DNeasy Plant Mini Kit (Qiagen, Hilden, Allemagne). Les amplifications PCR ont, dans ce cas, été réalisées avec la GoTaq G2 (Promega, Madison, USA) dans un volume final de 15 μ L (20 ng d'ADN matrice, 1X de tampon GoTaq, 1,5 mM de $MgCl_2$, 0,2 mM de chaque dNTP, 0,25 μ M de chaque amorce, et 0,3 U de polymérase GoTaqG2). Les conditions étaient un cycle initial de 3 minutes à 94 °C, suivi de 30 cycles de 30 secondes à 94 °C, 45 secondes à une température adaptée en fonction du couple d'amorces et une minute par kilobase à 72 °C, avec une extension finale de 5 minutes à 72 °C.

Après amplification des marqueurs CAPS ou dCAPS, 10 μ L d'un mélange tampon-enzyme de restriction (2 μ L de tampon de réaction, 0,4 μ L de l'enzyme de restriction appropriée, 0,2 μ L d'albumine sérique bovine et 7,4 μ L d'eau ultrapure) a été ajouté à 10 μ L de produit PCR et incubé 2h à 37 °C.

L'électrophorèse des produits de digestion ou des produits PCR a été réalisée sur agarose et révélée avec du bromure d'éthidium.

Certains fragments ont également été séquencés par séquençage Sanger pour révéler les polymorphismes présents. Ce génotypage a été utilisé lorsque plusieurs polymorphismes d'intérêt étaient présents dans une région de taille limitée (cas de *RYMV1*) ou lorsque le faible nombre d'accessions à cribler ne justifiait pas la conception de nouveaux marqueurs. Ce séquençage a été sous-traité auprès de Beckman Coulter Genomics.

L'ensemble des couples d'amorces utilisés en tant que marqueurs ou pour l'amplification de fragments à séquencer est détaillé dans le tableau 2.2.

TABLE 2.2 – Couples d’amorces utilisés pour le génotypage aux gènes de résistances.

Objectif	amorce sens	amorce anti-sens
Amplification MIF4G <i>RYMV1</i>	RYMV1-F16b : CCTTGGTCAGCTAGAAGAGGCA	RYMV1-R16 : CCTCGGTACAACCAAGAGAC
Séquençage MIF4G <i>RYMV1</i>	RYMV1-R15 : CTCTTCACGTCGAGGCACCCA	-
Marqueur CAPS <i>RYMV2</i> (CfoI)	CAPS1-F : CGAAGGACATTCCGAGAGG	CAPS1-R : CACGACGTCGCTTCGCCTCC
Marqueur dCAPS <i>RYMV3</i> (HaeIII)	LRR_RYMV3-F : CTACACAAACTCCGCCGTCT	LRR_RYMV3-R : AGTACAGTGAAGAAATACATGGGC
Amplification fragment 1 <i>RYMV2</i>	ID251179F : GATGATTCTGTGGCTGTGGA	5bloc3R : GAAAGAGACAGGGTGGGAAA
Amplification fragment 2 <i>RYMV2</i>	5ex1F : CTCGCCTGCTTCTTCGAG	252430F : CGAAGGACATTCCGAGAGG
Amplification fragment 3 <i>RYMV2</i>	5ex4F : TTATGTCAGCGTGCTTTGGA	5ex5R : CCGGCGAACCCAGTATGGTAAC

CHAPITRE

3

Cartographie d'un gène majeur de
résistance

Chapitre 3 : Cartographie d'un gène majeur de résistance

1 Introduction

Avant le début de cette étude, deux gènes de résistance contre le RYMV avaient été identifiés. Tous deux contrôlent des résistances élevées récessives. *RYMV1* code pour un facteur d'initiation de la traduction (eIF(iso)4G1), pour lequel quatre allèles de résistance distincts ont été décrits, correspondant à des substitutions ou de courtes délétions dans le domaine central conservé de la protéine (Albar et al., 2006; Thiémélé et al., 2010). *RYMV2* correspond à un allèle nul du gène *CPR5-1*, nommé par homologie avec le gène d'*A. thaliana constitutive expressor of pathogenesis related protein-5*, et est supposé coder un régulateur négatif des mécanismes de défense (Orjuela et al., 2013).

Cependant, Thiémélé et al. (2010) ont identifié, dans l'espèce *O. glaberrima*, des sources de résistance ne présentant aucun des allèles de résistance connus sur les gènes *RYMV1* et *RYMV2*, suggérant l'existence de nouveaux allèles de résistance à ces gènes ou de nouveaux gènes de résistance. Parmi ces accessions, l'accession Tog5307 a été étudiée plus en détail et une analyse génétique conventionnelle préliminaire a suggéré que sa résistance était indépendante de *RYMV1* et *RYMV2* (Thiémélé, 2012). L'objectif de cette première partie de ma thèse était donc d'identifier le déterminisme génétique de cette nouvelle source de résistance.

Ce travail a fait l'objet d'un article soumis dans la revue *Theoretical and Applied Genetics* et intégralement présenté ci-dessous. Cet article contient, dans une première partie, une analyse de la spécificité de la résistance de Tog5307 face à des isolats représentatifs de la diversité du RYMV, à laquelle je n'ai contribué que ponctuellement. Il présente ensuite l'approche de cartographie génétique puis d'analyse génomique que j'ai menée, et qui a permis l'identification d'un gène candidat. J'ai ensuite analysé la potentialité de la présence de ce gène dans l'accession Tog5672, possédant un allèle de résistance sur le gène *RYMV1* (Albar et al., 2006) mais supposée contenir un second gène de résistance (Thiémélé et al., 2010).

Dans la conclusion de ce chapitre, je résumerai les résultats de cet article ainsi que les principaux points de discussions, complétés par des résultats d'expérimentations menées dans le cadre de ma thèse ou d'autres projets de l'équipe.

2 Fine mapping of *RYMV3* : a new resistance gene to *Rice yellow mottle virus* from *Oryza glaberrima*

Hélène Pidon¹, Alain Ghesquière¹, Sophie Chéron¹, Souley Issaka^{3,4}, Eugénie Hébrard², François Sabot¹, Olufisayo Kolade^{1,3}, Drissa Silué³, Laurence Albar^{1*}

¹ DIADE UMR IRD/UM, 911 Avenue Agropolis BP 64501, F-34394 Montpellier Cedex 5, France

² IPME UMR IRD/CIRAD/UM, 911 Avenue Agropolis BP 64501, F-34394 Montpellier Cedex 5, France

³ Africa rice Center, 01 BP 2031, Cotonou, Benin

⁴ FSAE, Université de Tillabéri, BP 175 Tillabéri, Niger

* Corresponding author

Abstract *Rice yellow mottle virus* (RYMV) disease is a serious constraint to the cultivation of rice in Africa and selection for resistance is considered to be the most effective management strategy. The aim of this study was to characterize the resistance of Tog5307, a highly resistant accession belonging to the African cultivated rice species (*Oryza glaberrima*), that has none of the previously identified resistance genes to RYMV. The specificity of Tog5307 resistance was analyzed using 18 RYMV isolates. While three of them were able to infect Tog5307 very rapidly, resistance against the others was effective despite infection events attributed to resistance-breakdown or incomplete penetrance of the resistance. Segregation of resistance in an interspecific backcross population derived from a cross between Tog5307 and the susceptible *Oryza sativa* variety IR64 showed that resistance is dominant and is controlled by a single gene, named *RYMV3*. *RYMV3* was mapped in an approximately 15 kb interval in which two candidate genes coding for a putative transmembrane protein and a CC-NBS-LRR, were annotated. Sequencing revealed non-synonymous polymorphisms between Tog5307 and the *O. glaberrima* susceptible accession CG14 in both candidate genes. An additional resistant *O. glaberrima* accession, Tog5672, was found to have the genotype of Tog5307 for the CC-NBS-LRR but not for the putative transmembrane proteins. Analysis of the cosegregation of Tog5672 resistance with the *RYMV3* locus suggests that *RYMV3* is also involved in Tog5672 resistance, thereby supporting the CC-NBS-LRR gene as the best candidate for *RYMV3*.

Keywords fine mapping, RYMV, rice, *Oryza glaberrima*, resistance, NB-LRR

Authors Contributions AG, LA, SC developed the mapping populations. IS, DS, EH provided and characterized the viral material. HP, LA, AG conceived and designed the experiments. HP, SC, OK performed phenotyping and genotyping of the mapping

population. HP, FS performed the sequence analysis. HP, LA analyzed the data and wrote the paper. FS, EH revised the paper.

Key Message A new resistance gene against *Rice yellow mottle* virus was identified and mapped in a 15 kb interval. The best candidate gene is CC-NBS-LRR.

2.1 Introduction

Rice is the staple food of more than half the world population. Africa produced 24 million tons in 2004 and in 2013, which is only 3.5 % of world rice production but Africa accounts for 29.6 % of world imports (<http://faostat3.fao.org/>). Despite regular improvement, yields remain low and are not sufficient to feed the growing population.

Among the various abiotic and biotic stresses that reduce rice yield, *Rice yellow mottle virus* (RYMV) disease is a major cause of losses in sub-Saharan Africa (Issaka et al., 2012; Kam et al., 2013; Ochola and Tusiime, 2011). RYMV is a *Sobemovirus* endemic to Africa and is well established in all the sub-Saharan countries where rice is cultivated. Extensive characterization of its diversity has revealed different viral strains, i.e., genetic groups, with gradual differentiation from East Africa, where it originated, to West Africa (Pinel-Galzi et al., 2015). The epidemiology of RYMV is complex (Traoré et al., 2009) and the fight against the disease has thus been directed towards resistance genes.

Resistance to RYMV has been investigated in the two cultivated species of rice : *Oryza sativa*, originating from Asia and cultivated worldwide, and *O. glaberrima*, more recently domesticated and only cultivated locally in West Africa (Linares, 2002). Several accessions were found, mostly among the *O. glaberrima* species, which showed no symptoms after RYMV inoculation, nor detectable virus content with enzyme linked immuno-sorbent assay (ELISA), (Thiémmélé et al., 2010). So far, two genes conferring high resistance to RYMV have been identified. The first one, *RYMV1*, encodes a translation initiation factor, eIF(iso)4G1 (Albar et al., 2006). Four recessive resistance alleles have been described : *rymv1-2* in two resistant *O. sativa* accessions, *rymv1-3*, *-4* and *-5* in *O. glaberrima* accessions. As observed for eIF4E or eIF(iso)4E translation initiation factors involved in resistance against plant viruses (Sanfaçon, 2015), eIF(iso)4G1 interacts with the viral VPg in susceptible plants, while resistance is gained by the rupture of the interaction (Hébrard et al., 2010). The second resistance gene, *RYMV2*, also named *CPR5-1*, is homologous to *Arabidopsis thaliana constitutive expressor of pathogenesis-related genes-5* (*CPR5*). The resistance allele on the *RYMV2* gene is recessive and corresponds to a loss of function, it was discovered in seven *O. glaberrima* accessions (Orjuela et al., 2013). Studies of its homolog in *A. thaliana* (Boch et al., 1998; Bowling et al., 1997; Wang et al., 2014d) suggest that *CPR5-1* is a negative regulator of defense mechanisms.

Even if using genetic resistance is efficient and environment friendly, the durability of resistance genes is still a challenge as virus variants able to overcome a resistance gene

emerge frequently (Kobayashi et al., 2014). The *rymv1-2* resistance allele was recently transferred to high yielding varieties (Ndjiondjop et al., 2013), but none of the resistance genes or alleles against RYMV have yet been deployed in the field, so their durability is unknown. However, the resistance breakdown of accessions carrying *rymv1-2* and *rymv1-3* alleles has already been observed and characterized in experimental conditions. Mutations associated with *RYMV1* resistance breaking take place in the central domain of the VPg (Pinel-Galzi et al., 2007; Traoré et al., 2010). In addition, the ability of the different isolates to evolve toward virulence is determined by a polymorphism at amino acid 49 of the VPg and is associated with a rice species dependent selective advantage in susceptible accessions (Poulicard et al., 2012). The geographical distribution of this polymorphism in RYMV diversity may affect the durability of the resistance conferred by *rymv1-2* or *rymv1-3*.

Of the resistant accessions identified in *O. glaberrima* species by Thiémélé et al. (2010), some do not have a known resistant allele on either *RYMV1* or *RYMV2* and therefore represent diversity to be exploited against RYMV. We characterized the specificity of Tog5307 resistance using isolates representative of RYMV diversity and here report the fine mapping of the *RYMV3* gene, which is involved in the resistance of Tog5307 and of Tog5672.

2.2 Materials and Methods

2.2.1 Plant material

Four rice accessions were used in this study. Tog5307 and Tog5672 are *O. glaberrima* accessions resistant to RYMV. Tog5672 has the resistance allele *rymv1-4* for *RYMV1* (Albar et al., 2006) but is suspected to carry an additional unknown resistance gene (Thiémélé et al., 2010). *O. sativa* spp *indica* IR64 variety and *O. glaberrima* Tog6206 accession are highly susceptible to RYMV. An interspecific population derived from the Tog5307 x IR64 cross was developed to map Tog5307 resistance. Successive backcrosses, with IR64 as the recurrent parent and phenotypic selection for RYMV resistance at each step, were performed to introgress Tog5307 resistance into the IR64 genetic background. Independent lineages were derived from four resistant BC1F1 progenitors. The BC3F2 and BC3F3 generations were used for segregation analysis and fine genetic mapping of the resistance gene. The genetic basis of Tog5672 resistance was analyzed in a F2 population derived from the Tog5672 x Tog6206 cross.

2.2.2 Viral material

Eighteen RYMV isolates were used in this study. Their characteristics are listed in Table 3.1. Isolates were classified according to their country of origin, their viral strain, i.e. genetic group, and the presence at position 49 of their VPg of a glutamic acid (E49) or a threonine (T49). The sample included nine T49 isolates from West Africa

representing strains S1 and S2, four E49 isolates from West Africa representing strain S1, and five E49 isolates from East Africa, representing strains S4, S5 and S6. Isolates were propagated on IR64 susceptible plants and stored at -20°C . Four new isolates were included in the viral corpus. They were collected on susceptible *O. glaberrima* varieties in Togo and in susceptible *O. sativa* varieties in Niger. Total RNA was extracted with the RNeasy Plant Mini kit (Qiagen, Germany). The 720 nucleotide coat protein (CP) gene was amplified by reverse transcription-polymerase chain reaction (RT-PCR) (Fargette et al., 2002b) and the sequences were compared to the set of CP sequences for phylogenetic reconstruction using the maximum-likelihood method (Traoré et al., 2010). In addition, the VPg and its 5' and 3' neighboring regions of these isolates were sequenced as done earlier (Hébrard et al., 2006). All four isolates belonged to strain S1.

TABLE 3.1 – Characteristics of the RYMV isolates used in this study. The country of origin, the amino acid at position 49 of the VPg, and the strain, i.e. genetic group, are indicated for each isolate.

Name	Country of origin	VPg49 a. a.	Strain	Reference
BF1	Burkina Faso	T	S2	(Pinel et al., 2000)
CIa	Côte d'Ivoire	T	S2	(Fargette et al., 2004)
CI2	Côte d'Ivoire	T	S1	(Pinel et al., 2000)
CI3	Côte d'Ivoire	T	S1	(Pinel et al., 2000)
CI8	Côte d'Ivoire	T	S2	(Pinel et al., 2000)
CI4	Côte d'Ivoire	E	S1	(Pinel et al., 2000)
Ni1	Nigeria	T	S1	(Pinel et al., 2000)
Tg274	Togo	T	S1	this study
Ng106	Niger	T	S1	(Traoré et al., 2010)
Ng119	Niger	T	S1	this study
Ng109	Niger	E	S1	(Traoré et al., 2010)
Ng111	Niger	E	S1	this study
Ng113	Niger	E	S1	this study
Mg1	Madagascar	E	S4	(Pinel et al., 2000)
Tz5	Tanzania	E	S4	(Fargette et al., 2004)
Tz11	Tanzania	E	S6	(Abubakar, 2003)
Tz201	Tanzania	E	S4	(Fargette et al., 2008b)
Tz211	Tanzania	E	S5	(Fargette et al., 2008b)

2.2.3 Phenotypic evaluation

The resistance spectrum was characterized in four different trials conducted in similar conditions. The plants were cultivated in greenhouse conditions in trays holding 28 individual 350 ml pots. Between 40 and 56 Tog5307 plants and eight IR64 control plants were inoculated with each isolate. Inoculum was prepared by grinding IR64 infected leaves in 0.1M phosphate buffer (0.1g/ml) and adding 600-mesh carborundum. Plants were mechanically inoculated approximately two weeks after sowing by rubbing the inoculum on leaves. Symptoms were observed two, four and six weeks after inoculation.

Samples from the last emerged leaf were collected four and six weeks after inoculation and virus content was estimated by direct double antibody sandwich (DAS-) ELISA, as described by Ndjioudjop et al. (1999), using a polyclonal antiserum directed against a RYMV isolate from Madagascar (N'guessan et al., 2000). About 1 mg of fresh leaves in 1 ml buffer was used. Samples were considered negative when the OD values were less than three times the negative control. Considering that recovery from the infected state is highly unlikely, ELISA tests were first performed on samples collected four weeks after inoculation on symptomatic plants, and, if negative, on samples collected six weeks after inoculation ; in the case of non-symptomatic plants, ELISA tests were first performed on samples collected at six weeks and, if positive, on samples collected at four weeks .

For the genetic analysis of Tog5307 and Tog5672 resistance, culture and inoculation conditions were the same as described above. BF1 isolate was used. Resistance was assessed on the basis of the intensity of the symptoms and confirmed, when necessary, by ELISA tests. Symptoms were monitored between 10 days and three weeks after inoculation. Susceptible plants developed clear yellow mottles 10 to 15 days post-inoculation, whereas resistant plants had no symptoms at all.

2.2.4 Genomic resources

Three different genome sequences were used in this study : the *O. sativa* spp. *japonica* Nipponbare release Os-Nipponbare-Reference-IRGSP-1.0 (Kawahara et al., 2013), the *O. sativa* spp. *indica* IR64 sequence Os-IR64-Draft-CSHL-1.0 (Schatz et al., 2014) and the *O. glaberrima* CG14 genome assembly *Oryza_glaberrima_AGI1.1* (Wang et al., 2014c). The IR64 scaffold corresponding to the resistance gene region was identified with Blast 2.2.31+ (Camacho et al., 2009). Annotations on CG14 refer to the 2011-05-AGI conducted by Munich Information Center for Protein Sequences and browsed through Gramene release 49 (<http://www.gramene.org/>). Sequence comparison in the candidate region was visualized with Artemis software (Rutherford et al., 2000). Tog5307 and Tog5672 sequences at candidate open reading frames (ORFs) were obtained by Sanger sequencing of PCR amplification fragments at Beckman Coulter Genomics (Takeley, UK).

2.2.5 Development of molecular markers and PCR procedure

The single nucleotide polymorphism (SNP) chip described in Orjuela et al. (2014) was used for the first step of Tog5307 resistance mapping.

PCR markers based on polymorphisms between CG14 and Nipponbare or IR64 sequences were developed for the fine mapping of Tog5307 resistance. A marker specific to the *rymv1-4* allele was also developed to screen the Tog5672 x Tog6206 F2 population. Primers were designed using Primer3 (Untergasser et al., 2007) and checked for specificity in the Nipponbare genome using Primer blaster tool (Droc et al., 2009). Most of the

markers were either indel markers or cleaved amplified polymorphic sequence (CAPS) markers. One dCAPS marker was developed based on the dCAPS Finder 2.0 software (Neff et al., 2002). The complete characteristics of the markers are listed in Table 3.2.

TABLE 3.2 – Characteristics of the markers developed for mapping and segregation analysis. For CAPS and dCAPS markers, the restriction enzyme used to reveal polymorphism is in parentheses. All the markers are located on chromosome 11, except SNP_rymv1-4, which was designed on the *RYMV1* gene, on chromosome 4.

Name	Marker type	Forward primer (5'-3')	Reverse primer (5'-3')
C5	CAPS (HindII)	TCACTGACCCGCTTCTATGG	CGCTTGGATCATGCTTGCAA
C11	CAPS (HindII)	AACAAGCTTTGAATCGTTCGT	ACTCTCCGTTTTGGTCATATTACA
C10	CAPS (CfoI)	CTTTCCCGGACCTGTTCTTGA	ACAGAAGCAACGGTACAGCC
M6	InDel	CTGGTGACCCATCTTGCTT	TGTATTTACAGGGGCGGACC
C34	CAPS (AluI)	CTTTAGGCGCTGCGTACATC	GAGAGGGGAGGGGAGATTGT
C31	CAPS (HindII)	ATCTACCGGGGACAACCTCCA	GAGGCGCATCTTCACTCAGT
C16	CAPS (PstI)	TAACGGCATCCCAGCACATT	CCCCTTCTTCTCCAGCAACA
C26	CAPS (CfoI)	GCCTATCCATGTCCACACCA	GGTGACCATGATTGCCACCT
M4	CAPS (CfoI)	CCCTGCCTCCCTGATTGTTT	GCAAGGTAGAGTGTGTGCAA
LRR_RYMV3	dCAPS (HaeIII)	CTACACAAACTCCGCCGTCT	AGTACAGTGAAGAAATACATGGGC
SNP_rymv1-4	CAPS (EarI)	GAGCCACCTTCTGTCCGATG	AGTAGCTCACCAATTAGACGGA

Genotyping was performed by direct PCR amplifications on leaf extracts or by standard PCR amplifications on extracted DNA. PCR conditions are described in detail in Orjuela et al. (2013). After amplification of CAPS or dCAPS markers, 10 μ L of buffer-restriction enzyme mixture (2 μ L of reaction buffer, 0,4 μ L of appropriate restriction enzyme, 0,2 μ L of bovine serum albumen, and 7,4 μ L of double-distilled H₂O) was added to 10 μ L of PCR product and incubated at 37 °C for 2 h. Electrophoresis of PCR or digestion products was performed in agarose and revealed with ethidium bromide.

2.3 Results

2.3.1 Tog5307 resistance spectrum

Tog5307 accession was inoculated with 18 isolates of RYMV, selected as representative of the genetic diversity of the virus in Africa.

Clear symptoms were observed 12 days after inoculation on IR64 control plants, whatever the isolates, and on all the Tog5307 plants inoculated by Ng119 and Tz211 isolates. After Ni1 inoculation, symptoms were also observed on 60 % of the plants 12 days after inoculation, and on all the plants but one, four weeks after inoculation. ELISA tests on randomly selected Tog5307 plants confirmed the presence of the virus. For the other isolates, almost all the Tog5307 inoculated plants were symptom free two weeks after inoculation (Figure 3.1). Disease incidence, i.e. the percentage of diseased plants, ranged from 0 to 25 % four weeks after inoculation and from 0 to 45 % six weeks after inoculation. Disease incidences at these two dates were highly correlated ($R_2 = 0.95$).

Disease incidences with Ni1, Ng119 and Tz211 isolates ranged between 98 or 100 % four weeks after inoculation and were significantly different from those obtained with

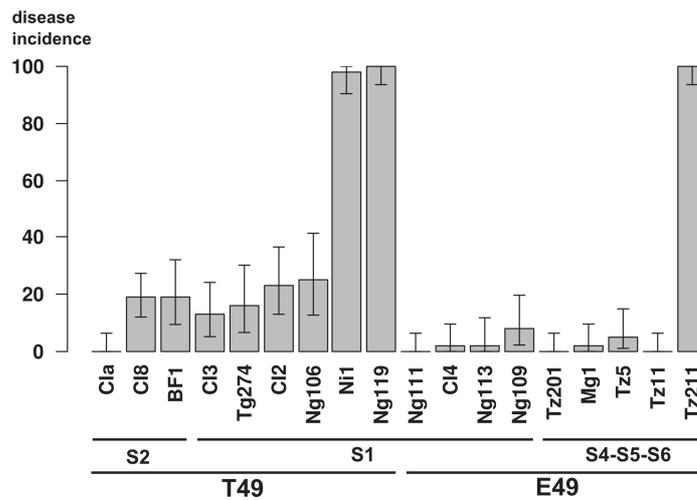


FIGURE 3.1 – Disease incidence, i.e. percentage of diseased plant on Tog5307 accessions four weeks after inoculation. Plants were inoculated with RYMV isolates representing different strains of the virus (S1, S2, S4, S5, S6) and distinguished by a T or a E amino acid at position 49 of the VPg. The error bars represent 95 % confidence intervals.

other isolates (Fisher exact test, $P < 0.001$), revealing a strong specificity of interactions between Tog5307 resistance and RYMV isolates. Among other isolates, four weeks after inoculation, no infection cases were observed with isolates Cl1a, Ng111, Tz201 and Tz11, and a single case of infection each with isolates Cl4, Ng113 and Mg1. These isolates differed significantly from BF1, Cl2, Cl8, and Ng106, which showed over 18 % incidence, while the incidence of Cl3, Ng109, Tg274 and Tz5 was between the two.

As the T/E polymorphism at codon 49 of the VPg is associated with *RYMV1* resistance-breakdown ability and with a selective advantage of T49 isolates on *O. glaberrima* susceptible accessions (Poulicard et al., 2012), its association with disease incidence on Tog5307 was analyzed. Considering all the isolates except Tz211, Ng119 and Ni1, significantly different incidences ($P < 10^{-3}$, Fisher exact test) were observed between T49 isolates (average incidence 41 %) and E49 isolates (7 %), suggesting that E/T polymorphism also plays a role in the interaction between RYMV isolates and Tog5307 resistance.

2.3.2 Inheritance of Tog5307 resistance

The inheritance of Tog5307 resistance was analyzed in populations derived from the Tog5307 x IR64 cross and subsequent backcrosses using IR64 as the recurrent parent. Resistance phenotype was assessed by inoculation with isolate BF1.

Preliminary observations were made on BC1F1, BC2F1 and BC3F1 hybrids in the course of the introgression process. Two contrasted phenotypes were observed. IR64 and some hybrids exhibited very clear symptoms as early as 10 to 14 days after

inoculation, while Tog5307 and other hybrids showed no symptoms even several weeks after inoculation.

The segregation of resistance was evaluated in five BC3F3 families derived from four independent BC1F1. Between 16 and 24 plants per family were phenotyped. A total of 82 resistant and 32 susceptible plants were identified, which fit a 3:1 resistant/susceptible segregation pattern ($\chi^2 = 0.57$; $P = 0.45$). This segregation pattern was also supported by each individual family (Table 3). This result is in agreement with monogenic and dominant inheritance of resistance.

TABLE 3.3 – Segregation of resistance in five BC3F3 families derived from the (IR64 x Tog5307) cross and subsequent backcrosses on IR64. The names of families refer to their lineage, and the BC1F1 and BC2F1 plants from which they are derived are listed in the first three columns. The results of the Chi squared test used to determine the goodness-of-fit to a 3:1 ratio, expected for a dominant monogenic gene, and the corresponding p-value (P) are given.

Lineage			Number of resistant plants	Number of susceptible plants	χ^2	P
BC1 line	BC2 line	BC3F3 family				
1	1.2	1.2.1.1	14	9	2.44	0.12
1	1.3	1.3.1.C	18	6	0	1
2	2.5	2.5.6.2	16	8	0.89	0.35
4	4.1	4.1.3.1	20	4	0.89	0.35
6	6.5	6.5.2.1	14	5	0.02	0.89

2.3.3 Identification of a candidate locus for Tog5307 resistance on chromosome 11

To identify the genomic region carrying the Tog5307 resistance gene, 18 BC1F1 plants resistant to RYMV were genotyped with the SNP chip described in Orjuela et al. (2014). Among the 384 SNP markers on the chip, 119 were polymorphic in the target population. As expected, BC1F1 were either heterozygous, or homozygous for IR64 alleles. Two markers showed a strong segregation distortion towards the heterozygous state (17 out of 18 individuals) (Figure 3.2). This segregation strongly diverged from the 1:1 ratio ($\chi^2 = 14.2$; $P = 1.6 \times 10^{-4}$), as expected for markers associated with resistance. These markers were on chromosome 6 at position 2,346,023 on the *O. sativa* Nipponbare reference genome and on chromosome 11 at position 26,619,446. On chromosome 6, the marker was located on the interspecific sterility locus *S1* (Garavito et al., 2010), which is known to induce major segregation distortions toward the *O. glaberrima* allele in interspecific crosses between *O. sativa* and *O. glaberrima* (Lorieux et al., 2000). Thus, the distortion observed on chromosome 6 is likely to be due to *S1*. As no sterility locus has ever been reported in this region of chromosome 11 (Garavito et al., 2010), the second distortion is assumed to result from the resistance gene.

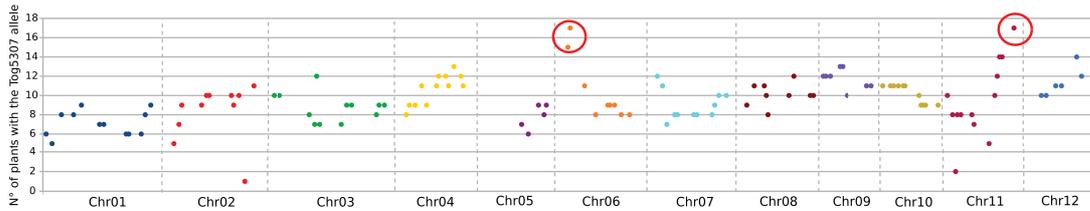


FIGURE 3.2 – Genomic regions in linkage disequilibrium with resistance in a population of 18 resistant BC1F1 plants from the cross (IR64 x Tog5307) x IR64. Plants were challenged with BF1 isolate and genotyped with the SNP chip described in Orjuela et al. (2014). The number of heterozygous BC1F1 plants at each marker over the genome is indicated in ordinates. The two encircled regions on chromosomes 6 and 11 show segregation distortion toward the heterozygous state.

Since none of the previously identified resistance genes (Albar et al., 2006; Orjuela et al., 2013) or QTLs (Boisnard et al., 2007) against RYMV cosegregated with resistance in this population, the resistance gene of Tog5307 was considered to be new and was named *RYMV3*, for *Resistance-to-yellow-mottle-virus-3*.

2.3.4 Fine mapping of *RYMV3*

Mapping of *RYMV3* was confirmed and refined on BC3F2 and BC3F3 plants. In the first step, 37 resistant BC3F2 plants were genotyped with 13 markers designed in the candidate region on chromosome 11. The cosegregation of resistance with a region comprised between markers C5 and M4, corresponding to an 880 kb-region on the Nipponbare genomic sequence was confirmed.

In the second step, 1363 BC3F3 plants were genotyped at markers C5 and M4 (Figure 3.3). Fifty-eight recombinant individuals were identified and genotyped with additional markers in the interval. The phenotyping of the most informative recombinants was performed on their progenies, with 10 to 16 plants tested per progeny. The interval containing *RYMV3* was finally bounded by markers M6 and C16. Two recombination events were located between marker M6 and *RYMV3*, and one event between *RYMV3* and C16.

2.3.5 *RYMV3* candidate genes

The length of the M6-C16 region containing *RYMV3* is 15 kb, 17 kb and 42 kb in the CG14, IR64 and Nipponbare genomic sequences, respectively. The differences in size were mainly explained by 2.1 kb and 3.5 kb insertions/deletions between the CG14 and IR64 genomes, and insertions of 2.4 kb, 6.7 kb and 16 kb in the Nipponbare genome that were absent in the IR64 genome (Figure 3.4).

Two ORFs were annotated in the CG14 genome in the M6-C16 interval (Figure 3.5). The first one (ORGLA11g0175700) was labeled as a gene coding for a hypothetical protein. Protein sequence analysis (<http://www.ebi.ac.uk/interpro/>) revealed

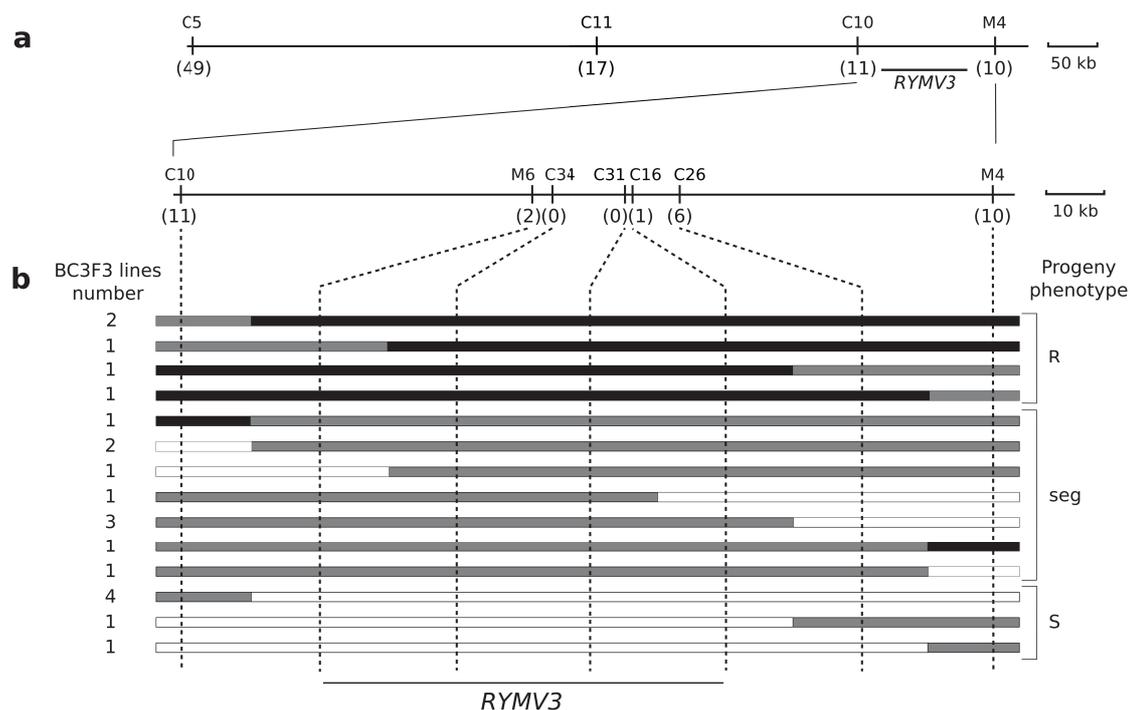


FIGURE 3.3 – Physical map of the *RYMV3* locus. (a) Fine mapping of the *RYMV3* locus in BC3F3 families. The number of recombinants between the corresponding marker and the gene is in parentheses. (b) Genotypes of recombinants and phenotype in their selfing progenies (R : resistant, seg : segregation of resistance, S : susceptible). Tog5307 homozygous, IR64 homozygous and heterozygous states are represented by black, white and gray bars, respectively. The number of recombinants corresponding to each genotype/phenotype pattern is indicated on the left. Recombination events indicate that *RYMV3* is bounded by M6 and C16 markers.

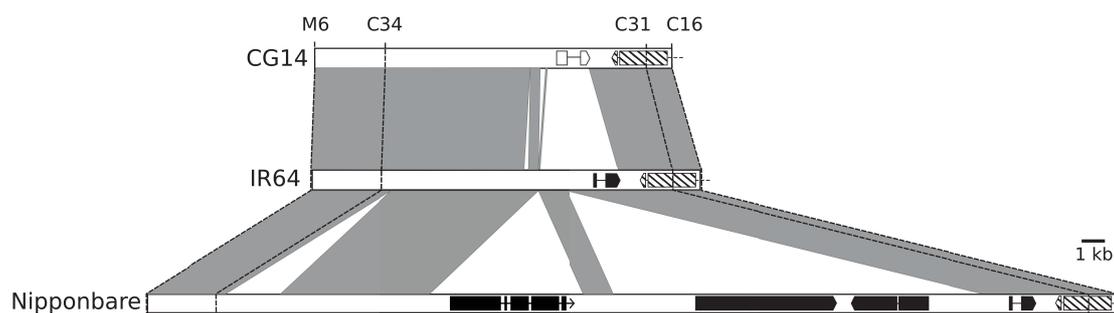


FIGURE 3.4 – Collinearity between the CG14, IR64 and Nipponbare genomes in the genomic region containing *RYMV3*, visualized with Artemis software (Rutherford et al., 2000). Collinear regions are in gray. Genes annotated in the different genomes are indicated by boxes representing exons. Annotations in the IR64 genome were obtained from Nipponbare ones. The white gene is ORGLA11g0175700. The hatched gene is the CC-NBS-LRR gene (ORGLA11G75800 on the CG14 genome and LOC_Os11g43700 on the Nipponbare genome). It is only partially included in the region M6-C16 containing *RYMV3*. Markers used for mapping are shown with dashed lines.

a transmembrane structure with a 69 amino acid (aa) cytoplasmic domain, a 28 aa transmembrane domain and a 129 aa non-cytoplasmic domain, but no known conserved functional domains. This ORF was absent in the Nipponbare and IR64 sequences. However, two homologs of ORGLA11g0175700 were identified in the CG14 genome (ORGLA11g0103500 and ORGLA11g0189200) and in the Nipponbare genome (LOC_Os11g25530 and LOC_Os11g47750) on chromosome 11, 8.3 Mb upstream and 1.5 Mb downstream of locus *RYMV3* in the CG14 genome.

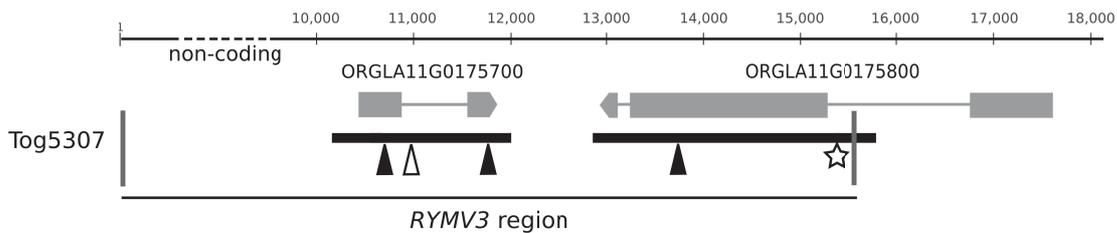


FIGURE 3.5 – Annotation of the *RYMV3* candidate region and polymorphisms detected between CG14 and Tog5307. The regions sequenced in Tog5307 are indicated by thick black bars. Black arrows represent non-synonymous SNPs, the white arrow a synonymous SNP, and the white star a deletion of 33 bp. ORGLA11G0175700 is a putative transmembrane gene. ORGLA11G0175800 is a typical CC-NBS-LRR resistance gene.

The second ORF (ORGLA11g0175800) was only partially located within the M6-C16 region, as C16 is located on the first intron. It is present in the Nipponbare sequence where it is named LOC_Os11g43700. The encoded protein showed a coiled-coil (CC), a nucleotide binding site (NBS) and a leucine rich repeat (LRR) domain, characteristic of the large family of resistance genes, and consequently named CC-NBS-LRR genes (Takken and Goverse, 2012).

These ORFs were sequenced by Sanger methodology in the Tog5307 accession. Three SNPs were detected between Tog5307 and the susceptible accession CG14 on gene ORGLA11g0175700. One of them was intronic, the other two resulted in a glycine to arginine change at aa position 28 in the cytoplasmic domain, and a proline to leucine change at aa position 212 in the non-cytoplasmic domain. Two polymorphisms were identified in the part of the ORGLA11g0175800 gene located within the M6-C16 region. The first one was a 33 bp deletion in the first intron, and the second one a non-synonymous SNP in the LRR domain inducing a conversion of a lysine into an arginine at aa position 779.

2.3.6 Tog5672 second resistance gene is linked to *RYMV3*

Tog5672 has the *rymv1-4* resistance allele of the *RYMV1* gene (Albar et al., 2006) but segregation analysis suggested that this accession contains another resistance gene (Thiémélé et al., 2010). As no evidence for a resistance allele on the *RYMV2* gene was reported by Orjuela et al. (2013), we investigated if Tog5672 carries *RYMV3*. The sequence of the *RYMV3* candidate genes in Tog5672 was assessed by Sanger sequencing.

Out of the five polymorphisms identified between Tog5307 and CG14, Tog5672 displayed the Tog5307 genotype only for the SNP located at amino acid position 779, in the LRR domain of the CC-NBS-LRR gene.

A cross was performed between Tog5672 and Tog6206, a susceptible *O. glaberrima* accession. F2 plants were inoculated with the BF1 isolate and cosegregation of the resistance phenotype with both *rymv1-4* and the *RYMV3* locus was assessed by the CAPS marker SNP_rymv1-4, targeting a polymorphism characterizing *rymv1-4* allele, and the dCAPS marker LRR_RYMV3, targeting the polymorphism in the LRR domain of the candidate gene for *RYMV3*. As expected, plants homozygous for *rymv1-4* turned out to be resistant. Of the others, the seven plants homozygous for the Tog6206 allele at the LRR_RYMV3 marker showed clear symptoms and had positive ELISA values, whereas the 12 plants homozygous for the Tog5672 allele showed no symptoms and did not have positive ELISA values. Independently from being homozygous or heterozygous for the susceptibility allele at *RYMV1*, most *RYMV3* heterozygous plants had a clearly resistant phenotype, but eight out of 20 showed symptoms and/or had low but positive ELISA scores, suggesting partial resistance. Results obtained on the homozygous plants for the LRR_RYMV3 marker clearly showed that the second resistance determinant of Tog5672 is linked to *RYMV3*. In addition, the susceptible heterozygous plants suggest more complex genetic determinism.

Some F2 plants were also inoculated with Ng119 isolate. All the plants that were non-homozygous for *rymv1-4* and heterozygous (eight plants) or homozygous for the Tog5672 allele (two plants) for *RYMV3* were homogeneously infected two weeks after inoculation. The second resistance determinant of Tog5672 thus showed the same specificity of interaction as Tog5307 with BF1 and Ng119 isolates, which is in agreement with *RYMV3* being involved in Tog5672 resistance.

2.4 Discussion

Rice yellow mottle virus is one of the main causes of losses in rice crops in Africa and Madagascar, and breeding resistant cultivars is likely the best strategy to control this disease. However, understanding the specificity of interactions between RYMV isolates, resistance genes or alleles, and the genetic background is crucial to achieving durable resistance. The analysis of the Tog5307 resistance spectrum revealed contrasted patterns of interaction with RYMV isolates. With some isolates, all the plants were rapidly infected, suggesting that *RYMV3* is not efficient against these isolates. With the other isolates, infection occurred later and in a limited number of plants. These observations may reflect incomplete penetrance of resistance, i.e. the occurrence of plants with the resistant genotype that develop the disease, as previously described in other plant/virus pathosystems (Chandra-Shekara et al., 2006; Ouibrahim et al., 2014; Poque et al., 2015). However, they may also result from the emergence in some plants of a viral variant able to overcome resistance, as frequently observed in host/pathogen interactions (Lecoq

et al., 2004; Quenouille et al., 2013). Resistance breakdown is well documented in the case of the *RYMV1* gene in the rice/*RYMV* interaction (Hébrard et al., 2006). The two hypotheses may coexist. Back inoculation and a search for the emergence of mutations in the viral populations are planned to determine which hypothesis applied.

Poulicard et al. (2012) reported a selective advantage of isolates with a T at position 49 of the VPg in susceptible *O. glaberrima* accessions when T49 and E49 isolates were inoculated in competition, and interpreted it as an adaptation to *O. glaberrima* species. This T/E polymorphism has also been previously identified as a key determinant of the ability to overcome resistance of Gigante (*O. sativa* accession carrying *rymv1-2*) and Tog5681 (*O. glaberrima* accession carrying *rymv1-3*) : E49 isolates cannot overcome the resistance of Tog5681, and T49 isolates rarely overcome the resistance of Gigante (Traoré et al., 2010). Here we observed that, excluding the isolates that infected all Tog5307 plants, incidence of the disease was significantly higher for T49 isolates than for E49 isolates. Hence, T/E specificity is independent of the resistance genes. The T49 isolates, which are fitter than E49 isolates on *O. glaberrima* accessions, are probably more prone to infect resistant accessions, irrespective of whether infection is due to incomplete penetrance of the resistance or to the acquisition of resistance breaking mutations, as suggested by Poulicard et al. (2012). We are currently investigating this pattern of specificity in the Tog7291 accession (*O. glaberrima*), whose resistance involves the *RYMV2* gene, and in near isogenic lines of IR64 (*O. sativa* spp. *indica*) introgressed with different resistance genes. However, the E/T polymorphism is not the only determinant of the specificity of interaction between Tog5307 and *RYMV* isolates. In particular, this polymorphism does not explain why the T49 Cl_a isolate did not infect Tog5307 and why 100 % incidence was observed with the E49 Tz211 isolate. The very high incidence observed with Tz211, Ng119 and Nil is presumed to be due to their virulence, according to the definition of VanderPlank (1968), on Tog5307, which behaves like a susceptible accession when challenged with these isolates.

Tog5307 resistance is under the control of *RYMV3*, the third major resistance gene to *RYMV* identified in *O. glaberrima*. Our mapping, sequencing and isolate specificity results suggest it is also involved in Tog5672 resistance. We think *RYMV3* is sufficient to explain the resistance of Tog5307, but we cannot totally exclude that an interaction between *RYMV3* and a gene closely linked to the *S1* locus, on chromosome 6, is involved. In *O. sativa* x *O. glaberrima* crosses, the *S1* locus is the main interspecific sterility locus and was shown to undergo a strong segregation distortion toward the *O. glaberrima* allele in progenies (Lorieux et al., 2000). This locus explains why such distortion was observed in BC1F1 and, in the same way, BC3F2 and BC3F3 used for *RYMV3* fine mapping are expected to be homozygous for the *O. glaberrima* allele at locus *S1*. This lack of segregation would prevent the detection of a gene linked to locus *S1* that may be involved in resistance in interaction with *RYMV3*. A small fraction of plants with the *O. sativa* allele at locus *S1* would be sufficient to exclude the implication of such a gene. To date, back crosses for introgression of *RYMV3* in IR64 are as advanced as

BC6. We confirmed the association of *RYMV3* with resistance in BC6F2 but, despite screening more than one hundred BC6F2 plants, no plants with the *O. sativa S1* allele were identified (data not shown).

The segregation of resistance in the Tog5672 x Tog6206 population does not exclude the involvement of another resistance gene, which could have been masked by *S1* distortion or by the absence of polymorphism in the Tog5307 x IR64 population. Some plants belonging to the Tog5672 x Tog6206 population that were heterozygous for *RYMV3* were only partially resistant, thereby supporting the involvement of such a gene. However incomplete penetrance of *RYMV3* mediated resistance could also explain the observed segregation. Penetrance may be affected by environmental factors and by a dosage effect, explaining why some heterozygous plants were partially resistant while all the homozygous ones were highly resistant. Dosage and temperature dependent resistance has already been described for the TIR-NBS-LRR resistance gene *I* in bean (Collmer et al., 2000). The effect of temperature on NB-LRR mediated resistance is well-known (Zhu et al., 2010) and the effect of light on penetrance was reported in the resistance of *Arabidopsis* to *Turnip Crinkle Virus* (Chandra-Shekara et al., 2006) and to *Watermelon mosaic virus* (Ouibrahim et al., 2014). Analysis of progenies rather than of individual plants would have prevented the detection of this dosage dependent response in Tog5307 derived progenies used for *RYMV3* fine mapping. We checked the penetrance of the resistance in 84 BC5F3 plants from three segregating families derived from the introgression process of *RYMV3* in IR64 and no heterozygous plants developed the disease (data not shown). All these results highlight the overriding role of *RYMV3* in resistance but suggest that complex interactions between the gene, the genetic background and the features of the RYMV isolates may affect the expression of resistance.

We mapped *RYMV3* in a 15 kb-long interval on the CG14 reference genome. Two candidate genes are annotated in this interval. ORGLA11g0175700, encodes a hypothetical protein with a probable transmembrane localization but this gene has no homology outside the *Oryza* genus or with any known functional domain. However, resistance genes with no significant homology to any characterized proteins have already been described (Wang et al., 2015; Xiao et al., 2001). If ORGLA11g0175700 is involved in resistance, two non-synonymous polymorphisms between Tog5307 and CG14, one in the probable cytoplasmic domain and the other outside, are candidates to explain the gain in resistance. The other gene, ORGLA11g0175800, is a typical NB-LRR gene belonging to the CC-NBS-LRR subfamily. Genes in this class are the most common dominant resistance genes in all types of pathosystems. Resistance to plant viruses is not an exception to the rule, as 20 NB-LRR genes against plant viruses have been cloned so far (Boualem et al., 2016). NB-LRR proteins are cytoplasmic. They are composed of a central LRR (leucine rich repeat) domain involved in pathogen specificity, linked to a C-terminal NB-ARC (nucleotide-binding adaptor shared by APAF-1, certain R gene products and CED-4) domain required for signal transduction, and either a TIR (toll

and interleukin-1 receptor) or a CC (coiled coil) domain in the N terminal, which play a role in the auto-inhibition of the protein in and protein-protein interactions (Takken and Govere, 2012). NB-LRR genes directly or indirectly specifically recognize a strain or isolate specific avirulence factor and trigger the resistance signaling cascade, which in the majority of cases leads to local cell death, named hypersensitive response. In the plant/virus pathosystem, an extreme resistance, characterized by no detectable viral accumulation and no visible necrosis, can also be observed (Bendahmane et al., 1999). In the ORGLA11g0175800 gene, we detected a single non-synonymous SNP between Tog5307 from CG14, leading to a conversion from a lysine to an arginine in the LRR domain. The LRR domain of NB-LRR proteins has been shown to play a central role in virus recognition (Farnham and Baulcombe, 2006; Sekine et al., 2012; Tomita et al., 2011) and such polymorphism can explain the difference between a resistant and a susceptible plant, as previously demonstrated in rice/blast interactions (Bryan et al., 2000; Liu et al., 2007). The involvement of this gene would be in agreement with the observed isolate-specific resistance response. Therefore, we believe ORGLA11g0175800 is the best candidate for Tog5307 resistance. The functional complementation of a susceptible *O. sativa* variety with the Tog5307 allele of the candidate genes will confirm the role of the candidates, and exclude the involvement of additional genes in Tog5307 resistance.

In conclusion, we mapped *RYMV3*, a new resistance gene to RYMV, and identified a NB-LRR as a very convincing candidate gene. We showed that *RYMV3*-associated resistance can be codominant or incompletely penetrant, depending on the genetic background or the environmental conditions. The three known resistance genes to RYMV are presumed to be involved in different mechanisms of interaction with the virus. Their pyramiding will probably render the emergence of resistance-breaking variants difficult. This is thus a very promising way to achieve durable resistance, in particular in the regions of Africa where specific interactions between the local isolates, the resistance genes and the genetic background may rapidly impair the efficiency of varieties with a single resistance gene.

Acknowledgements This work was financially supported by the Agropolis Foundation (project N° 1403-066, SURVEY) and the Global Rice Science Partnership (GriSP). The French Ministère de l'Enseignement Supérieur et de la Recherche provided a PhD grant to H. Pidon. We thank Christine Tranchant-Dubreuil, Agnès Pinel-Galzi and Harold Chrestin for technical help and Yacouba Séré for giving access to some RYMV isolates. We acknowledge François Anthony and Mathias Lorieux for helpful discussions.

3 Conclusion générale et perspectives

3.1 *RYMV3* confère une résistance monogénique dominante

La cartographie génétique de la résistance de Tog5307 dans un croisement Tog5307 x IR64 a révélé la présence d'un gène de résistance élevée, dominant, nommé *RYMV3*. Il s'agit du premier gène de résistance dominant identifié dans le pathosystème riz/*RYMV*. Nos résultats suggèrent également que ce gène contrôle la résistance au *RYMV* dans l'accession Tog5672, précédemment décrite comme possédant deux facteurs de résistance dont *rymv1-4* (Thiémélé et al., 2010).

Cependant, comme discuté dans l'article, la sensibilité de certains hétérozygotes de la population Tog6206 x Tog5672, pourrait suggérer l'implication d'un second gène dans cette résistance ou une pénétrance incomplète de la résistance conférée par *RYMV3*. En outre, la forte distorsion de ségrégation sur le locus *S1* et l'analyse du phénotype sur descendances et non sur plantes individuelles auraient pu masquer l'un ou l'autre de ces éléments au cours de l'analyse de la résistance de Tog5307.

Deux expérimentations complémentaires ont été réalisées afin d'évaluer l'éventualité d'un autre gène ou d'une pénétrance incomplète de la résistance. D'une part, comme mentionné dans la discussion de l'article, 84 plantes BC5F3 du croisement Tog5307 x IR64, issues de trois BC5F1 indépendantes et incluant 42 plantes hétérozygotes à *RYMV3*, ont été évaluées pour leur résistance à l'isolat BF1, sur la base de tests ELISA réalisés dix jours après inoculation. À l'exception d'une plante montrant une réponse ambiguë, tous les hétérozygotes se sont avérés résistants, confirmant ainsi la dominance de la résistance provenant de Tog5307. La résistance partielle observée dans le croisement Tog5672 x Tog6206 est donc spécifique à ce croisement. D'autre part, la résistance de 25 plantes hétérozygotes issues du croisement Tog5672 x Tog6206 a été évaluée avec l'isolat CIa, incapable d'infecter Tog5307 lors des expériences de spécificité de la résistance. Toutes les plantes présentaient un niveau de résistance élevé en ELISA, rendant très peu probable l'hypothèse d'un troisième gène de résistance dans cette population. La présence de résistants partiels à BF1 dans les hétérozygotes inoculés avec BF1 pourrait s'expliquer par des mécanismes de pénétrance incomplète ou de contournement, isolat spécifique et similaires à ceux décrits sur Tog5307 dans la première partie de l'article.

Le dosage moindre de la protéine de résistance dans les hétérozygotes, combiné à une interaction plus efficace entre un isolat T49 et une plante de fond génétique *O. glaberrima* (Poulicard et al., 2012), pourrait favoriser l'infection partielle de ces plantes dans un intervalle de temps relativement court. En revanche, les plantes du croisement de cartographie sont des back-cross avancés sur IR64. Leur fond génétique est donc majoritairement *O. sativa*, dans lequel l'interaction avec un isolat T49 est moins favorable au virus, pouvant expliquer la différence observée entre les deux populations.

Cette différence d'infection des hétérozygotes pourrait également être provoquée par un effet environnemental, comme la température ou la lumière, pouvant avoir varié

entre les différentes expériences, comme discuté dans l'article.

3.2 *RYMV3* est probablement un gène de type CC-NBS-LRR

Le gène *RYMV3* a été cartographié dans une région d'environ 15 kb sur le chromosome 11. Deux gènes candidats y ont été identifiés. Le premier gène n'a aucun homologue en dehors du genre *Oryza*. Des analyses de prédictions de domaines protéiques suggèrent qu'il s'agirait d'un gène codant pour une protéine transmembranaire. Le second gène est uniquement partiellement compris dans l'intervalle de cartographie. Il s'agit d'un gène codant pour une protéine de type CC-NBS-LRR, typique des gènes de résistance aux agents pathogènes.

Trois SNP codants séparent Tog5307 de l'accession *O. glaberrima* sensible CG14. Un seul d'entre eux est également polymorphe entre Tog5672 et CG14. Ce SNP est localisé dans le gène NB-LRR. Il induit le changement d'une lysine en arginine dans le domaine LRR de la protéine, connu pour jouer un rôle central dans la reconnaissance de l'agent pathogène (Farnham and Baulcombe, 2006; Sekine et al., 2012; Tomita et al., 2011). Un polymorphisme d'un seul acide-aminé dans ce domaine peut expliquer la différence entre résistance et sensibilité dans le cas de l'interaction riz/pyriculariose (Bryan et al., 2000; Liu and Chen, 2007). Comme discuté dans l'article, et bien qu'on ne puisse pas totalement exclure le rôle de SNP dans des régions non codantes, ces éléments nous ont amené à considérer le gène NB-LRR comme le meilleur candidat pour *RYMV3*.

Afin de le valider définitivement, une transformation stable par *Agrobacterium tumefaciens* avec l'allèle de Tog5307 du gène NB-LRR ORGLA11g0175800 est en cours. Les transformations de l'espèce *O. glaberrima* et de la sous-espèce *O. sativa* spp. *indica* étant encore mal maîtrisées, la transformation est faite dans la variété *O. sativa* spp. *japonica* Kitake, facile à transformer et sensible au RYMV. La synthèse de l'ADN complémentaire et son introduction dans le vecteur PCambia5300UbiNos, sous promoteur constitutif ubi, a été confiée à GeneCust (Dudelage, Luxembourg). Des essais préliminaires de phénotypage pourront éventuellement être faits sur des boutures de plantes T0, avant l'analyse des T1.

Par ailleurs, une analyse de l'association des polymorphismes de Tog5307 dans la région de cartographie avec la résistance au RYMV dans la collection de riz africains de l'IRD a été réalisée et est présentée au chapitre 4.

3.3 Contournement ou pénétrance incomplète de la résistance ?

Dans la première partie de l'article, la capacité d'une collection d'isolats de RYMV à infecter Tog5307 a été analysée. Il a été mis en évidence que certains isolats avaient une forte capacité à infecter Tog5307, et que cette capacité était associée au polymorphisme E/T en position 49 de la VPg, comme c'est également le cas dans le contournement des gènes *RYMV1* (Traoré et al., 2010) et *RYMV2* (Pinel-Galzi et al., sub). En revanche, les

résultats présentés ne permettaient pas de savoir si les cas d'infection relevaient d'une pénétrance incomplète de la résistance face à certains isolats ou d'un contournement lié à l'émergence de variants viraux.

Des expérimentations complémentaires réalisées au laboratoire suggèrent que l'émergence de variants contournant la résistance est l'hypothèse la plus probable.

D'une part, la rétro-inoculation de Tog5307 avec un inoculum issu de plantes Tog5307 précédemment infectées par le virus a confirmé qu'au moins une partie des inoculums était capable d'infecter Tog5307 plus rapidement et plus fréquemment que l'isolat initial, suggérant une forte évolution de la population virale.

D'autre part, le séquençage d'une partie des isolats ayant réussi à infecter Tog5307 a permis d'identifier une ou plusieurs mutations dans la protéine de capsid de chacun d'entre eux (Albar, communication personnelle). Le rôle de ces mutations doit encore être validé mais leur détection suggère l'émergence de variants virulents. La protéine de capsid (CP) du RYMV serait donc la protéine d'avirulence correspondant à *RYMV3*. Ce type de protéine d'avirulence est commun dans le cas des résistances aux virus conférées par des gènes de types NB-LRR (Boualem et al., 2016) et serait donc cohérent avec le gène candidat identifié.

Afin de valider le couple gène de résistance/facteur d'avirulence, leur interaction pourrait être testée en système hétérologue, comme le double hybride chez la levure ou par agro-infiltration dans *Nicotiana benthamiana*. Toutefois, l'interaction entre un NB-LRR et sa cible est souvent indirecte, comme cela a notamment été montré dans le cas de l'interaction entre le gène de résistance *Rx* et la protéine de capsid de *Potato virus X* (PVX) (Sacco et al., 2007). De plus, le facteur d'avirulence pourrait être la capsid assemblée, et non la CP, et son assemblage en système hétérologue n'est pas garanti. Une absence d'interaction en système hétérologue ne remettrait donc pas forcément en question le couple.

Cependant, certains inoculums issus de plantes Tog5307 infectées, ainsi que des variants issus de mutagenèse dirigée pour valider le rôle des mutations observées dans la CP, ne sont pas avérés infectieux sur Tog5307, voire sur le témoin sensible IR64 (Hébrard et Albar, communication personnelle). Ces résultats pourraient s'expliquer par une perte de fitness de ces variants. En effet, la protéine de capsid est primordiale à l'intégrité du virus. Une mutation dans cette protéine peut donc déstabiliser le virus et limiter son infectiosité, comme cela a été suggéré par plusieurs études sur la résistance du piment au *Pepper mild mottle virus* (Hamada et al., 2007; Tsuda et al., 1998). Cette hypothèse est encourageante pour la durabilité de la résistance conférée par *RYMV3* et devra être confirmée par des résultats complémentaires.

CHAPITRE

4

Analyse de la résistance au RYMV dans
l'espèce *O. glaberrima* et son apparenté
sauvage *O. barthii*

Chapitre 4 : Analyse de la résistance au RYMV dans l'espèce *O. glaberrima* et son apparenté sauvage *O. barthii*

1 Introduction

À ce jour, trois gènes de résistance majeurs contre le RYMV ont été identifiés, dont deux uniquement dans l'espèce *O. glaberrima*. Les travaux de Thiémélé et al. (2010) et Orjuela et al. (2013) ont permis d'identifier un total de 31 accessions *O. glaberrima* résistantes au RYMV, soit environ 8,5 % des accessions étudiées, ainsi que deux gènes de résistance. Le premier, *RYMV1* code pour eIF(iso)4G1, un facteur d'initiation de la traduction (Albar et al., 2006). Trois allèles de résistance ont été identifiés dans l'espèce *O. glaberrima*. Ils induisent tous des substitutions ou des délétions dans le domaine MIF4G de la protéine. Le second gène code pour un allèle nul d'une protéine du pore nucléaire impliquée dans la cascade de signalisation de la résistance, CPR5-1 (Gu et al., 2016; Orjuela et al., 2013). L'allèle de résistance a été identifié dans sept accessions *O. glaberrima*. Au chapitre précédent, nous avons relaté l'identification d'un troisième gène de résistance dans l'accession Tog5307, codant probablement pour une protéine de type CC-NBS-LRR.

L'IRD dispose d'une collection des 278 accessions *O. glaberrima*, dont font partie les 31 accessions résistantes identifiées par Thiémélé et al. (2010), et de 103 accessions de son apparenté sauvage *O. barthii*. Les 165 accessions *O. glaberrima* et 84 accessions *O. barthii* de la collection IRIGIN sont incluses dans cette collection IRD. Avant cette étude, le phénotypage de cette dernière était incomplet : 161 accessions *O. glaberrima* n'avaient jamais été phénotypées et la résistance de l'espèce *O. barthii* n'avait pas été étudiée.

L'objectif de cette étude était donc d'obtenir un panorama de la résistance élevée au RYMV dans la diversité de ces espèces en recherchant d'autres accessions résistantes dans la collection IRD et en étudiant leur génotype aux gènes de résistance connus.

Pour cela, nous avons complété le phénotypage des accessions *O. glaberrima* de la collection IRD et phénotypé une partie des accessions *O. barthii*. Nous avons ensuite recherché la présence d'allèles de résistance déjà identifiés ou candidats aux gènes de résistance connus dans les accessions résistantes. Ces accessions et leur déterminisme de résistance ont été positionnées dans la diversité phylogénétique et géographique de l'espèce afin de tenter de dégager des hypothèses quant à la dynamique évolutive de la résistance au RYMV.

2 Résultats

2.1 Identification des individus résistants

Nous avons complété le phénotypage de la collection IRD d'*O. glaberrima*. Pour cela nous avons inoculé environ quatre plantes par accession avec l'isolat BF1 du RYMV, connu pour être très agressif (Pinel et al., 2000). Le phénotype des accessions déjà identifiées comme résistantes a également été vérifié. Cela nous a permis d'identifier 13 nouvelles accessions résistantes. En revanche, trois accessions précédemment identifiées comme résistantes à l'isolat BF1 par Thiémélé et al. (2010) sont apparues sensibles dans nos tests, sûrement à cause d'un problème de graines. Le nombre total d'accessions résistantes est donc porté à 41 soit près de 15 % de la collection IRD.

La collection IRD contient également 103 accessions de l'espèce sauvage apparentée à *O. glaberrima* : *O. barthii*. La diversité génétique étant supérieure dans le compartiment sauvage par rapport au compartiment cultivé, cette espèce peut contenir des résistances intéressantes au RYMV. Ces accessions sont difficiles à phénotyper car leur germination est incertaine, souvent faible, et peu homogène. Pour la plupart d'entre elles, le nombre de graines disponibles est faible à cause d'un égrainage important lors de la multiplication. Lorsque le nombre de graines et leur germination nous ont permis de réaliser le test, nous avons également pu observer des symptômes hétérogènes entre les quatre plantes de certaines accessions. Cette hétérogénéité peut provenir d'une hétérogénéité des lots de graines ou bien d'une caractéristique intrinsèque de pénétrance incomplète de la résistance de certaines accessions *O. barthii*. Nous n'avons pu acquérir de manière à peu près solide que les phénotypes de 50 accessions. Nous y avons identifié quatre accessions hautement résistantes, mais le reste de la collection contient probablement d'autres sources de résistances.

2.2 Recherche d'allèles aux gènes de résistance connus

Nous avons recherché le génotype des accessions de la collection aux gènes de résistance connus. Dans le cas des accessions séquencées, nous avons recherché les polymorphismes dans les fichiers de polymorphismes IRIGIN. Dans les autres cas, seules les accessions résistantes ont été étudiées. Nous y avons recherché des allèles soit par séquençage Sanger de régions contenant une ou plusieurs mutations précédemment identifiées (dans le cas de *RYMV1* et *RYMV2*) soit en utilisant un marqueur spécifique du polymorphisme causal de la résistance (*RYMV3*). Nous avons ainsi pu mettre en évidence des accessions présentant des allèles de résistance précédemment identifiés mais également identifier de nouveaux allèles rares, candidats pour la résistance.

2.2.1 Allèles à *RYMV1*

Trois allèles de résistance à *RYMV1* avaient déjà été identifiés dans l'espèce *O. glaberrima* : *rymv1-3*, *rymv1-4* et *rymv1-5* (Albar et al., 2006; Thiémélé et al., 2010). Ces allèles correspondent à une délétion et/ou une substitution d'acide aminé dans le domaine MIF4G de la protéine. Il a été montré que la résistance conférée par ces allèles de résistance résultait d'une rupture d'interaction entre le domaine MIF4G de la protéine et la VPg du virus (Hébrard et al., 2010). Il est donc probable que l'ensemble des polymorphismes de résistance au RYMV à *RYMV1* soient situés dans ce domaine.

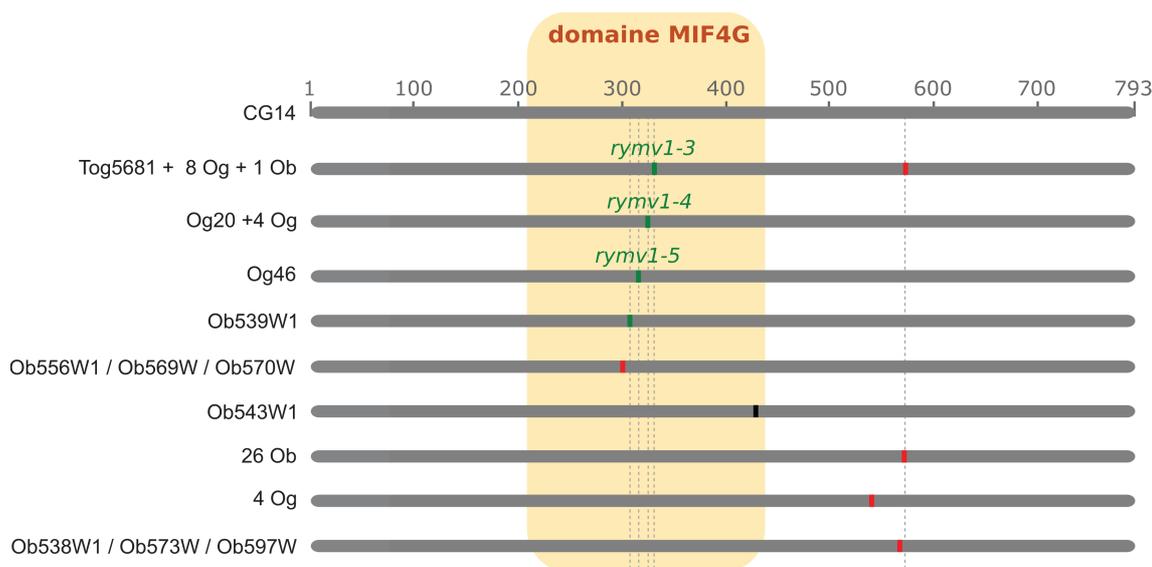


FIGURE 4.1 – Positions des polymorphismes non-synonymes dans la protéine RYMV1 présents dans la collection IRIGIN. Les positions correspondent à la séquence de référence de la variété *O. sativa* Nipponbare et les polymorphismes sont indiqués par rapport à ceux présents dans la séquence de CG14. Les polymorphismes spécifiques des variétés résistantes sont indiqués en vert, ceux qui ne le sont pas en rouge, et ceux dont l'association avec la résistance est inconnue en noir. Le domaine MIF4G de la protéine est indiqué en couleur.

Dans le domaine MIF4G, les séquences obtenues par le projet IRIGIN nous ont permis d'identifier une nouvelle accession *O. glaberrima* ayant l'allèle *rymv1-4* (Og423) et une accession *O. barthii* résistante porteuse de l'allèle *rymv1-3* (Ob589W) (Figure 4.1). Une autre accession *O. barthii* résistante (Ob539W1) possède une mutation rare dans ce domaine, pouvant être un cinquième allèle de résistance et induisant la mutation d'une leucine en arginine à l'acide-aminé 306 de la protéine. L'accession *O. barthii* Ob543W1 a une mutation non synonyme en limite du domaine MIF4G. Au niveau protéique, elle se traduit par la substitution d'une valine en isoleucine à l'acide aminé 429. Toutefois, le phénotype de cette accession n'a pu être évalué à ce jour et la mutation est relativement éloignée des polymorphismes de résistance validés.

Par ailleurs, nous avons identifié trois accessions *O. barthii* sensibles (Ob556W1, Ob569W et Ob570W) possédant une mutation dans le domaine MIF4G. Celle-ci induit la mutation d'un acide aspartique (D) en alanine (A) en position 303 de la protéine.

Cette mutation n'est pas associée à la résistance au RYMV. Toutefois, si l'ensemble des accessions *O. glaberrima* et les autres accessions *O. barthii* étudiées portent un D à cette position, au contraire, tous les *O. sativa* génotypés à ce jour, y compris au sein du projet 3000 génomes (Alexandrov et al., 2015), ont un A. Cela fait donc de ce polymorphisme une signature des riz asiatiques.

Le génotype des accessions résistantes ne faisant pas partie de la collection IRIGIN a été acquis en séquençant en Sanger la portion du gène correspondant au domaine MIF4G de la protéine. Aucun allèle de résistance connu ou candidat n'y a été identifié.

Dans la collection IRD, nous avons donc pu identifier au total :

- neuf accessions *O. glaberrima* et une accession *O. barthii* avec l'allèle *rymv1-3*,
- cinq accessions *O. glaberrima* avec l'allèle *rymv1-4*,
- une accession *O. glaberrima* avec l'allèle *rymv1-5*,
- et deux accessions *O. barthii* porteuses de nouveaux polymorphismes candidats pour la résistance.

Le séquençage de la collection IRIGIN a également permis d'étudier les polymorphismes sur le reste du gène.

Dans les accessions *O. glaberrima*, un SNP non-synonyme au codon 576 est présent dans toutes les accessions porteuses de *rymv1-3* et dans aucune autre. Il est également présent dans l'accession *O. barthii* porteuse de *rymv1-3* mais aussi dans 26 autres accessions *O. barthii* n'ayant aucun allèle de résistance à *RYMV1* et parmi lesquelles 14 au moins sont sensibles. Il se pourrait donc que l'allèle *rymv1-3* ait évolué à partir d'une plante sensible possédant ce polymorphisme.

Deux autres mutations non synonymes non associées à la résistance au RYMV ont également été identifiées aux codons 541 et 575, respectivement chez quatre accessions *O. glaberrima* et trois accessions *O. barthii*.

Les polymorphismes synonymes sont également très rares dans les exons du gène. Seuls trois sont présents dans la collection IRIGIN.

2.2.2 Allèles à *RYMV2*

Lors de l'identification de *RYMV2* chez Tog7291 par Orjuela et al. (2013), un marqueur spécifique de l'allèle de résistance identifié a été utilisé pour génotyper l'ensemble de la collection IRD. Au total sept accessions *O. glaberrima* portant l'allèle de Tog7291 ont été identifiées par ce criblage. L'utilisation d'un marqueur plutôt que d'un séquençage n'avait alors pas permis d'identifier d'autres allèles de résistance candidats à ce gène. Cependant, la résistance étant conférée par un allèle nul du gène et contrairement au cas de *RYMV1*, aucune région n'était particulièrement susceptible de contenir un autre allèle de résistance. Pour identifier d'autres allèles il aurait alors fallu séquencer le gène complet sur toutes les accessions, ce qui aurait été trop coûteux.

Disposant maintenant des séquences de la collection IRIGIN, nous avons pu faire cette recherche de polymorphismes (Figure 4.2). Nous avons identifiés cinq polymorphismes

TABLE 4.1 – Mutations non-synonymes candidates pour la résistance dans le gène *RYMV2*. Les accessions dont le nom est suivi d'une étoile sont celles ne faisant pas partie du projet de séquençage IRIGIN.
déc = décalage du cadre de lecture.

Nom IRD	Phénotype	Mutation	Codon	Effet
Og151	} résistant	AT/A	224	déc.
Og152*				
Og153*				
Og213				
Og406*				
Og498				
Og133	} résistant	CG/C	69	déc.
Og447*				
Og452*				
Og183	résistant	TGC/AGA	216	S216I et codon STOP
Og256	résistant	A/T	472	codon STOP
Og491	résistant	GC/G	378	déc.
Ob525W1	inconnu	G/GT	327	déc.
Ob608W	sensible	T/A	305	S305C
Ob608W	sensible	T/C	421	A421G
Ob557W	sensible	G/A	303	T303I
Ob587W	sensible	G/A	62	E62G

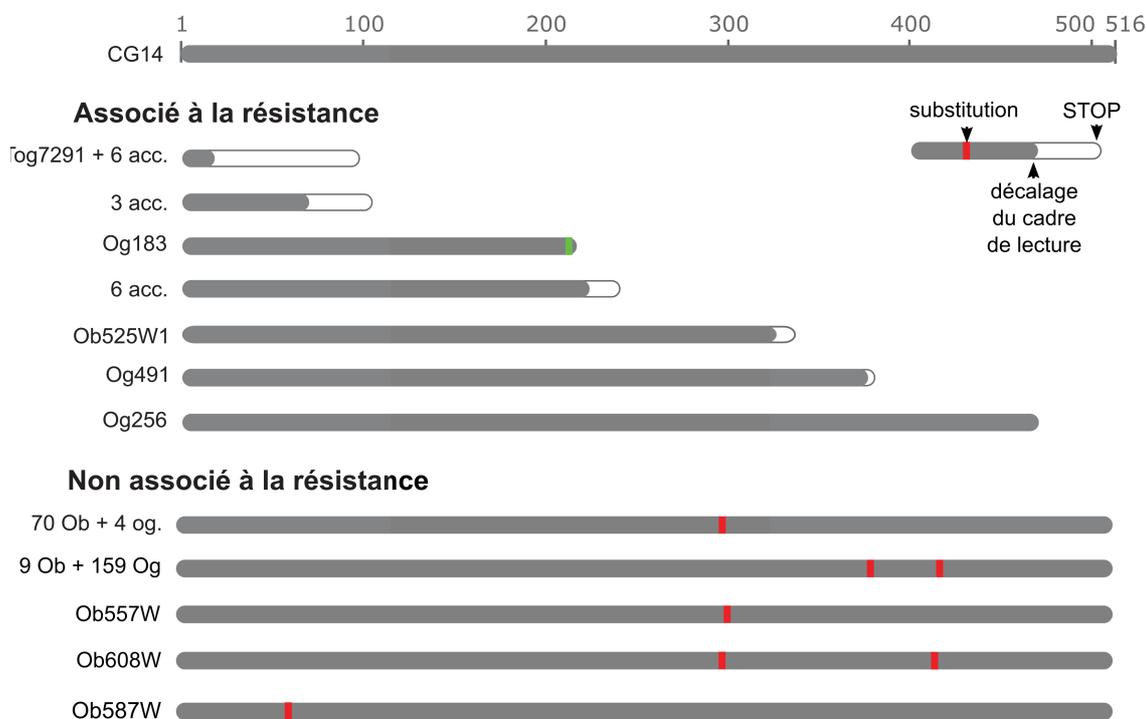


FIGURE 4.2 – Position des mutations non-synonymes sur le modèle protéique de RYMV2. La séquence de référence utilisée est celle de la variété *O. sativa* de référence Nipponbare. Les polymorphismes sont représentés par rapport à ceux de CG14. Les polymorphismes en rouge ne sont pas associés à la résistance.

rare distincts induisant un décalage du cadre de lecture ou un codon STOP dans le gène *RYMV2* (Table 4.1). Ces polymorphismes résultent en une protéine tronquée, probablement non fonctionnelle. Les sept accessions *O. glaberrima* présentant un de ces polymorphismes sont toutes résistantes au RYMV et n'ont aucun allèle de résistance connu par ailleurs. Ces cinq polymorphismes sont donc susceptibles d'induire leur résistance au RYMV.

Nous avons recherché ces allèles de résistance putatifs dans les accessions résistantes ne présentant aucun allèle de résistance connu et ne faisant pas partie du projet de séquençage IRIGIN. Pour cela nous avons fait séquencer en Sanger les régions spécifiques de ces allèles. Nous avons identifié cinq accessions supplémentaires possédant un des allèles induisant un décalage du cadre de lecture dans *RYMV2*.

Par ailleurs, plusieurs accessions *O. barthii* présentent également des polymorphismes rares non-synonymes. Dans le cas de l'accèsion Ob525W1 il s'agit d'une insertion d'une paire de bases induisant un décalage du cadre de lecture et pouvant conférer la résistance. Toutefois, faute de graines, nous n'avons pu établir son phénotype. Trois autres accessions, toutes sensibles, portent des SNP non-synonymes induisant des changements d'acides aminés. Les mutations qu'elles portent ne seraient donc pas suffisantes pour conférer la résistance.

Trois autres polymorphismes, induisant des substitutions d'acides aminés, semblent distinguer les espèces *O. glaberrima* et *O. barthii*. Les substitutions F384L et D421N sont présentes dans 159 accessions *O. glaberrima* et seulement neuf accessions *O. barthii*, alors que la substitution T298K est présente dans 70 accessions *O. barthii* et seulement quatre accessions *O. glaberrima*. Aucune accession ne présente à la fois la mutation T298K et le couple F384L-D421N.

Comme dans le cas de *RYMV1*, le faible nombre de mutations non-synonymes est associé à un faible nombre de mutations synonymes dans les régions codantes : seulement deux sont présentes dans l'ensemble de la collection IRIGIN.

Afin de distinguer clairement les allèles les plus fréquents à *RYMV2*, nous avons choisi de les nommer de manière similaire à celle utilisée pour les allèles à *RYMV1*. L'allèle *rymv2-1* est donc l'allèle de sensibilité à *RYMV2*. L'allèle de résistance d'Orjuela et al. (2013), porté par sept accessions dont Tog7291, est *rymv2-2*. L'allèle d'Og151, porté par six accessions au total, est *rymv2-3*, et l'allèle d'Og133, porté par trois accessions, est *rymv2-4*. Enfin, les autres allèles identifiés ne l'ont été que dans une accession chacun et nous avons choisi de ne pas les nommer pour l'instant.

2.2.3 Polymorphismes à *RYMV3*

Au chapitre précédant, nous avons mis en évidence, dans l'accèsion *O. glaberrima* Tog5307, une région d'environ 15 kb contenant le gène dominant de résistance *RYMV3*. Nos résultats suggèrent également que ce gène est le second déterminant de la résistance de l'accèsion Tog5672, qui possède également l'allèle de résistance *rymv1-4*. L'intervalle

de cartographie comprend deux gènes : un gène putatif contenant un domaine transmembranaire et un gène de type CC-NBS-LRR. Par rapport à l'accèsion *O. glaberrima* de référence CG14, Tog5307 présente deux SNP codants et un SNP intronique dans le gène transmembranaire, ainsi qu'un SNP codant dans le domaine LRR et une délétion intronique de 33 pb dans le gène NB-LRR. Parmi ces mutations, Tog5672 ne possède que le SNP codant du gène NB-LRR.

Nous avons étudié l'association de ces polymorphismes avec le phénotype de résistance au RYMV dans la collection IRIGIN.

Comme présenté dans la figure 3.4, la divergence entre la séquence de référence *O. sativa* et la référence *O. glaberrima* est très grande dans la région de cartographie du gène *RYMV3*. Le gène transmembranaire est notamment absent du génome *O. sativa*. Afin d'obtenir le génotype à ce gène, nous avons donc utilisé les données de séquençage IRIGIN *mappées* sur le génome de CG14 (Wang et al., 2014c). Toutefois ces données ne comprennent que 120 accèsions *O. glaberrima*. Le NB-LRR étant présent dans le génome de Nipponbare, les séquences IRIGIN de l'ensemble des 163 accèsions *O. glaberrima* et 84 *O. barthii* *mappées* sur le génome de Nipponbare ont pu être étudiées.

L'haplotype de Tog5307 au gène transmembranaire est présent dans huit autres accèsions sur les 120 analysées. Parmi elles, seule Og126 est résistante au RYMV. Ces polymorphismes ne sont donc pas associés à la résistance et le gène transmembranaire n'est probablement pas le gène de résistance.

Dans le NB-LRR, la délétion de 33 paires de bases dans l'intron est présente dans 20 accèsions *O. glaberrima* et 44 accèsions *O. barthii*. Parmi elles, une seule accèsion *O. glaberrima* et trois accèsions *O. barthii* sont résistantes. Le SNP du domaine LRR est porté par six accèsions *O. glaberrima* et deux accèsions *O. barthii*. Ces huit accèsions sont résistantes au RYMV. Ce SNP est donc le seul des cinq polymorphismes de Tog5307 associé au phénotype de résistance.

À l'aide d'un marqueur dCAPS spécifique, ce SNP a été recherché parmi les accèsions *O. glaberrima* résistantes de la collection IRD non séquencées dans le projet IRIGIN. Deux accèsions, Og26 et Og51, portent ce SNP.

Cette association unique du SNP du gène NB-LRR avec la résistance est un élément supplémentaire pour affirmer que le gène NB-LRR est le gène de résistance et que ce polymorphisme est responsable de la résistance.

À l'image des analyses réalisées pour les deux autres gènes de résistance, nous avons recherché l'ensemble des polymorphismes codants du gène NB-LRR dans les séquences IRIGIN. Au total, 39 mutations non synonymes et 18 mutations synonymes avec l'accèsion de référence CG14 ont été identifiées. Neuf sont présentes dans la majorité des riz africains. Les quelques individus n'ayant pas ces polymorphismes ont donc le génotype de Nipponbare à ces positions. Parmi ces 39 polymorphismes, quatre sont localisés dans le domaine CC, 13 dans le domaine NB-ARC, et 14 dans le domaine LRR (Figure 4.3).

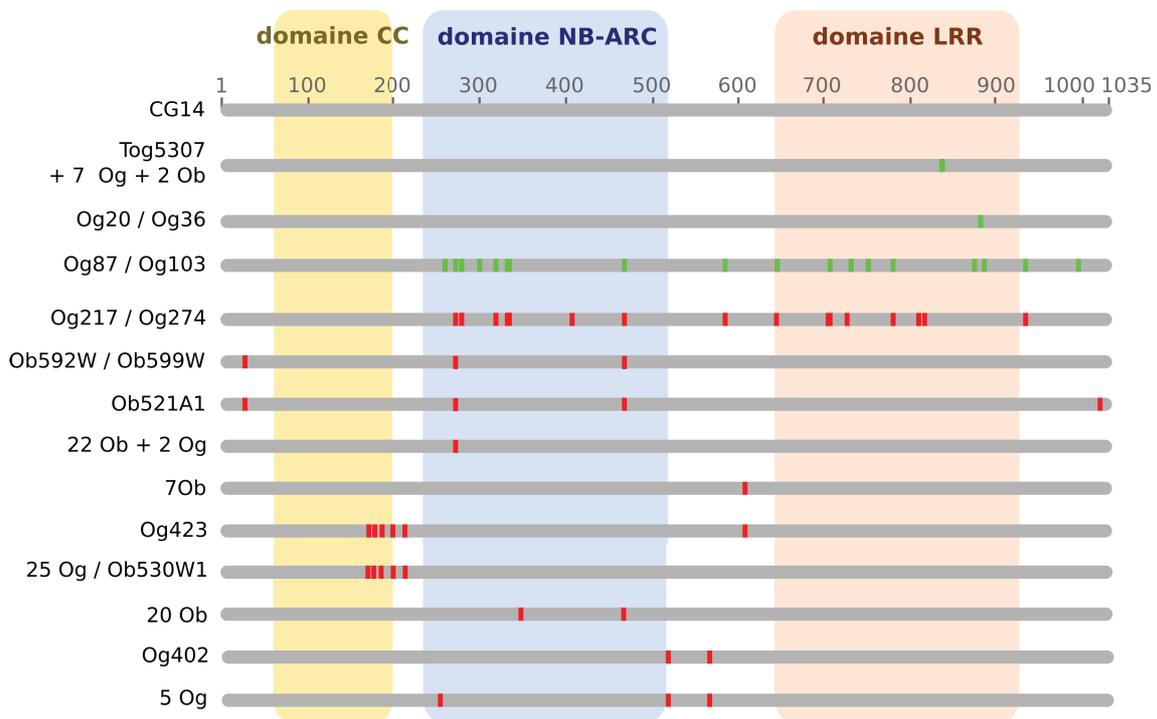


FIGURE 4.3 – Position des mutations codantes dans le NB-LRR sur le modèle protéique de RYMV3. La séquence de référence utilisée est celle de la variété *O. sativa* de référence Nipponbare et les polymorphismes sont indiqués par rapport à la séquence de CG14. Les différents domaines de la protéine, identifiés par NCBI Conserved Domains, sont indiqués en couleurs. Les polymorphismes en vert sont associés à la résistance, ceux en rouge ne le sont pas.

En plus de l'haplotype de résistance de Tog5307, on peut distinguer huit haplotypes différents. Parmi eux, deux sont associés à la résistance. Les accessions résistantes Og87 et Og103, porteuses de l'allèle de résistance *rymv2-2*, présentent ainsi 18 polymorphismes non-synonymes par rapport à CG14. Le second haplotype associé à la résistance est celui des accessions Og20 et Og36. Respectivement porteuses des allèles de résistance *rymv1-4* et *rymv1-3*, elles présentent un SNP codant dans le domaine LRR du gène.

2.2.4 Conclusion sur les allèles de résistance identifiés

À l'issue de cette étude, nous avons donc identifié cinq allèles potentiels de résistance au gène *RYMV2*, différents de celui décrit par Orjuela et al. (2013) mais correspondant également à des protéines tronquées. Nous avons confirmé que le polymorphisme du gène NB-LRR candidat pour *RYMV3* était le seul polymorphisme non-synonyme distinguant Tog5307 de CG14 dans la région candidate associée à la résistance. Nous avons également identifié un polymorphisme responsable, ou potentiellement responsable, de la résistance pour les 41 accessions *O. glaberrima* de la collection IRD qui sont hautement résistantes au RYMV. Ainsi, 15 accessions portent un allèle de résistance à *RYMV1*, 19 ont un allèle conduisant à une protéine tronquée à *RYMV2* et huit possèdent le polymorphisme identifié dans le domaine LRR du gène candidat pour *RYMV3*. Par ailleurs, quatre accessions résistantes porteuses d'allèles de résistance à *RYMV1* ou *RYMV2* ont également un génotype particulier au NB-LRR candidat pour *RYMV3*.

Dans l'espèce *O. barthii*, une accession a l'allèle *rymv1-3* et deux ont l'allèle de résistance à *RYMV3*. Nous avons également identifié au moins un allèle de résistance putatif original à *RYMV1* et un à *RYMV2*.

2.3 Répartition des résistances dans la diversité des riz africains

2.3.1 Diversité génétique

La diversité des déterminismes de résistance au RYMV identifiés dans le riz africain est importante. Nous nous sommes demandé comment cette diversité se structurerait phylogénétiquement. Pour cela, nous avons utilisé deux représentations de la diversité de l'espèce *O. glaberrima* et une de l'espèce *O. barthii*. Les premières sont des arbres phylogénétiques réalisés par Orjuela et al. (2014) sur la collection IRD selon la méthode Neighbor-joining basée sur 235 SNP polymorphes dans le riz africain. L'un représente la diversité des accessions *O. glaberrima*, l'autre celle des *O. barthii*. Le deuxième type de représentation est une analyse en composantes principales (ACP) réalisée sur les polymorphismes SNP des accessions *O. glaberrima* issus du séquençage IRIGIN. Cent mille SNP ont été échantillonnés au hasard puis l'ACP a été réalisée dans le logiciel R avec la fonction "pca" du package "LEA" (Frichot and François, 2015).

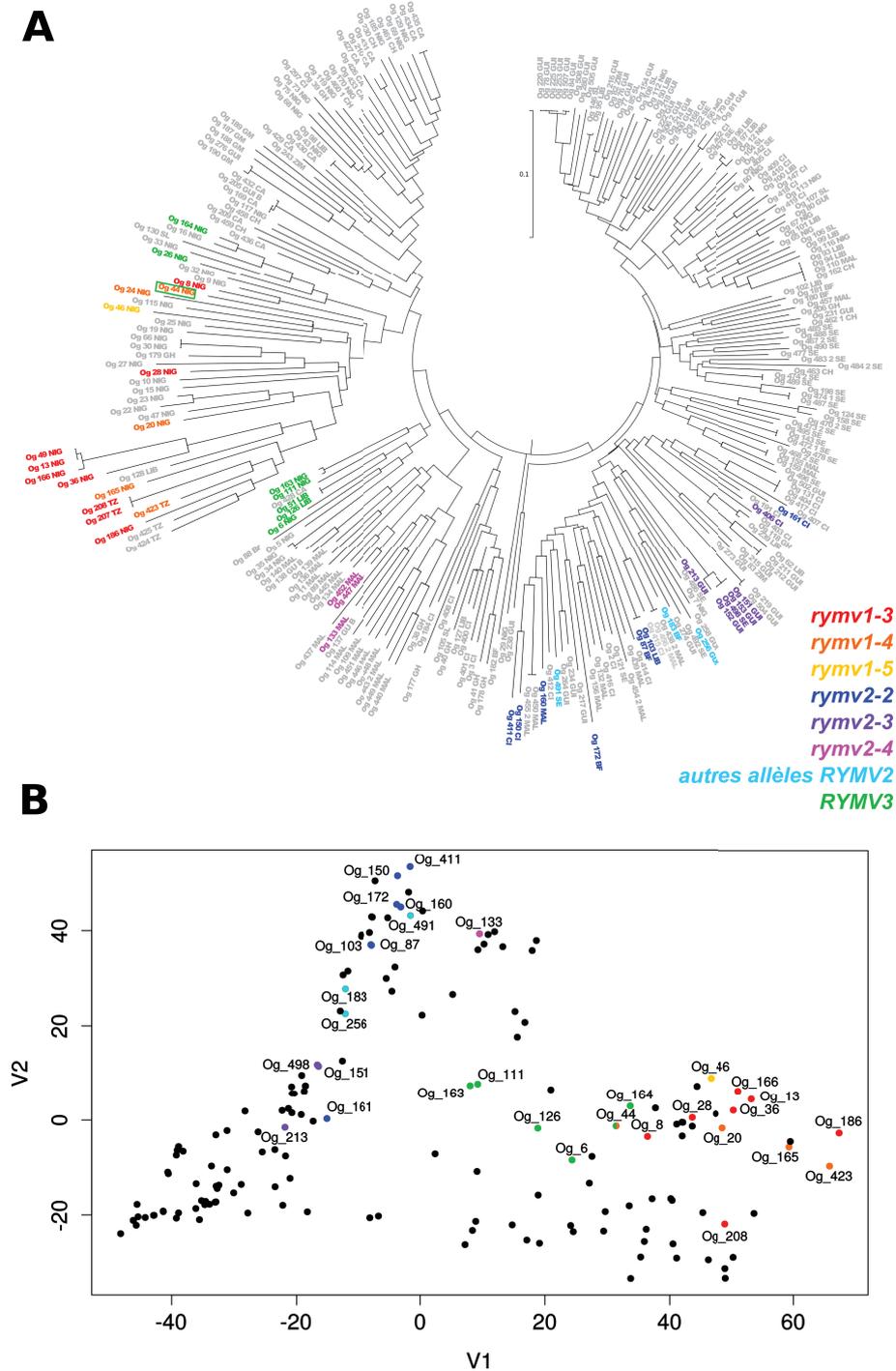


FIGURE 4.4 – Accessions *O. glaberrima* résistantes dans la diversité de l'espèce. **A** : Arbre phylogénétique d'*O. glaberrima* publié dans Orjuela et al. (2013). **B** : Graphique des dimensions 1 et 2 de l'analyse en composante principale de la diversité d'*O. glaberrima*, basée sur le séquençage du projet IRIGIN. Les accessions résistantes sont indiquées en couleur. Og44 (=Tog5672) a à la fois l'allèle *rymv1-4* et le polymorphisme candidat à *RYMV3*.

Les accessions *O. glaberrima* résistantes ont été placées sur ces représentations de la diversité (Figure 4.4). Comme attendu, les accessions ayant un même allèle de résistance sont phylogénétiquement proches dans l'arbre et dans l'ACP. Il est toutefois plus surprenant de voir que les accessions portant des allèles de résistance indépendants au même gène sont également proches. En effet, la séquence des allèles de résistance à *RYMV1* et *RYMV2* suggère que les différents allèles identifiés ne dérivent pas les uns des autres mais proviennent de mutations indépendantes.

Les accessions porteuses de l'allèle *rymv1-4* sont notamment souvent plus proches d'accessions portant l'allèle *rymv1-3* que d'une autre accession avec le même allèle. De même, les accessions ayant un des allèles de résistance candidat à *RYMV2* sont proches des accessions ayant l'allèle de résistance connu à ce gène. Les accessions portant l'allèle *rymv2-4* sont plus distantes des autres dans l'arbre mais la seule accession IRIGIN porteuse de *rymv2-4*, Og133, est groupée avec les autres dans l'ACP.

Les quatre accessions *O. barthii* résistantes identifiées ont également été placées sur l'arbre de diversité (Figure 4.5). Les deux accessions porteuses de l'allèle de résistance à *RYMV3* sont très proches. Toutefois, contrairement à ce qui est observé dans l'espèce *O. glaberrima* et bien que faisant partie du même groupe génétique défini par Orjuela et al. (2014), l'accession porteuse d'un allèle de résistance putatif à *RYMV1* n'est pas particulièrement proche dans la diversité de celle portant *rymv1-3*.

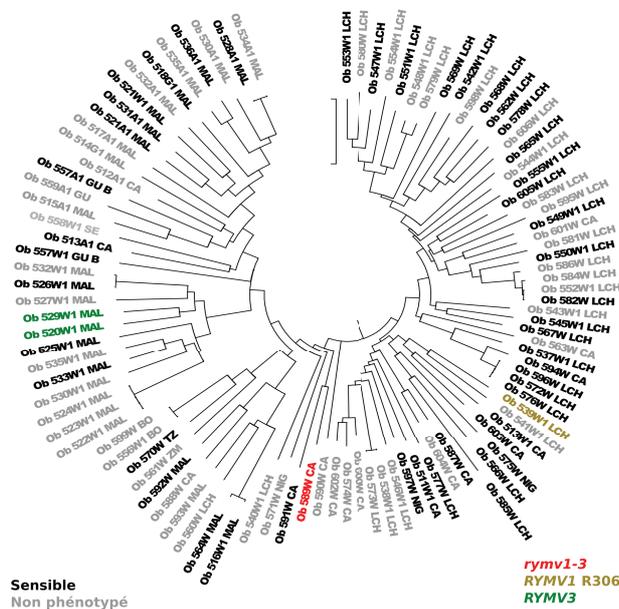


FIGURE 4.5 – Accessions *O. barthii* résistantes dans la diversité de l'espèce. Les quatre accessions *O. barthii* résistantes ont été placées sur l'arbre phylogénétique publié par Orjuela et al. (2013). Les accessions sensibles sont indiquées en noir et les accessions non phénotypées en gris.

2.3.2 Répartition géographique

Nous nous sommes également demandé comment se structurait cette diversité de résistance géographiquement. Nous disposons pour cela du pays d'échantillonnage de

l'ensemble des accessions de la collection (Figure 2.1).

Nous avons donc étudié la distribution géographique des différents allèles de résistance (Figure 4.6). On observe une délimitation claire de la zone géographique des accessions possédant un allèle nul à *RYMV2* et de celles possédant un allèle de résistance à *RYMV1*. Les allèles de résistance à *RYMV2* sont présents au Mali, au Sénégal, en Guinée, en Sierra Leone, en Côte d'Ivoire et au Burkina Faso. Les allèles de *RYMV1* sont plus à l'est : au Nigeria, au Cameroun, au Tchad et en Tanzanie. Les accessions porteuses de résistance à ces deux gènes sont séparées par un axe longitudinal passant par le Bénin. Les accessions porteuses de l'allèle de résistance à *RYMV3* sont réparties de manière plus homogène : en Sierra Leone, au Mali et au Nigeria.

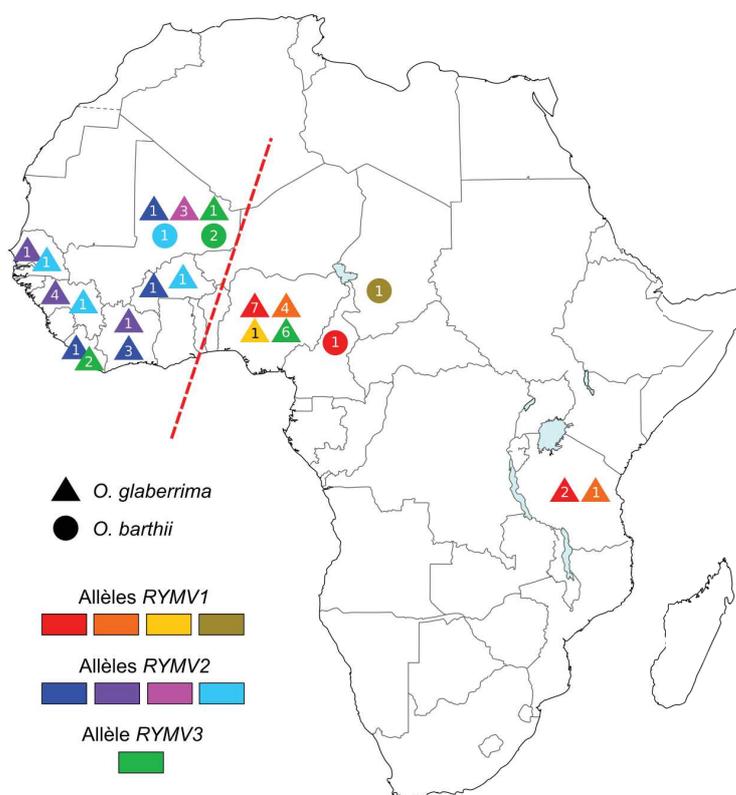


FIGURE 4.6 – Répartition géographique des accessions résistantes. Le nombre d'accessions *O. glaberrima* (triangle) et *O. barthii* pour chaque allèle de résistance validé ou putatif (couleurs) est indiqué par pays. La ligne pointillée rouge figure la séparation entre accessions portant un allèle de résistance à *RYMV1* et *RYMV2*.

3 Discussion

3.1 Fréquence des sources de résistance élevée chez les riz africains

Les riz africains présentent une proportion d'accessions résistantes au RYMV très importante, largement supérieure à celle identifiée dans l'espèce *O. sativa*. En effet,

Albar (1998) a évalué une collection d'environ 130 variétés *O. sativa* dont aucune n'a montré de résistance élevée, dans des conditions de phénotypage très similaires à celles utilisées ici. Jusqu'à ce jour, les seules sources de résistance élevée dans l'espèce *O. sativa* sont les variétés Gigante et Bekarosaka (Ndjondjop et al., 1999; Rakotomalala et al., 2008).

Les premiers, Thottappilly and Rossel (1993) ont identifié cinq accessions *O. glaberrima* et quatre accessions *O. barthii* hautement résistantes. Plus tard, Thiémélé et al. (2010) ont étudié la résistance de 337 accessions *O. glaberrima* et y ont identifié 29 accessions résistantes. Lors de notre crible de la collection IRD, nous avons identifié 41 accessions *O. glaberrima* résistantes sur les 278. Parmi les 29 accessions résistantes identifiées par Thiémélé et al. (2010), trois se sont avérées sensibles lors de notre crible, des lots de graines hétérogènes étant probablement à l'origine de ces résultats contradictoires. Les pourcentages apparents de résistance élevée dans la collection IRD (14,55 %), et dans une moindre mesure dans la collection de Thiémélé et al. (2010) (8,6 %) sont cependant une surestimation du taux réel de résistance au sein de l'espèce. En effet, des accessions décrites comme étant résistantes ont été incluses à ces collections, biaisant ainsi les effectifs. Sur la base d'une sous partie de la collection de Thiémélé et al. (2010), établie sans *a priori* sur le niveau de résistance des accessions, il semble que le pourcentage d'accessions hautement résistantes dans l'espèce *O. glaberrima* soit plutôt de l'ordre de 7 %. Nous avons également mis en évidence quatre accessions *O. barthii* résistantes sur les 103 présentes dans la collection. Cependant le phénotypage de la collection d'*O. barthii* n'est encore que très partiel. La multiplication en est difficile à cause d'un égrainage important, la levée est plus aléatoire que celle des accessions *O. glaberrima* et les lots de graines sont parfois hétérogènes. Toutes ces caractéristiques rendent difficile le phénotypage et son lien avec le génotype. Malgré ces difficultés, il semble cependant que le pourcentage de résistance chez *O. barthii* soit du même ordre de grandeur que celui observé chez *O. glaberrima*. La diversité génétique du compartiment sauvage étant plus importante que celle du compartiment cultivé, l'intérêt de la collection d'accessions *O. barthii* est important pour identifier de nouveaux gènes ou allèles de résistance, mais également pour mieux comprendre la dynamique évolutive de la résistance au RYMV. Un phénotypage rigoureux de la résistance au RYMV de l'ensemble de la collection *O. barthii* est donc souhaitable.

La différence de fréquence de résistance dans les riz africains et asiatiques n'est à ce jour, pas clairement expliquée. En effet, la divergence entre les souches de RYMV suggère que l'origine du RYMV remonte à 200 ans (Fargette et al., 2008b) et les symptômes du RYMV ont été rapportés pour la première fois en 1974 en Afrique de l'Est (Bakker, 1974) et en 1977 en Afrique de l'Ouest (Fauquet and Thouvenel, 1977). À cette période, la culture de variétés *O. sativa* était déjà bien établie en Afrique de l'Ouest (Portères, 1962). La pression de sélection exercée sur l'espèce *O. sativa* ne devrait donc pas avoir été moindre que celle exercée sur *O. glaberrima*. Il faut cependant noter que les deux sources de résistance identifiées dans l'espèce *O. sativa* sont des

variétés traditionnelles africaines (Awoderu, 1991; Michel et al., 2008; Rakotomalala et al., 2008). Il serait intéressant de cribler spécifiquement ce type de variétés, conservées par exemple au centre de ressources génétiques du Centre du Riz pour l'Afrique, pour évaluer la fréquence de résistance chez ces variétés et la comparer à ce qui est observé chez *O. glaberrima*

3.2 Gènes et allèles de résistance identifiés

Les résistances des accessions que nous avons identifiées sont conférées par trois gènes différents.

3.2.1 *RYMV1*

Le premier gène identifié, *RYMV1*, est récessif. Il code pour eIF(iso)4G1, un facteur d'initiation de la traduction (Albar et al., 2006). Il s'agit d'un gène de sensibilité de la plante : la protéine eIF(iso)4G1, qui est nécessaire au virus pour compléter son cycle dans la plante. Le virus détourne l'utilisation de cette protéine à son profit *via* une interaction entre le domaine MIF4G d'eIF(iso)4G1 et sa VPg. L'interaction du virus *via* une de ses protéines, souvent la VPg, avec un facteur d'initiation de la traduction est un mécanisme conservé dans les processus d'infection virale des plantes (Sanfaçon, 2015). Le facteur d'initiation recruté et la protéine interagissant avec lui ne sont toutefois pas toujours de même nature. Dans le cas de l'interaction du riz/RYMV, les polymorphismes de résistance conduisent à une rupture de l'interaction entre la VPg et *RYMV1*, aboutissant à la résistance de la plante (Hébrard et al., 2010).

Jusqu'à cette étude, trois allèles de résistance avaient été identifiés dans l'espèce *O. glaberrima* et un dans l'espèce *O. sativa* (Albar et al., 2006; Thiémélé et al., 2010). Nous avons montré que les allèles *rymv1-3*, *rymv1-4* et *rymv1-5* sont respectivement présents dans neuf, cinq et une accessions *O. glaberrima* de la collection IRD résistantes et que *rymv1-3* est également porté par une accession *O. barthii*. Les quatre allèles de résistance codent pour des substitutions et/ou des délétions dans le domaine MIF4G de la protéine, interagissant avec la VPg. Dans le cas de la résistance du riz au RTSV, la résistance est également acquise par des mutations dans le facteur eIF4G (Lee et al., 2010), cependant celles-ci ne sont pas situées dans le domaine MIF4G mais légèrement en amont de celui-ci. De la même manière, les mutations d'eIF4E conférant la résistance du piment à plusieurs potyvirus sont situées dans deux hélices en surface de la protéine (Charron et al., 2008). Les mutations de résistance dans les facteurs d'initiation de la traduction affectent donc les régions impliquées dans l'interaction avec les protéines virales et dépendent du pathosystème. Par ailleurs, il est important de noter que les facteurs d'initiation de la traduction jouent un rôle clé dans le fonctionnement de la cellule végétale. Les mutations permettant la résistance ne doivent pas affecter les fonctions et interactions clés dans la cellule végétale et doivent donc être en nombre limité.

Nous avons identifié un nouveau polymorphisme dans le domaine MIF4G dans une accession *O. barthii* résistante. Une autre mutation a été observée, mais elle est localisée en bordure du domaine MIF4G et est plus éloignée des mutations de résistance précédemment identifiées. Des études de génétique ou d'interaction avec la VPg devront être réalisées pour valider le rôle de ces mutations dans la résistance. Nous avons également identifié, dans trois accessions *O. barthii* sensibles, une mutation codante dans le domaine MIF4G, induisant la substitution d'un acide aspartique par une alanine à l'acide aminé 303 de la protéine. Ce polymorphisme ne confère pas la résistance et correspond à l'acide aminé porté par les accessions *O. sativa*. Il pourrait être impliqué dans l'interaction eIF(iso)4G1/VPg et dans le contournement de la résistance (Poulicard et al., 2012).

3.2.2 RYMV2

Le second gène de résistance, *RYMV2*, est également récessif. Il a été identifié par Orjuela et al. (2013) dans sept accessions *O. glaberrima*. Le gène candidat pour cette résistance code pour une protéine CPR5-1, tronquée et non fonctionnelle dans les accessions résistantes. Ce gène de résistance n'a pas été validé par complémentation fonctionnelle d'une lignée résistante mais son rôle a été confirmé en utilisant des mutants d'insertion T-DNA (Albar, communication personnelle). Deux lignées indépendantes de mutants d'insertion T-DNA de la variété *O. sativa* spp. *japonica* Hwayoung (Jeon et al., 2000; Jeong et al., 2006) ont été utilisés. Ils présentent chacun une insertion dans un intron du gène et sont résistants au RYMV alors que le sauvage est sensible.

L'homologue de *CPR5-1* chez *A. thaliana*, *CPR5*, est une nucléoporeine faisant partie du complexe formant le pore nucléaire et sa surexpression diminue les réactions de défense (Gu et al., 2016; Wang et al., 2014b). CPR5 forme des homo-oligomères qui sont rompus lors de la signalisation, induite notamment par les NB-LRR. Cela rend le nucléopore plus perméable aux signaux moléculaires, activant notamment les processus avals de l'ETI. Dans le cas du mutant *cpr5*, cette perméabilité est constitutive. Ce mutant présente ainsi une résistance constitutive à des agents pathogènes bactériens et à des oomycètes (Bowling et al., 1997). Cette résistance est associée à une augmentation du niveau d'acide salicylique dans la plante et à l'activation du gène *pathogenesis-related 1*.

Lors de l'étude de l'infection de l'accession Tog7291, présentant un allèle nul au gène *CPR5-1*, avec plusieurs agents pathogènes du riz étudiés à Montpellier (*Magnaporthe oryzae*, *Xanthomonas oryzae* pv. *oryzae* et *oryzicola* et *Meloidogyne graminicola*), cette accession s'est révélée sensible (Albar, communication personnelle). De plus, le phénotype des mutants *cpr5* d'*A. thaliana* est très altéré et leur croissance est réduite, ce qui n'est pas le cas de Tog7291 ou des mutants T-DNA. Le riz, comme le maïs, le sorgho ou encore *Brachypodium distachyon*, a deux copies de *CPR5*. Les mécanismes de régulation cellulaire contrôlés par les homologues de *CPR5* pourraient donc y être plus complexes et impliquer notamment des phénomènes de redondance fonctionnelle entre

les deux protéines homologues.

Deux hypothèses peuvent donc être formulées sur le mécanisme de la résistance conférée par un allèle nul de *CPR5-1* dans le riz. La première possibilité est que, comme chez *A. thaliana*, la perte de fonction de *CPR5-1* pourrait entraîner une résistance constitutive. La sensibilité aux autres agents pathogènes testés n'est pas en faveur de cette hypothèse, mais ne suffit pas non plus à la rejeter. En effet, le mutant *cpr5* d'*A. thaliana* n'est pas résistant à l'ensemble des agents pathogènes testés. La deuxième hypothèse est que *RYMV2* pourrait être, comme *RYMV1*, un facteur de sensibilité avec lequel le virus doit interagir pour infecter la plante. Cette interaction pourrait par exemple être nécessaire pour que le virus pénètre dans le noyau de la cellule pour y accomplir une partie de son cycle. La présence du virus dans le noyau des cellules infectées a en effet été établie, bien que son rôle dans le cycle viral soit inconnu à ce jour (Opalka et al., 1998). Dans cette hypothèse, les mutations de la protéine P2a du virus qui sont impliquées dans le contournement de la résistance de Tog7291 (Pinel-Galzi et al., sub) pourraient permettre l'établissement d'une interaction avec la protéine homologue *CPR5-2*. L'interaction entre la polyprotéine 2a, et les protéines *CPR5-1* et *CPR5-2* pourrait notamment être testée en système hétérologue afin de vérifier cette hypothèse. Une autre possibilité pourrait être d'évaluer la capacité d'infection des variants mutés dans la P2a sur des mutants perte de fonction de *CPR5-2*.

Nous avons mis en évidence cinq autres polymorphismes rares conduisant à une protéine *CPR5-1* tronquée dans l'espèce *O. glaberrima*. Les 12 accessions porteuses d'un de ces polymorphismes sont toutes résistantes et ne portent pas d'autre allèle de résistance connu. Ces cinq polymorphismes constituent donc des allèles de résistance candidats. L'un d'eux est présent dans six accessions et un second dans trois. Les trois autres allèles sont portés par une seule accession de la collection IRD. Dans l'espèce *O. barthii*, une accession présente également un polymorphisme conduisant à une protéine tronquée. Des croisements sont actuellement en cours entre les accessions porteuses des allèles candidats, Tog7291 et des accessions sensibles afin de réaliser des tests d'allélisme et de colocalisation des déterminismes de résistance. D'autres croisements avec des accessions *O. glaberrima* sensibles ont également été réalisés afin de confirmer la colocalisation entre *RYMV2* et la résistance.

Si les allèles identifiés confèrent, comme attendu, la résistance au RYMV, *RYMV2* aurait alors sept allèles de résistance. Les décalages du cadre de lecture qu'ils engendrent sont situés à divers niveaux de la protéine. La délétion d'une paire de base initialement identifiée chez Tog7291 est située au tout début du gène alors que le codon STOP d'Og256 est situé après le codon 471, sur les 516 codons du gène chez les accessions sensibles. La protéine est donc tronquée de moins de 10 % de sa longueur. Une courte délétion pourrait donc suffire à conférer la résistance au RYMV. Chez *A. thaliana* certains mutants perte de fonction de *CPR5* ne présentent qu'une substitution d'acide aminé (Gu et al., 2016). Cependant, les substitutions que nous avons observées chez certaines accessions *O. barthii* ne sont pas spécifiques d'accessions résistantes.

À notre connaissance, aucun autre gène de résistance naturel décrit ne correspond à un homologue de *CPR5*, faisant de *RYMV2* le premier gène de résistance de ce type associé à des allèles de résistance naturels.

3.2.3 *RYMV3*

Le troisième gène identifié confère une résistance dominante. L'analyse du polymorphisme des gènes candidats identifiés au chapitre précédent conforte le gène CC-NBS-LRR comme meilleur candidat. Les gènes de résistance de type NB-LRR sont très fréquents : vingt gènes NB-LRR conférant une résistance à un virus ont déjà été clonés (Boualem et al., 2016). L'allèle de résistance, présent dans huit accessions *O. glaberrima* et deux accessions *O. barthii*, serait une substitution dans le domaine LRR, or il a été montré que le domaine LRR des NB-LRR joue un rôle central dans la reconnaissance de l'agent pathogène (Farnham and Baulcombe, 2006; Sekine et al., 2012; Tomita et al., 2011).

Deux autres haplotypes au gène sont également associés à la résistance. L'un est porté par Og87 et Og103, également porteurs de l'allèle *rymv2-2*, et l'autre est porté par les accessions résistantes Og20 (*rymv1-4*) et Og36 (*rymv1-3*). Des études de diversité à des gènes de résistance de type NB-LRR ont montré que des allèles différents, caractérisés par des mutations distinctes dans le domaine LRR ou dans les domaines NB-ARC et LRR pouvaient aboutir à la résistance, parfois avec une spécificité différente (Bhullar et al., 2010; Ellis, 1999; Ravensdale et al., 2012; Seeholzer et al., 2010; Thakur et al., 2015; Vasudevan et al., 2015; Xu et al., 2014). Ces accessions devront être étudiées dans des croisements avec des *O. glaberrima* sensibles afin d'étudier la ségrégation de leur résistance et vérifier la colocalisation d'un potentiel deuxième déterminisme de résistance avec *RYMV3*. Toutefois le contournement de la résistance d'Og20 par le RYMV est très facile (Hébrard, communication personnelle), ce qui n'est pas cohérent avec la présence de deux gènes de résistance aux mécanismes différents dans cette accession. Par ailleurs, plusieurs gènes de résistance partielle clonés sont des NB-LRR (Broglie and Butler, 2008; Giannakopoulou et al., 2015; Hayashi et al., 2010). Ces mutations pourraient donc également conférer une résistance partielle.

3.2.4 Conclusion sur les sources de résistance de la collection IRD

À l'issue de cette étude, nous avons probablement identifié tous les allèles de résistance présents dans la collection d'accessions *O. glaberrima* de l'IRD. Les déterminants de résistance identifiés sont nombreux et couvrent une diversité de mécanismes intéressante avec notamment un gène de sensibilité (*RYMV1*), un gène de résistance classique (*RYMV3*) et un gène original dont le mécanisme de résistance reste encore à élucider.

Il est possible que certaines accessions *O. glaberrima* ne faisant pas partie de notre collection possèdent des allèles rares originaux. Il serait donc intéressant de rechercher les variations alléliques aux trois gènes identifiés, ainsi qu'à d'autres gènes similaires

tels que des gènes des mêmes familles multigéniques ou intervenant dans les mêmes complexes, dans une collection plus large comme celle de l'AfricaRice qui compte plus de mille accessions *O. glaberrima*. Des approches de séquençage ciblé ou par capture à un coût acceptable permettent actuellement d'envisager ce type d'étude (Cronn et al., 2012).

Par ailleurs, l'espèce *O. barthii* est plus diverse que l'espèce *O. glaberrima* et le nombre d'accessions étudiées reste faible vis-à-vis de cette diversité. Il est donc probable que la diversité naturelle de plantes *O. barthii* recèle d'autres allèles ou gènes de résistance au RYMV. De même, le RYMV infecte un grand nombre de poacées sauvages. D'autres espèces présentes en Afrique dans les zones de présence du RYMV, comme *Oryza longistaminata*, ont certainement subi une pression de sélection exercée par le RYMV et pourraient contenir des sources de résistance. En effet, cette espèce est pérenne et allogame, elle est donc supposée être très diverse et est relativement éloignée des espèces autogames du genre *Oryza*. Elle pourrait donc avoir accumulé une diversité de résistance intéressante. Le gène de résistance *XA21* à *Xanthomonas oryzae* a notamment été identifié dans cette espèce (Ikeda et al., 1990). Par ailleurs, huit accessions de l'espèce sont en cours de séquençage dans le cadre du projet IRIGIN et une analyse préliminaire de phénotype suggère que plusieurs d'entre elles pourraient être résistantes au RYMV.

3.3 Diversité des gènes et évolution

3.3.1 Profils de diversité aux gènes de résistance

Les gènes *RYMV1* et *RYMV2* montrent un profil de diversité génétique relativement similaire.

Au sein des 163 accessions *O. glaberrima* séquencées dans le projet IRIGIN et toutes phénotypées pour la résistance au RYMV, cinq mutations non synonymes ont été observées sur le gène *RYMV1*, dont trois sont localisées dans le domaine MIF4G et impliquées dans la résistance au RYMV. Pour comparaison, nous avons regardé la variabilité du domaine MIF4G dans les données des "3 000 génomes" de riz (Alexandrov et al., 2015; Li et al., 2014), incluant presque exclusivement des variétés *O. sativa*. Seules deux mutations non synonymes ont été observées et sont présentes à une fréquence inférieure ou égale à 3 %, même si on ne peut exclure que des polymorphismes de type INDEL n'aient pas pu être détectés par les approches utilisées. Ces mutations sont distinctes de la mutation dans le même domaine responsable de la résistance des variétés *O. sativa* Gigante et Bekarosaka, absentes du jeu d'accessions du projet "3 000 génomes". Ces résultats suggèrent que le gène *RYMV1*, et en particulier son domaine MIF4G, sont très conservés chez les accessions sensibles, ce qui est cohérent avec le rôle de la protéine dans la plante, et que la variabilité observée est en grande partie associée à la résistance au RYMV.

Les polymorphismes observés à *RYMV2* sont également peu nombreux dans notre collection *O. glaberrima*. On dénombre au total neuf mutations non-synonymes. Parmi

elles, six sont des allèles de résistance candidats. Dans les "3 000 génomes" ; nous n'avons détecté aucune mutation conduisant à une protéine tronquée. La protéine est donc également très conservée dans les accessions sensibles et la majorité du polymorphisme codant est associé à la résistance. Il semble donc y avoir une sélection balancée entre une protéine fonctionnelle très conservée et une protéine non fonctionnelle permettant l'acquisition de la résistance au RYMV. L'apparition des allèles de résistance pourrait être permise par une redondance fonctionnelle partielle avec *CPR5-2* qui diminuerait la contrainte évolutive à *RYMV2*. En effet, *CPR5-2* est également très conservée dans la collection *O. glaberrima* et ne présente qu'un polymorphisme non-synonyme conduisant à une substitution d'une alanine en valine, deux acides aminés très proches.

Par comparaison, Charron et al. (2008) ont observé un grand nombre de polymorphismes non-synonymes dans les exons et un faible nombre de polymorphismes dans les introns, les conduisant à suggérer une sélection diversificatrice au gène. Il est possible que nous soyons dans le même cas, malgré un moins grand nombre de polymorphismes non synonymes, qui pourrait notamment être causé par un temps de cohabitation moins grand entre le couple riz/RYMV que piment/potyvirus (Fargette et al., 2008b; Gibbs et al., 2008).

Le profil de polymorphisme observé au NB-LRR candidat pour *RYMV3* est très différent. Nous avons identifié 39 polymorphismes codants. Ce nombre de polymorphismes est très important, en particulier face au faible nombre observé dans les autres gènes de résistance au RYMV, et peut s'expliquer par la nature du gène. En effet les NB-LRR n'ont pas d'autre fonction dans la cellule que la résistance à un nombre d'agents pathogènes restreint. Ils subissent donc peu de contraintes évolutives et sont très souvent sous sélection diversificatrice (McDowell and Simon, 2006).

Des études plus complètes devront être réalisées sur les séquences à ces gènes de résistance en comparant les résultats obtenus dans nos séquences d'accessions *O. glaberrima* et un échantillon de variétés *O. sativa* issues des "3 000 génomes", afin de mieux comprendre les scénarios évolutifs en jeu.

Par ailleurs, quelque soit le gène considéré, nous avons observé un plus fort taux de polymorphismes dans l'espèce *O. barthii*. Ce résultat est attendu lorsque l'on compare une plante cultivée et son apparenté sauvage. Toutefois, si l'espèce *O. barthii* présente certains allèles de résistance originaux, elle n'a pas plus de déterminismes de résistance au RYMV que l'espèce *O. glaberrima*. La fixation des allèles de résistance dans l'espèce cultivée a pu être accélérée par une sélection anthropique. Par ailleurs, l'habitat des plantes sauvages est plus dispersé, elles attirent donc moins les insectes vecteurs du RYMV et n'ont que peu de contacts directs entre elles. L'expansion des allèles dans l'espèce pourrait donc être plus limitée.

3.3.2 Structuration de la diversité des allèles de résistance au RYMV

Nous avons noté que les accessions *O. glaberrima* porteuses des allèles de résistance à *RYMV1* et celles porteuses d'allèles nuls à *RYMV2* sont chacune regroupées dans une partie de la diversité de l'espèce. Or ces différents allèles semblent avoir évolué de manière indépendante. Deux hypothèses pourraient expliquer la proximité phylogénétique des accessions les portant : d'une part l'impact du fond génétique, en particulier la diversité allélique sur un gène agissant en interaction avec le gène de résistance, pourrait influencer sur la capacité de ce dernier à muter ; d'autre part l'impact de conditions environnementales pourrait à la fois avoir contribué à structurer l'espèce et à sélectionner des variants sur les gènes de résistance.

L'indice de fixation (F_{st}) entre le groupe d'accessions portant un allèle de résistance à *RYMV1* et celles portant un allèle de résistance à *RYMV2* est un moyen d'identifier des gènes ou régions structurant la différenciation entre ces groupes et qui seraient donc potentiellement candidats pour influencer la capacité de mutation d'un gène dans un groupe plutôt que dans l'autre. Cet indice de fixation, calculé par fenêtres glissantes de 5kb le long du génome (Figure 4.7), met en évidence de nombreuses régions divergentes entre les deux groupes, ce qui était attendu compte tenu de leur éloignement dans la diversité de l'espèce. Cependant, aucun des principaux gènes appartenant à la même famille multigénique que les gènes de résistance *RYMV1* et *RYMV2* (*eIF4G*, *eIF(iso)4G2* ou *CPR5-2*), ou intervenant dans le même complexe (*eIF4E*, *eIF(iso)4E*) n'est dans une région particulièrement divergente entre les deux groupes d'accessions. Cette analyse reste très préliminaire et devra être complétée par des analyses du F_{st} le long des différents gènes des familles multigéniques ainsi que par l'étude de ces F_{st} dans des accessions sensibles proches. Les gènes codant pour des protéines du pore nucléaire pourraient également être ciblés dans cette analyse.

L'apparition d'allèles de résistance dans un gène plutôt qu'un autre pourrait également provenir d'une pression de sélection s'exerçant différemment sur les gènes selon l'environnement. En effet, nous avons pu remarquer que les accessions porteuses d'allèles nuls à *RYMV2* sont originaires de régions plus à l'ouest que celles portant un allèle de résistance à *RYMV1*. Les accessions dont la résistance est conférée par *RYMV3* sont réparties de manière plus homogène, ce qui pourrait s'expliquer par la faible contrainte évolutive à *RYMV3*. La fixation d'allèles de résistance à ce gène pourrait donc être moins influencée par l'environnement.

La scission géographique entre les allèles à *RYMV2* et ceux à *RYMV1* se fait selon un axe longitudinal passant par le Bénin. Il est intéressant de noter que cet axe sépare également deux groupes phylogénétiques distincts de RYMV. Les analyses phylogéographiques du RYMV suggèrent en effet un goulot d'étranglement le long de l'axe longitudinal 0-2° (Pinel-Galzi et al., 2015; Trovão et al., 2015), en lien avec une discontinuité de la culture du riz entre ces deux espaces. À l'ouest, la culture du riz est historiquement importante et la région est nommée "Rice Belt" par Portères (1957)

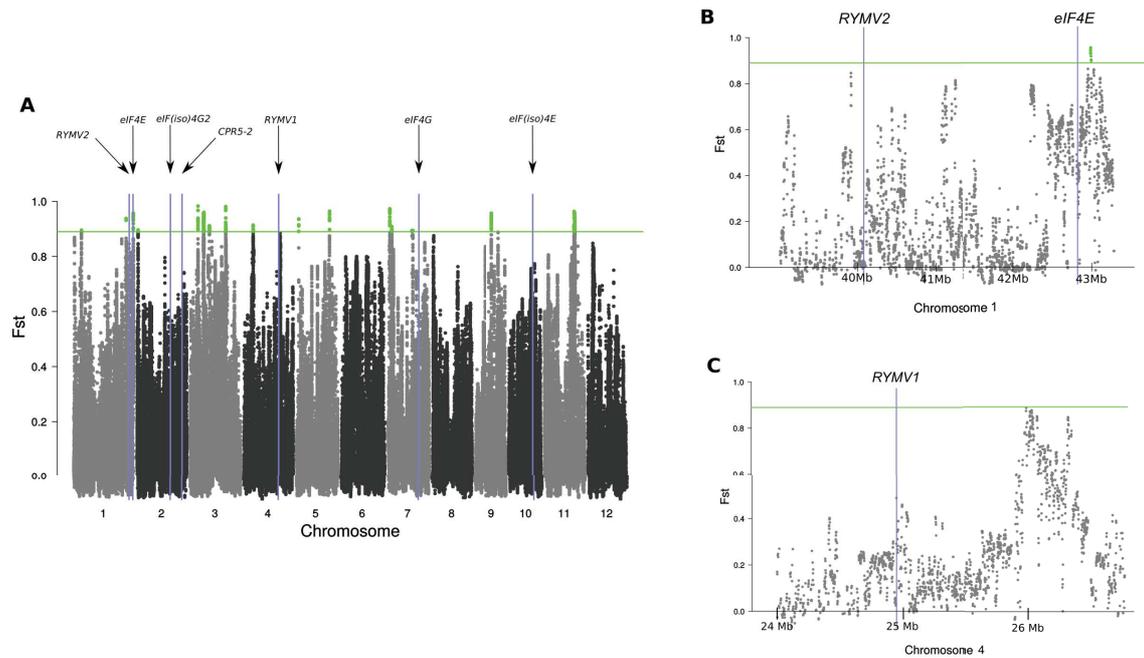


FIGURE 4.7 – Diagramme des F_{st} entre les accessions porteuses d'un allèle de résistance à *RYMV1* et à *RYMV2*. **A** : sur l'ensemble du génome. **B** : au chromosome 1, à proximité des gènes *RMYV2* et *eIF4E*. **C** : au chromosome 4, à proximité du gène *RMYV1*. Les SNP en vert sont ceux dont la valeur de F_{st} fait partie des 0,1 % les plus élevés. Les gènes correspondant à des facteurs d'initiation de la traduction ou aux homologues de *CPR5* sont placés sur le génome.

alors que la région située à l'est de l'axe, et qualifiée de "Yam Belt", est caractérisée par une culture importante d'igname et une culture plus clairsemée de riz. Par ailleurs, la phylogéographie de l'hôte semble avoir influencé celle du RYMV (Trovão et al., 2015).

Est-il possible que la diversité du virus ait influé sur l'apparition des résistances ? Le début de la diversification du RYMV à l'est de l'axe a été daté de la seconde moitié du XIX^e siècle, avec une expansion au XX^e siècle. Cette durée d'évolution d'un siècle est très longue à l'échelle de l'évolution du virus, en revanche elle est relativement courte vis-à-vis de celle de la plante. On ne peut cependant pas exclure qu'un ou plusieurs autres virus proches du RYMV aient sélectionné ces résistances dans le riz avant l'expansion du RYMV. Une autre possibilité est que d'autres facteurs environnementaux ou d'écotype influencent la capacité de *RYMV1* et *RYMV2* à évoluer vers la résistance.

Bien que la phylogénie de l'espèce *O. glaberrima* ne soit que peu influencée par sa géographie (Orjuela et al., 2014), il se peut également que la distribution géographique des résistances ne soit qu'un reflet de leur phylogénie et de la diversité à d'autres gènes, discutée plus haut, influencée par l'histoire de l'espèce.

CHAPITRE

5

Génétique d'association dans l'espèce
O. glaberrima

Chapitre 5 : Génétique d'association dans l'espèce *O. glaberrima*

1 Introduction

La résistance élevée au RYMV dans l'espèce *O. glaberrima* est bien caractérisée et Thiémélé et al. (2010) ont montré qu'elle contient des sources de résistance partielle. Par ailleurs, elle présente plus de sources de résistance élevée qu'*O. sativa*. Il est donc possible qu'elle recèle également une diversité originale de sources de résistance partielle. Jusqu'à présent, la résistance partielle au RYMV n'avait été étudiée que chez *O. sativa*, dans un croisement entre une variété de la sous-espèce *indica* (variété sensible IR64) et une de la sous-espèce *japonica* (variété partiellement résistante Azucena) (Albar et al., 1998; Boissard et al., 2007). Ces études ont permis d'identifier sept QTL dont trois majeurs sur les chromosomes 1, 2 et 12. Cependant, les deux espèces présentent une forte hétérogénéité. Les QTL détectés peuvent donc être en lien avec des caractères morphologiques et adaptatifs spécifiques des *japonica*, peu exploitables en sélection. L'espèce *O. glaberrima* est en revanche peu structurée et est donc bien adapté à la recherche de QTL de résistance.

La présente étude se base sur le séquençage d'accessions *O. glaberrima* du projet IRIGIN afin de réaliser une étude d'association sur génome entier (GWAS) de la résistance partielle au RYMV dans l'espèce *O. glaberrima*. Ce séquençage a été réalisé en Illumina sur 165 accessions *O. glaberrima* dont 30 hautement résistantes. Les séquences ont été mappées sur le génome de référence d'*O. sativa* Nipponbare et la profondeur de séquençage moyenne est de 35x.

Dans un étude d'association, un phénotypage de qualité est primordial. Nous avons pour cela mis en place un dispositif expérimental composé de six répétitions de l'ensemble des accessions et avons utilisé la mesure du contenu viral par test ELISA, plus répétable que la mesure des symptômes dans nos conditions expérimentales, comme évaluation du phénotype. Nous avons ensuite synthétisé les mesures obtenues puis appliqué des filtres sur les données de séquençage afin de ne retenir que l'information utile et limiter le risque de détection de faux positifs. Avec ces données, nous avons ensuite recherché des QTL de résistance partielle en utilisant plusieurs modèles de GWAS.

2 Matériel et méthodes spécifiques

2.1 Obtention des données de phénotype

2.1.1 Matériel végétal

L'étude de génétique d'association sur l'ensemble du génome a porté sur 125 accessions *O. glaberrima* non-hautement résistantes séquencées dans le projet IRIGIN (Table

A.1 en annexe). Le stock de graines multipliées à l'IRD était parfois trop faible pour le phénotypage et des graines multipliées au Burkina-Faso à partir de stocks IRD ont été utilisées pour 28 accessions.

Deux témoins *O. sativa* hautement sensibles (IR64 et Nipponbare) et un témoin *O. glaberrima* hautement résistant (Tog5681) ont également été utilisés.

2.1.2 Matériel viral

Les inoculations ont été réalisées avec l'isolat Ci4, décrit comme contournant peu les résistances des accessions d'*O. glaberrima* et moyennement agressif (Pinel et al., 2000; Poulicard et al., 2012).

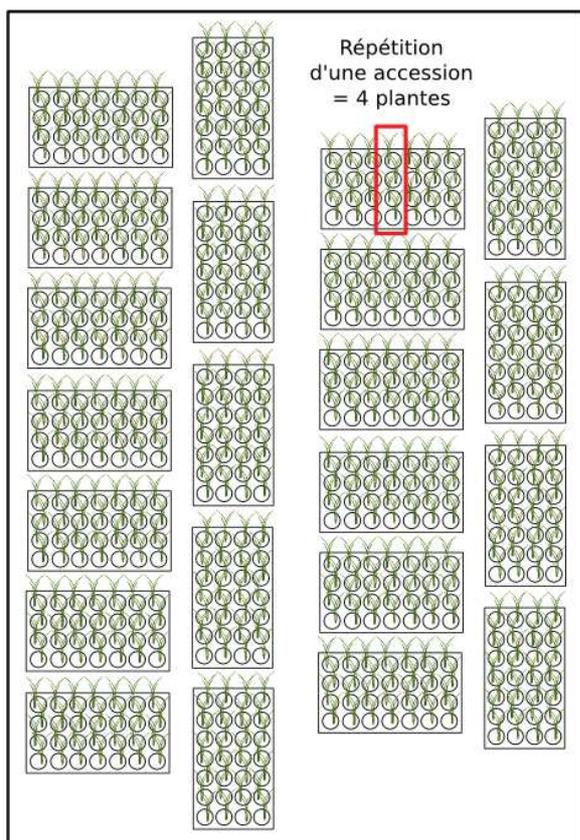
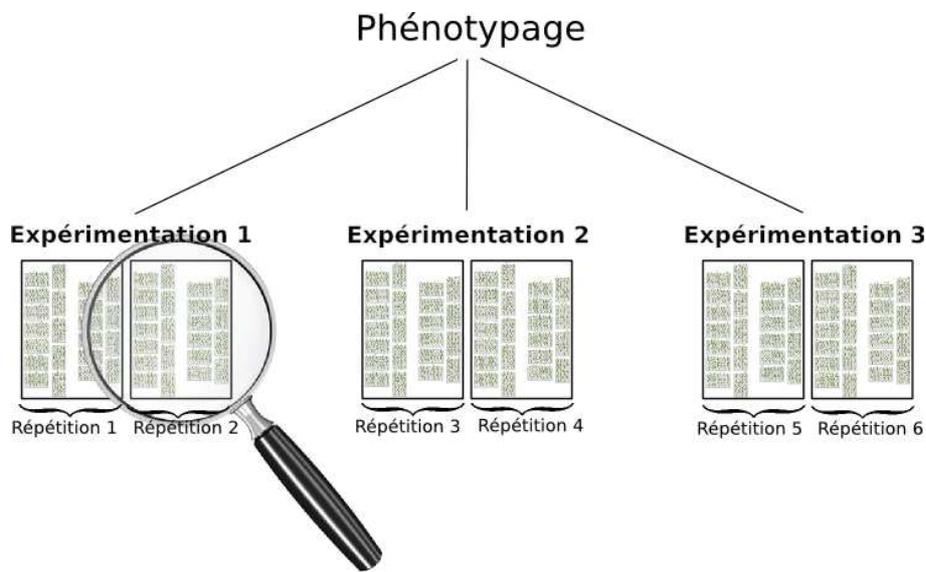
L'inoculation a été réalisée trois semaines après semis par infiltration de l'inoculum filtré à la seringue sans pointe en deux points à la base de la dernière feuille développée, comme décrit dans la partie 3.1 du chapitre 2. Cette méthode d'inoculation est supposée donner une meilleure répétabilité de la quantité d'inoculum utilisée et permet de repérer le point d'inoculation, différenciant la feuille inoculée des feuilles systémiques.

2.1.3 Dispositif expérimental

L'acquisition des échantillons de phénotypage s'est déroulée en serre de confinement tropicalisée et a impliqué six répétitions. Pour des raisons de place en serre, les six répétitions ont été divisées en trois expérimentations, chacune composée de deux répétitions. Chaque échantillon était composé de quatre plantes (Figure 5.1). Afin de limiter au maximum l'étendue temporelle entre les trois expérimentations, et donc la variabilité des conditions de luminosité, les semis des deux dernières expérimentations ont été réalisés en serre conventionnelle et n'ont été transférés en serre de confinement qu'au moment de l'inoculation, après l'élimination de l'expérimentation précédente. L'ensemble des expérimentations s'est ainsi déroulée sur une période de quatre mois. Pour limiter la variabilité environnementale, les positions des différentes accessions ont été randomisées au sein de chaque répétition et une rotation des terrines en serre a été appliquée deux fois par semaine.

Les symptômes étant peu caractéristiques sur *O. glaberrima*, ils n'ont pas été utilisés pour évaluer le niveau de résistance partielle. Un test ELISA semi-quantitatif a permis de quantifier l'accumulation virale dans la dernière feuille systémique. Pour mettre au point le test ELISA, nous avons sélectionné cinq accessions *O. glaberrima* (Og80, Og82, Og94, Og159 et Og405) ayant présenté différents niveaux de résistance au cours d'expériences préliminaires. Nous appellerons par la suite ces accessions "accessions références".

Deux échantillons par accession référence ont été inclus au sein de chaque répétition. Un premier échantillon a servi au calibrage de la mesure de l'accumulation virale, l'autre à l'évaluation du phénotype de l'accession. L'utilisation d'échantillons différents a permis d'éviter des événements de décongélation-recongélation des échantillons de



1 répétition = 137 accessions + témoins

FIGURE 5.1 – Plan de l'expérience de phénotypage. L'échantillonnage s'est déroulé en 3 expérimentations de 2 répétitions chacune. Une répétition se compose de 4 plantes de chacune des 131 accessions testées et d'accessions témoins.

phénotypage, risquant de les dégrader par rapport à ceux des autres accessions.

Certaines accessions confirmées ou mises en évidence comme hautement résistantes lors de la première expérimentation n'ont pas été intégrées aux autres expérimentations.

2.1.4 Mise au point et réalisation de la mesure de l'accumulation virale

Les prélèvements des échantillons nécessaires à l'ELISA ont eu lieu à cinq dates différentes : 10, 13, 17, 20 et 24 jours après inoculation (jai). Le prélèvement d'un échantillon était composé de 16 disques de feuille de 4mm de diamètre. Ces derniers étaient prélevés, à raison de quatre disques par plante, sur la dernière feuille développée.

Les prélèvements de feuilles ont tout d'abord été broyés au broyeur à billes puis suspendus dans 750 μ L de PBS-T 1X. À partir de cette suspension initiale, plusieurs dilutions ont été réalisées selon les besoins. La mesure de l'accumulation virale dans ces échantillons a été réalisée par le test DAS-ELISA décrit dans la partie 3.3 du chapitre 2. Le temps d'incubation avant lecture de la densité optique (DO) a été ajusté à chaque manipulation afin que la DO maximale lue sur une plaque soit comprise entre 2,3 et 2,6.

La courbe de mesure de la concentration virale en ELISA se présente comme une sigmoïde : elle comporte des zones de non-linéarité dans les concentrations faibles et dans les concentrations élevées. Afin d'obtenir une mesure proportionnelle à la concentration en virus dans les échantillons, nous avons recherché la dilution pour laquelle la DO de la majorité des échantillons se trouvait dans la phase linéaire. Pour choisir cette dilution, nous avons utilisé les accessions références.

Le premier échantillon de ces accessions a été testé à différentes dilutions pour chaque date et chaque répétition. À une date donnée, le facteur de dilution qui permettait de différencier au mieux les différentes accessions a été utilisé pour l'ensemble des échantillons à cette date.

Le nombre d'échantillons a nécessité la réalisation d'un grand nombre de plaques ELISA. Afin de comparer les données entre elles, une même gamme étalon, réalisée à partir de dilutions de broyat de feuilles de la variété *O. sativa* sensible IR64 infectées par l'isolat Ci4, a été déposée sur toutes les plaques ELISA. Elle est composée de 10 dilutions : 1g/100 ml (1:100), 1:400, 1:750, 1:1000, 1:2000, 1:5000, 1:10000, 1:25000, 1:50000 et 1:100000.

2.1.5 Traitement des données de phénotypage

L'ensemble des analyses des données de phénotypage a été réalisé avec le logiciel R (www.r-project.com).

Les mesures de DO de la gamme étalon ont été utilisées pour normaliser les valeurs de DO des accessions. Pour chaque plaque ELISA, les DO de gamme étalon ont été centrées-réduites et la transformation utilisée sur la gamme a été appliquée à l'ensemble des données de la plaque ELISA. Cette transformation est donc la suivante :

$$Mesure_{accession} = \frac{DO_{accession} - DO_{moyenne_{gamme}}}{\acute{E}cart\text{-}type_{gamme}}$$

Pour intégrer les données de contenu viral aux trois dates, une mesure inspirée de l'aire sous la courbe de progression de la maladie (AUDPC) a été calculée (Figure 5.2). Elle est calculée comme l'AUDPC mais ne peut être considérée comme une véritable AUDPC, les dilutions utilisées pour l'ELISA à chaque date étant différentes. Dans la suite cette mesure sera nommée pseudo-AUDPC (pAUDPC) La formule est :

$$pAUDPC = \sum_{i=1}^{N_i-1} \frac{y_i + y_{i+1}}{2} (t_{i+1} - t_i)$$

où N_i est le nombre de dates de mesure, t_i la date de mesure et y_i la valeur de DO normalisée par la gamme à la date i à laquelle la plus faible valeur de DO normalisée obtenue sur l'ensemble de l'expérience a été soustraite.

Nous avons réalisé des analyses en composantes principales pour observer la corrélation des différentes dates de prélèvements et répétitions. Pour cela, nous avons utilisé la fonction "PCA" du package "FactoMineR" dans le logiciel R. Les pourcentages de corrélations ont été visualisés graphiquement grâce au package "corrplot".

Afin d'obtenir une unique valeur de DO normalisée par accession pour chaque date et pour le calcul de la pAUDPC, le moyenne des valeurs aux différentes répétitions a été calculée.

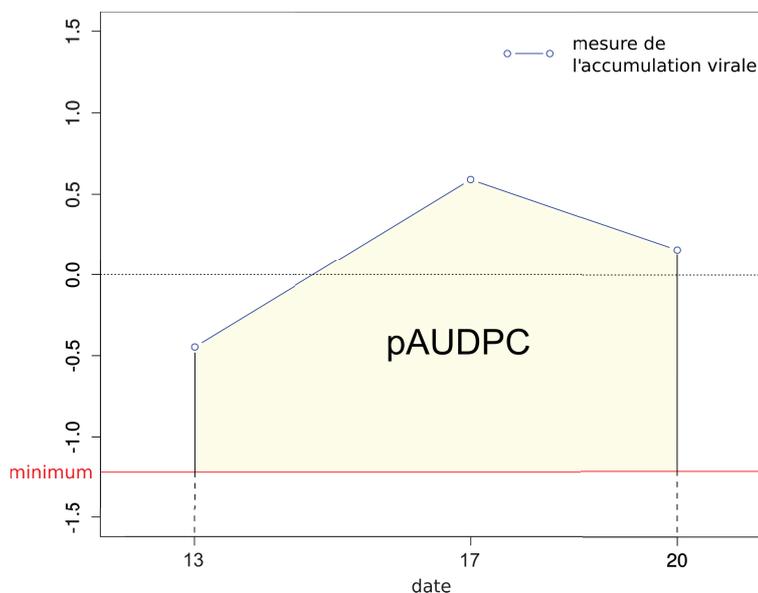


FIGURE 5.2 – Représentation de la pseudo-AUDPC.

Le plan d'expérience étant équilibré, l'héritabilité au sens large a été calculée pour

chaque data de prélèvement et pour la pAUDPC selon la formule :

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{B \times G}^2}{z} + \frac{\sigma_\epsilon^2}{nz}}$$

où σ_G^2 est la variabilité génotypique, $\sigma_{B \times G}^2$ est la variabilité de l'interaction génotype-expérimentation, σ_ϵ^2 la variabilité résiduelle, z le nombre d'expérimentations (3) et n le nombre de répétitions par expérimentation (2). Afin d'éviter de surévaluer l'héritabilité, les trois témoins (Tog5681, IR64 et Nipponbare) ont été retirés pour ce calcul.

Les différentes variances ont été obtenues avec un modèle mixte, établi dans le logiciel R avec le package "lme4". Les facteurs du modèle ont été traités comme des effets aléatoires.

2.2 Étude de génétique d'association

2.2.1 Traitement des données de séquences et analyse de la structure de population

Afin de réaliser l'étude de génétique d'association, nous avons utilisé les polymorphismes issus du séquençage IRIGIN que nous avons filtrés sur des critères supplémentaires. Nous n'avons conservé que les polymorphismes de type SNP *via* un filtre du logiciel "GATK". Nous avons éliminé les sites ayant plus de 5 % de données manquantes ou plus de deux allèles avec le logiciel "vcftools". Nos individus étant sensés être homozygotes, les sites présentant un grand nombre d'individus hétérozygotes sont certainement le reflet d'une duplication. Nous avons donc utilisé l'outil "SnpSift" pour supprimer les locus présentant plus de 10 % d'individus apparemment hétérozygotes.

Afin de choisir le modèle de génétique d'association adapté à la population utilisée, nous avons analysé la structuration de la population. En effet, celle-ci peut générer des faux positifs lorsqu'un SNP est spécifique d'une sous-population plus résistante qu'une autre en moyenne. En revanche, l'utilisation de données de structures comme co-facteurs dans le modèle d'association peut également générer des faux négatifs. Afin de choisir le modèle le plus adapté aux données, nous devons donc connaître la structure de la population et son importance. Nous avons pour cela échantillonné 10 000 SNP répartis de manière homogène sur le génome dans le fichier de polymorphismes obtenus. Nous avons ensuite utilisé l'outil "sNMF" implémenté dans le package R "LEA" pour obtenir la matrice de structure de population Q (Frichot and François, 2015).

Pour réaliser l'étude d'association, nous avons appliqué des filtres supplémentaires aux polymorphismes de la collection. Les allèles rares peuvent être artificiellement associés à la résistance et générer des faux positifs dans l'analyse. Nous avons donc filtré le fichier de polymorphismes obtenu sur la fréquence minimale des allèles (MAF) dans la population avec "vcftools" et n'avons conservé que les sites présentant des allèles présents dans plus de 5 % des individus. La détection de QTL dans un GWAS

reposant sur le déséquilibre de liaison (DL) entre les sites associés à la résistance et le polymorphisme causal, il est inutile d'étudier des SNP en DL trop important, porteurs d'une information redondante. Nous avons donc utilisé "PLINK" pour ne conserver que les locus présentant un r^2 inférieur à 0,8 dans une fenêtre de 50 SNP glissante par pas de 5 SNP. Nous avons ainsi obtenu un nombre final de 51 027 SNP.

2.2.2 Réalisation de l'étude de génétique d'association

Plusieurs modèles d'association génétique existent et se différencient par les co-facteurs utilisés. Afin de choisir le modèle le plus adapté à nos données, nous avons testé plusieurs modèles de génétique d'association : le modèle linéaire global (GLM), ne prenant pas en compte l'apparentement entre individus (matrice de *kinship* K), et un modèle mixte, le prenant en compte. Le modèle mixte utilisé est le modèle linéaire mixte compressé (CMLM) (Zhang et al., 2010) implémenté dans le package R "GAPIT" (Lipka et al., 2012). La matrice d'apparentement qu'il utilise a été calculée par le package "GAPIT" selon la méthode de VanRaden (2008). Le modèle GLM est implémenté dans TASSEL (Bradbury et al., 2007).

Nous avons testé chacun des deux modèles avec et sans utilisation de la structure (Q) de population comme co-facteur. Les composantes d'apparentements utilisées provenaient de l'analyse réalisée avec "sNMF". Le meilleur modèle a été sélectionné sur la base du critère d'information Bayésien (BIC) (Schwarz et al., 1978) et de l'apparence des Q - Q plots. Les Q - Q plots et *Manhattan plots* de résultats ont été générés avec le package R "qqman" (Turner, 2014).

Afin de connaître l'étendue d'un QTL, nous avons calculé le DL dans un intervalle de 2 Mb de part et d'autre des SNP significatifs avec "PLINK". Nous avons ensuite visualisé cette matrice avec le package R "LDheatmap" (Shin et al., 2006).

3 Résultats

3.1 Évaluation de la résistance partielle

3.1.1 Choix des dates et des dilutions

Afin de limiter le nombre de tests ELISA à réaliser, nous avons cherché à sélectionner les dates les plus informatives ainsi que les dilutions adaptées à chacune d'elles. Pour cela, plusieurs dilutions ont été testées sur les cinq accessions références pour chaque répétition et chaque date.

À 10 jai, une grande partie des échantillons a une DO très faible, quelle que soit la dilution. Il est donc difficile de distinguer un effet accession. À 24 jai l'ensemble des accessions a un contenu en virus élevé à toutes les dilutions. Aux trois autres dates en revanche, on distingue les différentes accessions. Au vu de ces résultats, les ELISA ont été réalisés sur l'ensemble des accessions sur les échantillons prélevés à 13, 17 et 20 jai.

La dilution permettant de mieux discriminer les différences entre accessions a été choisie pour chaque répétition et chaque date sur la base de la mesure de concentration virale des cinq accessions références avec plusieurs dilutions (Figure 5.3). Pour une date donnée, la dilution choisie a été la même pour l'ensemble des répétitions. Il s'agit de :

- 1:1 000 (1 g/1000 mL) pour 13 jai,
- 1:2 500 pour 17 jai,
- 1:5 000 pour 20 jai.

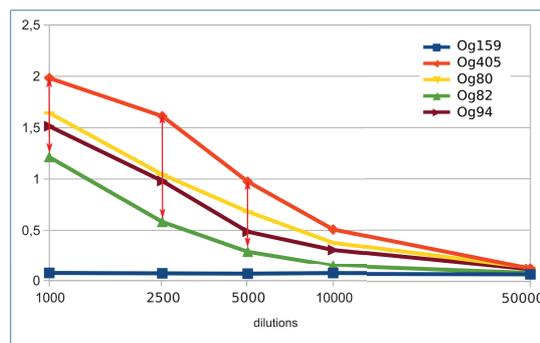


FIGURE 5.3 – Exemple de choix de dilution pour le test ELISA. Ce graphique représente la DO des accessions références en fonction de la dilution pour la répétition 5 à 17 jai. L'échelle des abscisses est logarithmique. À la dilution 1/1 000^e l'accession ayant la concentration la plus élevée commence à saturer. Les accessions avec un fort contenu en virus seront donc difficiles à différencier. À partir de la dilution 1/10 000^e la D.O. des différentes accessions est proche. La dilution 1/2 500^e est la plus adaptée.

3.1.2 Phénotypes obtenus

Lors des prélèvements, les symptômes sur les témoins sensibles *O. sativa* IR64 et Nipponbare se sont développés au bout d'environ deux semaines de manière homogène et l'accession résistante Tog5681 n'a présenté aucun symptôme. La plupart des accessions *O. glaberrima* ont exprimé des symptômes, bien que peu caractéristiques. Toutefois pour certaines accessions, les quatre plantes d'un échantillon montraient des symptômes contrastés.

Lors de ce test phénotypique, nous n'avons pas détecté de virus dans certaines accessions, de même que dans le témoin de résistance élevé Tog5681. Ces accessions ont alors été considérées comme hautement résistantes et éliminées de la suite de l'analyse présentée ici mais leur caractérisation a été décrite au chapitre précédent.

L'évolution de la charge virale relative au cours du temps est représentée par accession dans la figure 5.4 pour quelques accessions, et dans la figure B.1 en annexe pour l'ensemble des accessions. La mesure de l'accumulation virale est cependant relative, car réalisée sur des dilutions différentes pour les différentes dates de prélèvements. La comparaison entre dates de prélèvement doit donc être interprétée avec précaution. On peut distinguer plusieurs profils d'évolution de charge virales dans les différentes accessions. Certaines accessions ont un contenu viral relatif stable. Ainsi l'accession

Og56 a une résistance partielle élevée, tandis que l'accèsion Og86 est très sensible (Figure 5.4). D'autres accessions présentent un profil en cloche ou une augmentation du contenu viral relatif au cours du temps. On note que le contenu moyen en virus des accessions *O. glaberrima* n'est jamais supérieur à celui observé chez les témoins sensibles de l'espèce *O. sativa*, Nipponbare et IR64.

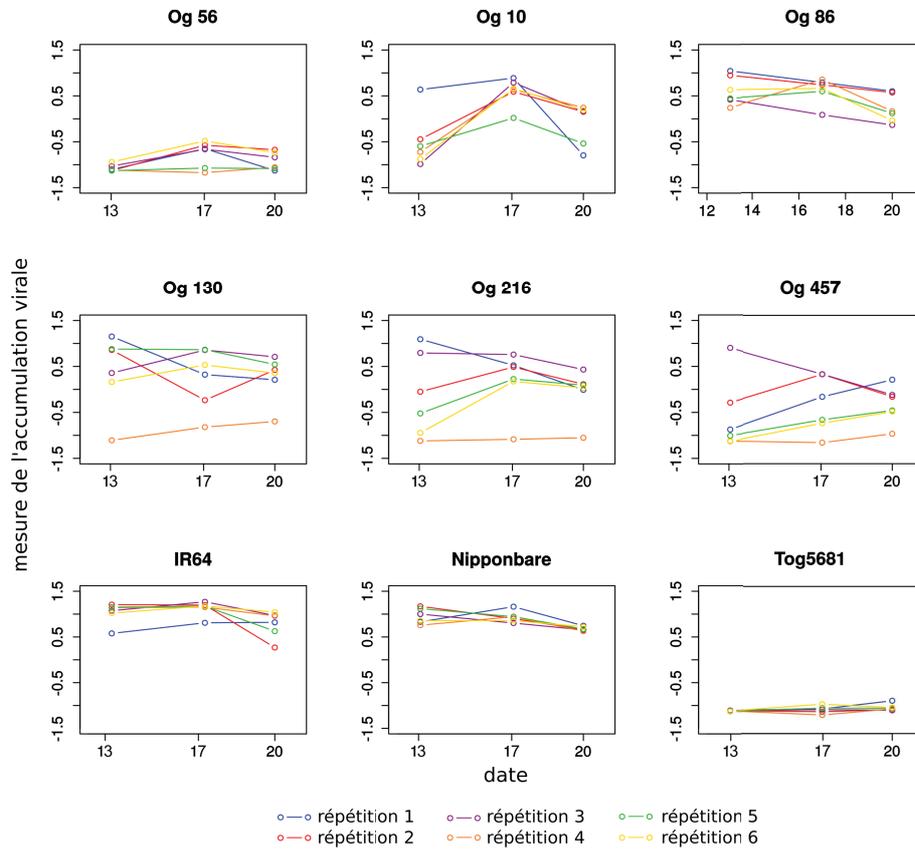


FIGURE 5.4 – Accumulation virale mesurée dans certaines accessions du test de phénotypage aux trois dates et six répétitions.

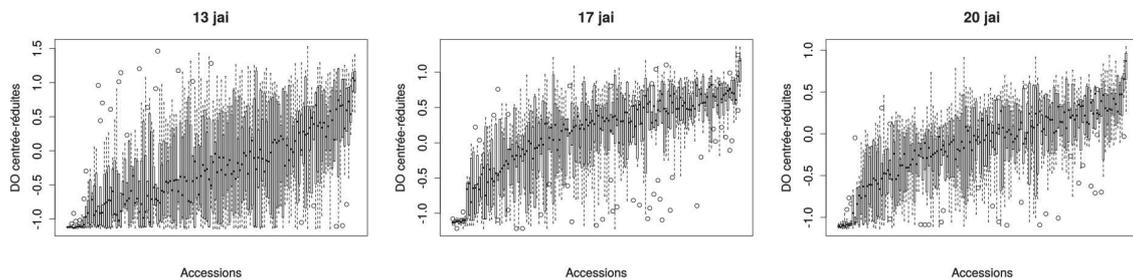


FIGURE 5.5 – Boxplot de la concentration virale aux trois dates dans les six répétitions en fonction des accessions. On note une variabilité importante au sein des accessions, plus marquée à 13 jai.

On observe également une variabilité importante entre les différentes répétitions de certaines accessions. La figure 5.5 représente, sous une forme plus synthétique, les variations de charge virale dans les accessions testées aux trois dates de prélèvements

et confirme l'importance de cette variabilité. Nous avons recherché les causes de la variabilité observée, notamment en recherchant son lien avec une date de prélèvement ou des répétitions particulières.

Les figures B.1 et 5.5 montrent que la mesure à 13 jai est notablement la plus variable.

Les corrélations entre répétitions confirment la variabilité observée : à la date 13 jai la corrélation moyenne entre répétitions est de 27 %, contre 42 % et 39 % à 17 et 20 jai.

Cette variabilité importante à 13 jai peut s'expliquer par la phase dans laquelle est la progression de la maladie à 13 jai. En effet, dans nos conditions, le niveau de multiplication à 10 jai est très faible. À 13 jai, le virus est donc en phase de propagation importante dans la plante. Un léger décalage de cette propagation entre les différentes plantes peut donc engendrer une différence importante de l'accumulation virale dans la feuille systémique.

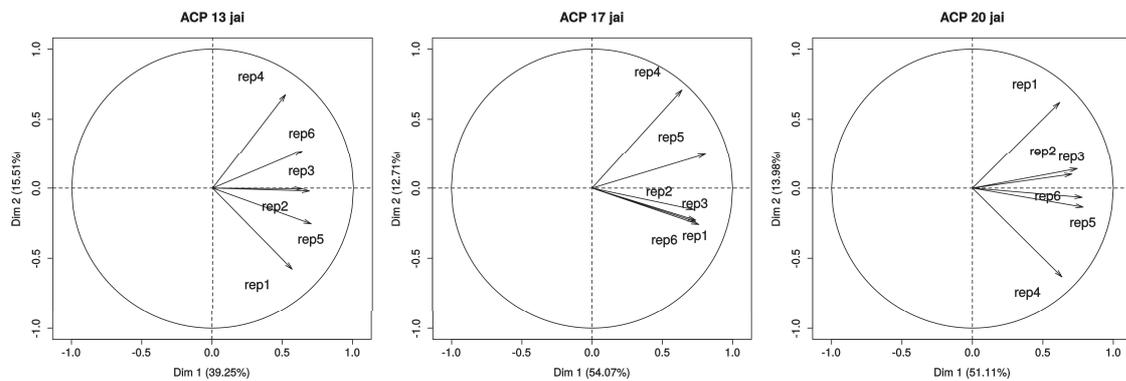


FIGURE 5.6 – Projection des répétitions selon les composantes 1 et 2 des ACP aux trois dates. Certaines répétitions sont peu corrélées aux autres. À 13 jai et 20 jai, les répétitions 1 et 4 se détachent et sont anti-corrélées. À 17 jai, la répétition 4 reste très peu corrélée aux autres mais la répétition 1 est elle bien corrélée. L'axe 1 de l'ACP à 17 jai et 20 jai exprime un pourcentage plus important des données qu'à 13 jai.

Par ailleurs, les ACP réalisées sur chacune des dates de prélèvements permettent d'observer les corrélations entre répétitions (figure 5.6). Il apparaît ainsi que la répétition 4, et dans une moindre mesure la répétition 1 sont moins corrélées aux autres répétitions. Elles sont également très peu corrélées entre elles.

À 13 jai, le contenu viral des plantes de la répétition 1 est en moyenne plus important que lors des autres répétitions. À 17 jai, l'effet est encore visible mais très peu marqué. Au contraire, les plantes de la répétition 4 ont un contenu viral moindre que celles des autres répétitions. L'effet décroît au cours du temps mais est encore marqué à 20 jai. La différence observée pour la répétition 4 peut s'expliquer par des conditions environnementales différentes. Les plantes de cette répétition ont subi une attaque d'acariens en début de croissance, maîtrisée avant leur inoculation et n'ayant pas affecté la croissance des plantes. De manière plus générale, les différences entre répétitions

peuvent être dues à des différences d'environnement ou à de légères différences lors de l'inoculation.

Nous avons calculé l'héritabilité au sens large aux trois dates de mesure et pour la pAUDPC. L'héritabilité calculée à 13 jai est de 0,388 ce qui est faible mais explicable par les deux répétitions extrêmes à cette date et la variabilité importante observée. En revanche à 17 et 20 jai, les héritabilités sont plus élevées, respectivement de 0,654 et 0,622. L'héritabilité de la pAUDPC est de 0,618, du même ordre de grandeur.

Pour chaque accession et pour chaque mesure d'accumulation virale, nous avons calculé une moyenne des six répétitions. L'allure des distributions obtenues montre une diversité du niveau de résistance partielle des accessions (Figure 5.7). La distribution à 20 jai est toutefois plus resserrée et les mesures à cette date présentent donc une diversité moindre. Par ailleurs, elles sont corrélées entre elles et sont liées par des relations linéaires (Figure 5.8). Le niveau relatif de résistance partielle d'une accession ne change donc pas fortement entre les dates.

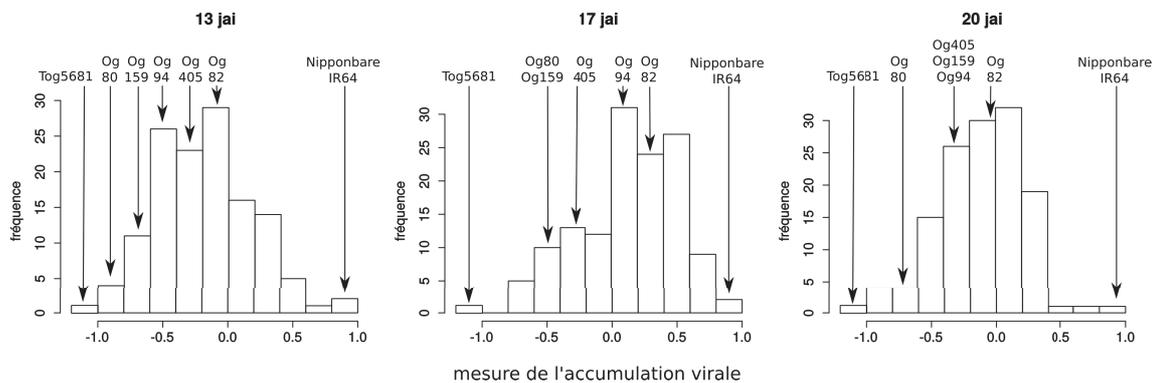


FIGURE 5.7 – Histogrammes de la distribution des moyennes de concentration virales des accessions aux trois dates. La position des accessions références ainsi que des témoins est indiquée.

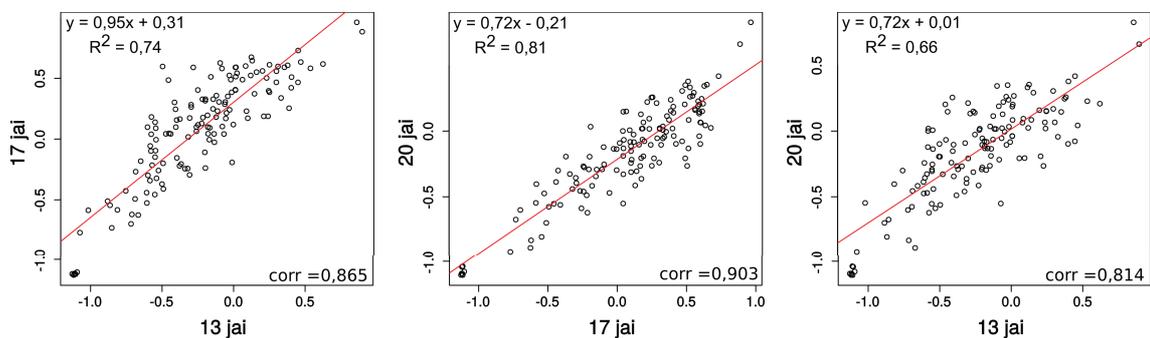


FIGURE 5.8 – Régression linéaire des valeurs de concentrations virales entre les trois dates. Les équations et les coefficients de détermination (R^2) des courbes, ainsi que la corrélation calculée entre les deux variables sont indiquées

Les héritabilités calculées pour les mesures à 17 et 20 jai ainsi que pour la pAUDPC sont supérieures à 0,5 et nous les considérons donc suffisantes pour une étude d'associa-

tion. De plus les niveaux d'accumulation virale obtenus couvrent une large gamme de valeur et devraient permettre de détecter des QTL de résistance.

3.2 Étude d'association

3.2.1 Structure de la population

Nous avons utilisé l'outil "sNMF" pour rechercher la structuration de notre population. Afin de déterminer le nombre de populations ancestrales le plus probable, cet outil utilise le critère de l'entropie croisée, qui doit être minimale. Nous avons donc tracé ce graphique pour une à 25 populations (Figure 5.9). Les valeurs d'entropie-croisée sont toutes faibles et décroissent jusqu'à 18 populations. Au vu de la faible diversité et de la faible structuration connue dans l'espèce (Orjuela et al., 2014), la structuration en 18 populations ancestrales est très improbable et nous ne la retenons pas.

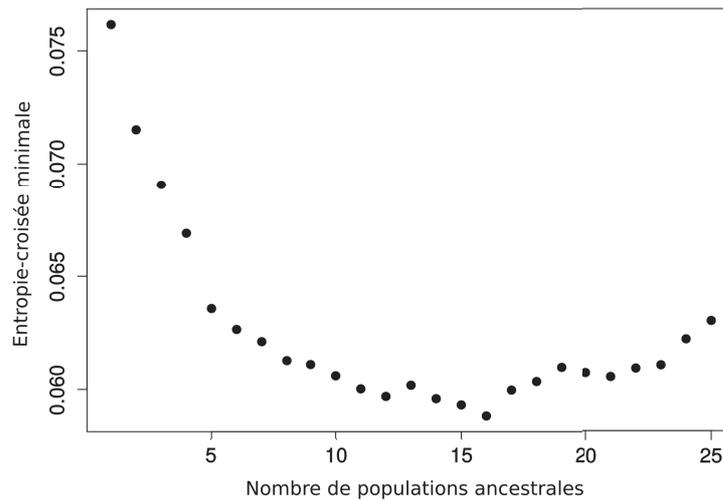


FIGURE 5.9 – Graphique de l'entropie-croisée calculé pour la structuration de la population en un à 25 groupes.

En revanche, l'ACP réalisée avec la fonction "pca" du package "LEA" semble former deux à quatre groupes relativement proches (Figure 5.10). Nous avons donc utilisé "sNMF" pour calculer le pourcentage de parenté de chacune des accessions avec deux, trois et quatre groupes. Le groupe pour lequel les accessions avaient le plus fort pourcentage d'appartenance leur a été assigné.

La structuration en deux groupes ne paraît pas particulièrement pertinente sur l'ACP de la diversité tant ces deux groupes sont proches. En revanche, les groupes d'accessions réalisés avec trois ou quatre populations ancestrales semblent plus distincts.

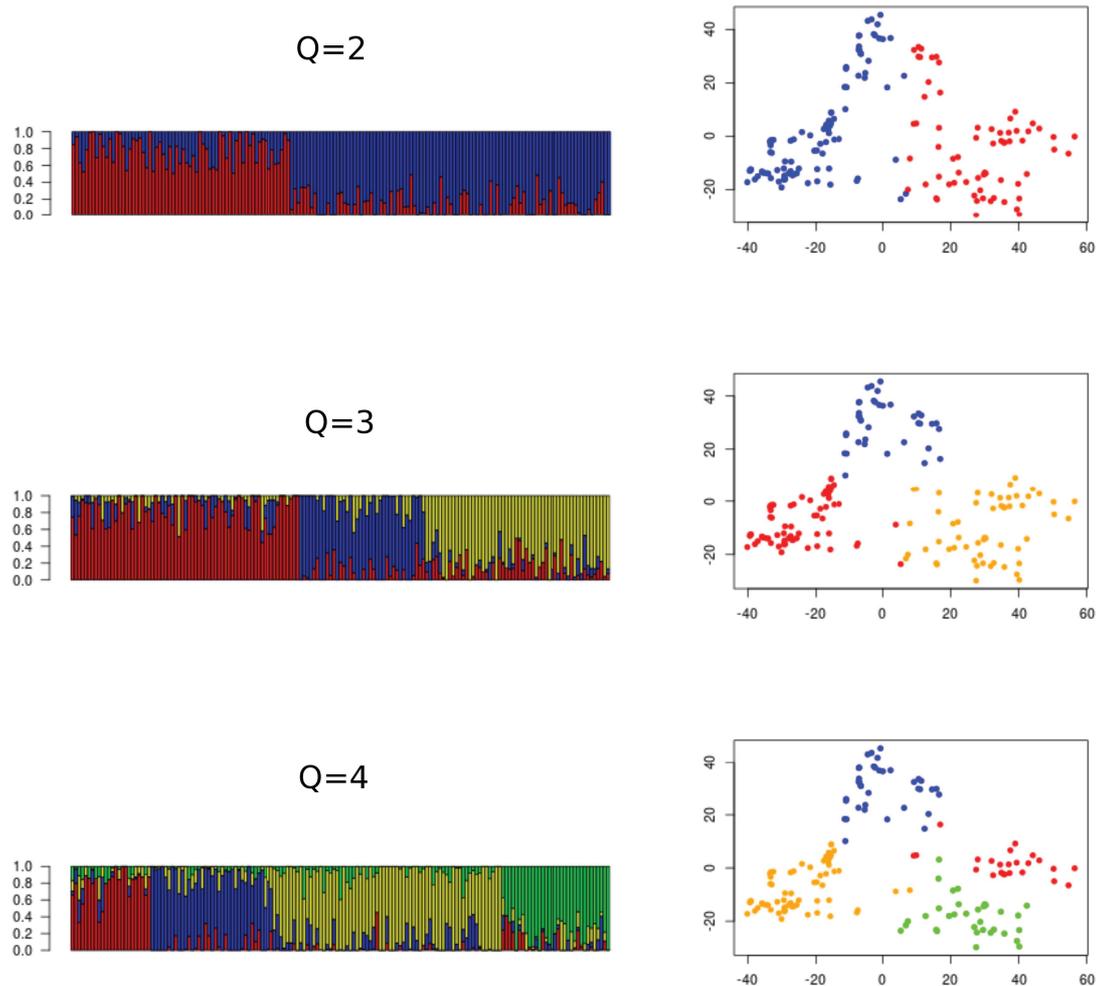


FIGURE 5.10 – Assignations des accessions *O. glaberrima* à deux, trois ou quatre groupes de structure. Les groupes de structure sont représentés en couleur sur l'ACP de la diversité et sur le diagramme en barres.

3.2.2 Choix du modèle

Afin de choisir le modèle le plus adapté à nos données, nous avons comparé les résultats obtenus avec le modèle GLM, ne prenant pas en compte l'apparentement entre accessions (K), à ceux obtenus avec le modèle CMLM, utilisant la matrice d'apparentement comme facteur aléatoire. Pour cela, nous avons comparé les $Q-Q$ plots obtenus avec chacun des modèles. Lors de cette comparaison, nous n'avons pas pris en compte la structure de la population et utilisé les données à 17 jai, les plus héritables. Le $Q-Q$ plot du modèle GLM montre une déviation importante des données par rapport à la normalité. En revanche, dans le modèle CMLM, les données sont très proches de la normalité (Figure 5.11).

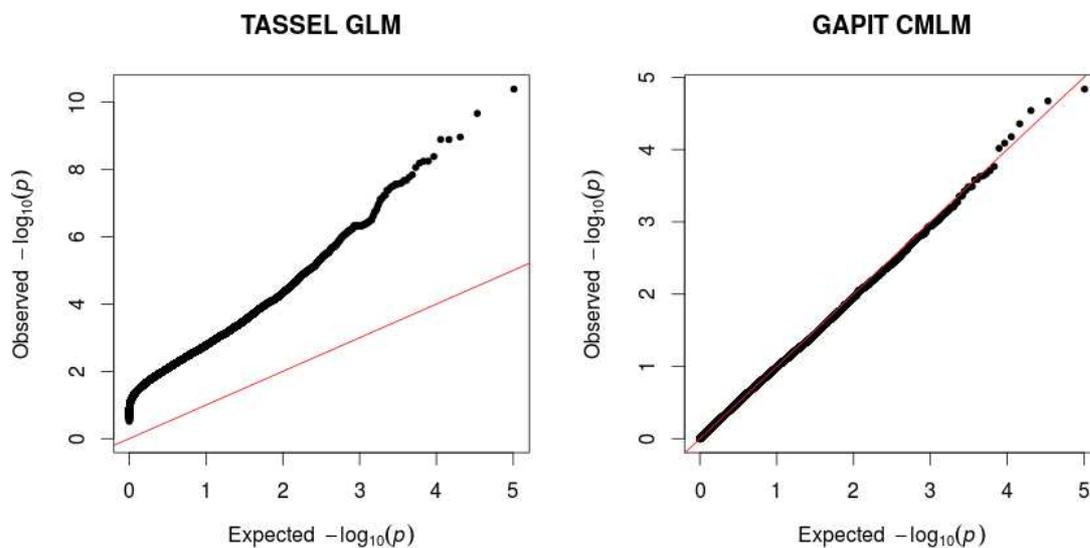


FIGURE 5.11 – $Q-Q$ plots des GWAS réalisés sur les données de mesure de l'accumulation virale à 17 jai avec les modèles GLM et CMLM

Nous avons ensuite cherché à savoir si l'ajout de la structure de population (Q) dans le modèle permettait de l'améliorer. Pour cela nous avons utilisé les pourcentages d'apparentement des accessions à trois ou quatre populations ancestrales, obtenus avec "sNMF". Nous avons comparé les modèles CMLM avec ou sans ajout de ces données de structure de population sur la base du BIC du modèle, calculé par "GAPIT". À 17 jai, le modèle sans structure de population présente un BIC de -49,78 alors que ceux des modèles avec trois et quatre groupes de structures ont respectivement des BIC de -56,20 et -58,13. Le meilleur modèle est donc celui sans structure de population. Les résultats obtenus pour les autres mesures sont similaires.

Nous avons donc, par la suite, utilisé ce modèle sur l'ensemble des variables de résistance partielle obtenues.

3.2.3 Détection des QTL

L'application du seuil de significativité de Bonferroni pour un seuil global $\alpha = 0,05$, calculé à 10^{-6} dans notre expérience, ne permet pas d'identifier de SNP associés à la résistance. Toutefois ce seuil est connu pour être beaucoup trop stringent car il considère que les tests réalisés sont indépendants, ce qui n'est pas le cas. Nous avons donc défini un seuil de p-value de 10^{-4} ($-\log_{10}(p) > 4$), malgré le risque de détecter des faux positifs.

Malgré cet abaissement du seuil, aucun SNP n'est significatif pour la mesure à 13 jai. L'héritabilité du caractère à cette date étant faible, le phénotype est donc imprécis et ce résultat n'est guère surprenant. En revanche à 17 jai, trois régions sont significativement associées à la résistance partielle, sur les chromosomes 3, 6 et 11 (Figure 5.12). Le profil du QTL du chromosome 11, formant un pic, est plus conforme au profil attendu pour un QTL réel que ceux des chromosomes 3 et 6 qui ne comprennent qu'un seul SNP supérieur au seuil que nous avons fixé. Il est intéressant de noter que le pic du chromosome 11 colocalise avec le gène *RYMV3*.

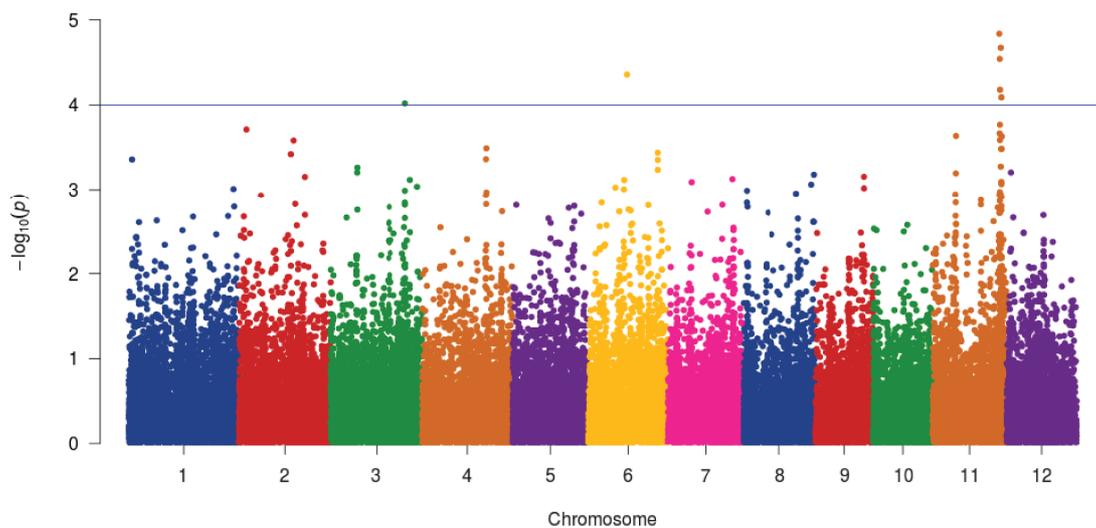


FIGURE 5.12 – *Manhattan plot* du GWAS réalisé sur les données de mesure de l'accumulation virale à 17 jai avec le modèle CMLM corrigeant par l'apparentement entre individus.

Les QTL des chromosomes 6 et 11 sont également significatifs pour la mesure à 20 jai et la pAUDPC, mais pas celui du chromosome 3 (Figure B.2 en annexe). Par ailleurs, la répétition 4 étant en nette déviation par rapport aux autres, nous avons également réalisé l'analyse de génétique d'association sur les valeurs moyennes des mesures d'accumulation virale aux cinq autres répétitions. La significativité des SNP est en moyenne légèrement moindre. Toutefois, on identifie également les pics des chromosomes 6 et 11, ainsi que des SNP significatifs supplémentaires sur les chromosomes 2 et 10 pour la mesure à 20 jai (Figure B.3 en annexe).

Nous avons réalisé un test de corrélation entre les différents phénotypes et le génotype obtenus avec les différentes variables de résistance partielle au SNP le plus significatif de chaque région où un QTL a été identifié (Table 5.1). Les SNP des chromosomes 6 et 11 sont corrélés à plus de 50 % aux différentes mesures de phénotype. Parmi les autres SNP, seul celui du chromosome 10 est corrélé à plus de 50 % à une des mesures (20 jai).

L'effet des SNP des chromosomes 6 et 11 sur l'accumulation virale à 17 et 20 jai est représenté sur la figure 5.13. "GAPIT" calcule que ces SNP expliquent une baisse d'environ 0,2 points de la concentration virale mesurée à 17 jai. Ces deux éléments confirment le gain de résistance des accessions présentant des polymorphismes sur les locus testés.

TABLE 5.1 – Corrélation entre les mesure de phénotype et le génotype aux marqueurs associés dans les différents modèles.

	Chr11 26 180 684	Chr6 15 132 378	Chr3 29 638 797	Chr10 18 549 929	Chr2 8 584 315
17 jai	0,55	0,54	0,32	0,39	0,20
20 jai	0,54	0,50	0,29	0,51	0,22
pAUDPC	0,54	0,53	0,26	0,43	0,23

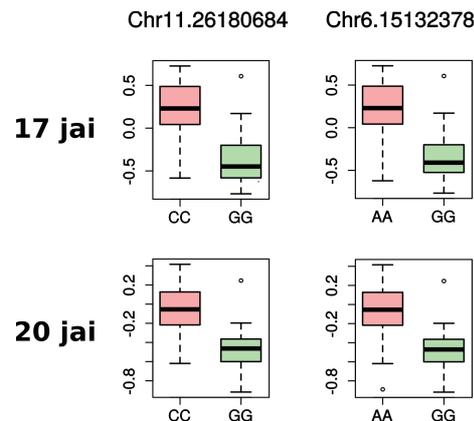


FIGURE 5.13 – Effet des SNP significativement associés à la résistance sur le phénotype d'accumulation virale à 17 et 20 jai.

Les QTL des chromosomes 2, 3 et 10 étant rarement détectés comme significatifs et moins associés au phénotype, nous nous sommes focalisés sur les QTL des chromosomes 6 et 11 pour la suite de l'analyse.

3.2.4 Étude préliminaire des QTL identifiés

Afin d'évaluer la largeur de l'intervalle des QTL des chromosomes 6 (QTL₆) et 11 (QTL₁₁), nous avons calculé le DL entre les SNP situés dans un intervalle de 4 Mb autour des SNP significatifs (Figure 5.14). Cette analyse a inclus des SNP présentant des DL élevés entre eux, non pris en compte lors de l'étape de détection des QTL. Le SNP significatif sur le chromosome 6 est très proche du centromère et le DL dans

la région est très faible. Les centromères recombinant peu, on pourrait s'attendre à un DL important. Cependant la densité en marqueurs est faible, pouvant expliquer cette différence. En revanche, le DL dans la région du chromosome 11 est relativement important sur une région de 2,1 Mb autour des cinq SNP significatifs, entre les positions 25 800 000 et 27 900 000.

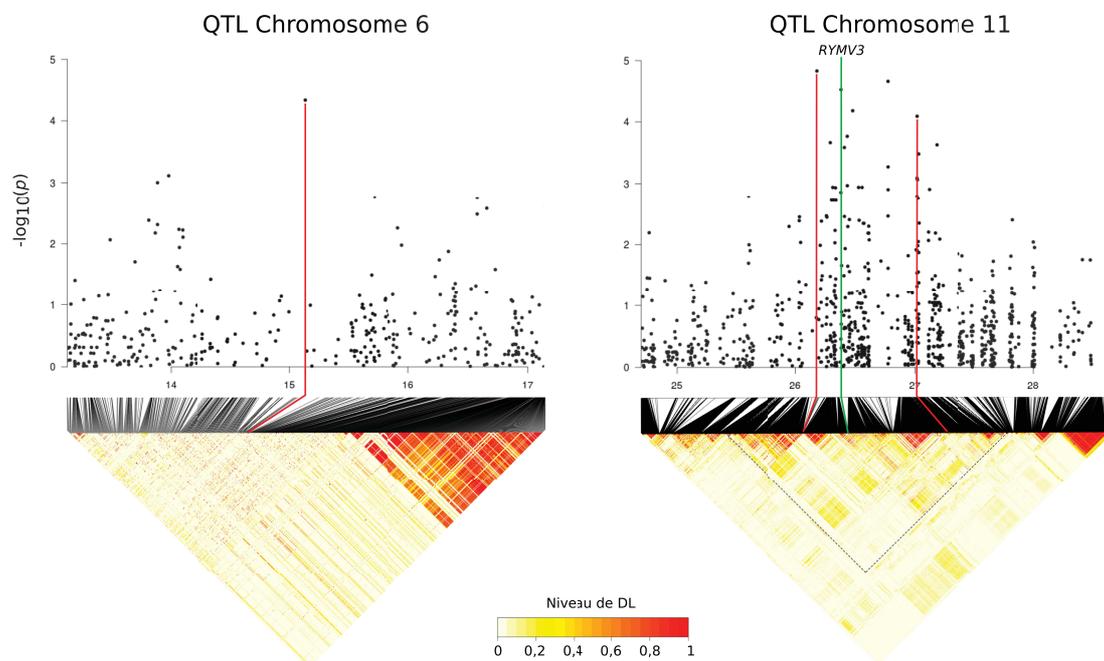


FIGURE 5.14 – DL et *Manhattan plot* autour des QTL identifiés. Le DL autour du SNP du QTL 6 est très faible. En revanche celui autour du QTL₁₁ est plus fort et l'on peut définir une région de 2,1 Mb autour des SNP les plus significatifs (représentée par le triangle pointillé sur le graphique de DL). Le gène *RYMV3* est inclus dans l'intervalle de ce QTL.

Par ailleurs, les SNP significatifs des QTL des chromosomes 6 et 11 présentent un fort DL entre eux : sur les 18 individus présentant l'allèle favorable au chromosome 11 et 17 présentant l'allèle favorable au chromosome 6, seize individus sont en commun.

Nous avons recherché les gènes annotés sur le génome de Nipponbare dans les régions des deux principaux QTL (Tables A.3 et A.4 en annexe). Dans le cas du chromosome 6, nous avons extrait les gènes dans un intervalle de 500 kb, de la position 1 488 000 à la position 15 380 000. L'annotation de cet intervalle contient 19 gènes dont 13 annotés comme "protéines exprimées" et quatre comme "protéines hypothétiques". Les deux autres gènes sont annotés comme codant respectivement pour une enzyme "1,4-alpha-glucane-branching" et pour une "prefoldin", protéine intervenant dans le repliement protéique. L'enzyme "1,4-alpha-glucane-branching" est située à environ 200 kb du SNP significatif et la "prefoldin" à environ 250 kb.

Dans l'intervalle de 2,1 Mb du QTL₁₁, 178 gènes sont annotés dont 87 comme "protéines exprimées" ou "protéines hypothétiques". Parmi les autres, on peut noter 26 gènes appartenant à la famille des NB-LRR, dont *RYMV3*, six codant pour des protéine-kinases, un codant pour une défensine, ou encore un gène contenant un domaine

MATH, similaire à celui du gène de résistance *RTM3* au TEV, LMV et PPV (Cosson et al., 2010a). De nombreux autres gènes de l'intervalle codent pour des protéines pouvant être impliquées dans des mécanismes de défense et sont donc de potentiels candidats.

4 Discussion et perspectives

4.1 Phénotypage de la résistance partielle

Au cours de l'analyse présentée dans ce chapitre, nous avons réalisé une évaluation précise de la résistance partielle au RYMV dans une collection d'accessions *O. glaberrima*. Ce type de résistance avait été précédemment identifié dans l'espèce par Thiémélé (2012), qui avait caractérisé finement dix accessions pour l'expression des symptômes, le contenu en virus, ainsi que l'impact de la maladie sur le développement des plantes. Cette étude avait mis en évidence qu'à l'exception d'une accession particulièrement sensible, les symptômes différenciaient assez mal les accessions *O. glaberrima*. Ce résultat, ainsi que nos propres observations sur des tests préliminaires, nous ont donc amenés à ne pas retenir ce caractère. Chez *O. glaberrima*, les symptômes semblent en effet relativement variables d'une plante à l'autre et s'expriment souvent par un jaunissement généralisé des feuilles, difficile à quantifier. En revanche, dans une population bi-parentale chez *O. sativa*, les symptômes s'expriment par des panachures qui se sont avérées suffisamment fiables et corrélées au contenu en virus pour réaliser une analyse QTL (Albar et al., 1998; Boisnard et al., 2007).

L'étude de Thiémélé (2012) avait suggéré que l'impact de la maladie sur le nombre de feuilles émises pouvait être un critère d'évaluation de la résistance partielle intéressant, tandis que les impacts sur la hauteur des plantes et sur le tallage étaient peu reproductibles et ne constituaient pas des estimateurs de résistance valables. Cependant, l'étude de l'impact de la maladie sur le nombre de feuilles nécessite de disposer d'autant de plantes inoculées que de plantes non inoculées et nos conditions d'expérimentations ne permettaient pas d'envisager ce type de mesures sur une collection de plus d'une centaine d'accessions. Nous avons donc retenu le contenu en virus comme seul critère d'évaluation de la résistance partielle.

Le contenu en virus dans la plante peut être mesuré selon plusieurs méthodes, notamment le test ELISA, et la RT-PCR quantitative. Cette seconde méthode a notamment été utilisée pour comparer l'accumulation virale de mutants de RYMV dans des accessions sensibles ou résistantes (Traoré et al., 2010) mais elle est coûteuse, relativement lourde à mettre en œuvre et aurait nécessité des tests préliminaires importants. Malgré les limites du test ELISA, généralement considéré comme semi-quantitatif uniquement, ce test est bien maîtrisé au laboratoire et avait précédemment permis de détecter des QTL de résistance partielle chez *O. sativa* (Albar, 1998).

Une variabilité intra-accession assez forte a été observée entre les différentes répéti-

tions. Des différences de traitement ou environnementales sont susceptibles d'expliquer cette variabilité. En effet, malgré les précautions prises, la taille de l'expérimentation a généré certaines différences de traitement entre plantes et entre répétitions. Ainsi, les répétitions paires et impaires n'ont pas été inoculées par le même manipulateur. De plus, les inoculations se sont déroulées sur une journée entière alors que la réponse immunitaire des plantes peut être influencée par le rythme circadien (Wang et al., 2011). Par ailleurs, deux mois se sont écoulés entre la première et la dernière expérimentation, la germination et la croissance des plantes avant inoculation n'ont donc pas toujours été réalisées dans les mêmes conditions de température, d'hygrométrie et de durée du jour. Ces paramètres sont également notoirement importants dans les réponses aux stress (Chandra-Shekara et al., 2006; Ouibrahim et al., 2014; Wang et al., 2011; Zhu et al., 2010). Dans le cas particulier de la répétition 4, une attaque d'acariens avant l'inoculation a pu déclencher des mécanismes de défense généraux, rendant les plantes moins sensibles à l'inoculation avec le RYMV et pouvant expliquer le plus faible contenu viral observé.

Lors des échantillonnages, nous avons également remarqué des différences assez nettes de symptômes entre les quatre plantes d'une même répétition pour certaines accessions. Afin de mieux quantifier cette variabilité, des tests ELISA devront être réalisés sur des prélèvements de plantes individuelles.

Malgré ces sources de variabilité, les six répétitions réalisées ont permis d'obtenir une héritabilité au sens large supérieure à 60 % à 17 et 20 jai.

Nous avons effectué des prélèvements à cinq dates après inoculation et estimé le contenu en virus à trois dates différentes, afin (i) d'évaluer sa progression au cours du temps et en fonction des accessions et (ii) de pouvoir choisir l'estimateur le plus héritable possible parmi les trois mesures ou la pAUDPC intégrant l'ensemble des mesures. L'utilisation de dilutions différentes ne permet pas une comparaison directe de l'accumulation virale aux différentes dates. Cependant, l'utilisation de dilutions plus importantes aux dates les plus tardives, de même que la faible réponse à 10 jai et la saturation à 24 jai observée sur les accessions de référence, montrent que le contenu viral a augmenté en moyenne entre toutes les dates de prélèvements. À 13 jai, la forte variabilité entre répétitions suggère que le virus est probablement en phase d'expansion rapide dans la plante. De plus, les mesures réalisées aux trois dates sont bien corrélées entre elles. Cela suggère que le niveau relatif de résistance partielle des accessions change peu entre les dates et que la progression de la maladie dans les plantes suit le même type de profil, quelle que soit l'accession. Il semble donc que la résistance partielle chez *O. glaberrima* corresponde, comme cela a été décrit chez la variété *O. sativa* Azucena (Ioannidou et al., 2003), à un retard dans l'accumulation virale dans la plante mais que le virus atteigne à terme la même concentration dans les accessions partiellement résistantes et les accessions sensibles.

La diversité des phénotypes observés au sein de cette collection confirme donc, qu'en plus de la résistance élevée, il existe des sources de résistance partielle au RYMV dans

l'espèce *O. glaberrima* et la distribution du phénotype suggère que ce caractère est sous un contrôle polygénique.

4.2 Génétique d'association

L'apparition des marqueurs moléculaires dans les années 1990 a permis l'identification et la caractérisation de QTL dans les populations bi-parentales. Cependant, le manque de transférabilité de certains marqueurs (RAPD, AFLP) entre accessions, et la lourdeur des techniques (RFLP) qui limitaient le nombre de marqueurs sur le génome ne permettait pas, à cette époque, la recherche de QTL dans des collections. Ce n'est qu'à la fin des années 2000 que le développement des marqueurs haut-débit, et notamment les marqueurs SNP générés par des approches NGS, a permis la multiplication des études de génétique d'association dans des collections de variétés ou d'accessions (Ingvarsson and Street, 2011). Le séquençage grande profondeur d'une partie de la collection *O. glaberrima* de l'IRD dans le cadre du projet IRIGIN nous a permis d'envisager ce type d'approche pour étudier la résistance partielle dans l'espèce.

Les analyses de génétique d'association font toutefois face à plusieurs facteurs limitants (Ingvarsson and Street, 2011; Korte and Farlow, 2013; Lipka et al., 2015).

Le premier est le nombre de marqueurs disponibles pour génotyper les plantes. En effet, la génétique d'association repose sur le DL entre un marqueur et le polymorphisme causal. Si le nombre de marqueurs est insuffisant, même un QTL à effet majeur pourrait ne pas être identifié. Nos données de séquençage à haute profondeur nous ont permis d'éviter cet écueil. Il nous a au contraire fallu réduire le nombre de marqueurs dont nous disposions en filtrant nos données sur plusieurs critères afin d'accélérer l'analyse et de limiter le nombre de faux positifs. De manière générale, avec la baisse des coûts de séquençage, ce point n'est plus le plus problématique (Ingvarsson and Street, 2011), cédant le pas aux difficultés de phénotypage, évoquées plus haut.

La taille de population est un autre défi du GWAS. Plus les effectifs sont importants, plus le pouvoir de détection de QTL est élevé. Ainsi les effets dus aux allèles rares ne peuvent être détectés dans une petite population. En augmentant les effectifs globaux, ceux des allèles rares augmentent en conséquence, permettant une meilleure détection. Notre taille de population est petite, comparée à certaines études dans l'espèce *O. sativa* pouvant compter jusqu'à un millier d'individus (Huang et al., 2012b).

La structuration de la population étudiée est également une problématique très importante en génétique d'association. Si les sous-populations d'un échantillon ont des phénotypes différents, des SNP liés à la structure peuvent être identifiés par l'analyse GWAS comme associés au caractère (Ingvarsson and Street, 2011; Korte and Farlow, 2013). Des modèles statistiques intégrant la structure de la population (K+Q) comme co-facteur permettent de réduire cet effet mais ils risquent également de créer des faux négatifs (Kang et al., 2008). Nous n'avons détecté dans notre échantillon qu'une faible structuration de population, ne permettant pas une meilleure adaptation des modèles

K+Q à nos données. Cela confirme les résultats d'Orjuela et al. (2013) sur la très faible diversité de l'espèce. Au contraire, ce problème est particulièrement marqué dans le cas des études d'association chez *O. sativa* du fait de la forte structuration entre les sous-espèces *indica* et *japonica*. Pour le contrer, la plupart des auteurs divisent leur échantillon pour réaliser un GWAS sur chacune des sous-espèces (Crowell et al., 2016). La structuration est donc également liée au problème de taille de population.

Une faible taille de population permet toutefois de repérer les QTL ayant des effets forts. Cette notion d'effet du caractère est liée à son héritabilité. Ainsi, plus un caractère est héritable, plus les QTL seront faciles à distinguer dans la population. Dans notre cas, l'héritabilité au sens large des meilleurs caractères est de l'ordre de 60 %. Il ne s'agit pas d'une des plus fortes héritabilités utilisées pour réaliser un GWAS mais elle permet toutefois de repérer les QTL les plus significatifs.

Par ailleurs, les héritabilités au sens strict calculées en sortie de modèle permettent de matérialiser un problème commun des études de génétique d'association : l'héritabilité manquante (Ingvarsson and Street, 2011; Korte and Farlow, 2013). La différence avec l'héritabilité au sens large du caractère repose sur le fait que les études d'association ne permettent pas de capturer l'ensemble des facteurs explicatifs de la variation du caractère. Cela est notamment dû aux facteurs limitants précédemment évoqués mais également à d'autres facteurs, non étudiés, comme les effets épigénétiques (Johannes et al., 2009) ou les interactions entre locus (Korte and Farlow, 2013). Dans notre cas, les héritabilités au sens strict pour les mesures à 13, 17, 20 jai et la pAUDPC sont respectivement de 34,5 %, 31,9 %, 34,3 % et 32,8 %, soit une baisse d'environ 30 % par rapport à l'héritabilité au sens large.

Les limitations de nos données en terme d'effectif et d'héritabilité ont conduit à une baisse de puissance de détection. Ainsi les QTL détectés l'ont été à un seuil très inférieur à celui de Bonferroni. Ce critère est la plupart du temps jugé trop conservatif car il considère que les tests réalisés (dans notre cas un par SNP) sont indépendants, ce qui au vu de l'existence de déséquilibre de liaison, est faux dans le cas d'une étude d'association génétique. L'utilisation de la proportion de *p-values* ajustées par le taux de faux positifs (FDR) peut être une alternative à celle du seuil de Bonferroni (Dolejsi et al., 2014; Yoav Benjamini, 1995). Toutefois cette méthode ne nous a pas non plus permis de détecter des SNP significatifs. Nous avons donc abaissé le seuil de détection à un niveau qui nous paraissait crédible au vu de l'apparence du *Manhattan plot*. Les QTL détectés doivent toutefois être considérés avec prudence.

La puissance de détection a également pu être affectée par des problèmes de concordance génotype/phénotype. Pour éviter ce type d'erreur, nous avons préférentiellement utilisé des graines directement récoltées sur la plante séquencée. Cependant, dans certains cas, nous avons dû utiliser des graines issues d'une multiplication ultérieure, parfois réalisée au champ. Pour deux accessions en particulier, nous avons observé des phénotypes différents (hautement résistants contre sensibles) entre deux lots de graines, suggérant des erreurs ponctuelles sur certains lots. Ces deux accessions ont été retirées

de l'analyse GWAS, mais d'autres ont pu être concernées sans que l'on soit en mesure de les repérer.

La capacité de détection varie également selon la mesure de résistance partielle utilisée. En effet, les différentes variables n'apportent pas tout à fait la même information. Ainsi, à 13 jai, l'imprécision de la mesure explique l'absence de détection de QTL. Les résultats d'association avec la mesure à 17 jai sont les plus significatifs. La résistance partielle au RYMV étant caractérisée par un retard de l'accumulation virale, la mesure à 17 jai doit être celle présentant la meilleure combinaison de répétabilité et de variabilité de la mesure. À 20 jai, une plus grande proportion d'accessions a atteint la saturation en virus et la réponse est donc moins quantitative. Nous nous attendions à ce que la mesure de pAUDPC apporte une information supplémentaire en intégrant les données aux trois dates de prélèvements. Toutefois, la forte corrélation des trois mesures ne la rend pas plus informative : elle est elle-même très corrélée à la mesure à 17 jai et sa prise en compte de la mesure à 13 jai la rend plus imprécise.

Nous avons pu détecter un total de cinq QTL dont la fiabilité est variable. Parmi les QTL identifiés, les QTL des chromosomes 2, 3 et 10 n'ont été détectés qu'avec certaines variables. Toutefois leur détection est plus fréquente en abaissant légèrement le seuil que nous avons défini. S'il ne s'agit pas des QTL les plus fiables il peut toutefois s'agir d'une association réelle peu significative dans notre analyse à cause d'une puissance de détection limitée. Par ailleurs, d'autres SNP présentent un profil intéressant, caractérisé par plusieurs SNP formant un pic, notamment au début du chromosome 11 (9,2 Mb environ) et à la l'extrémité du chromosome 4 (25,2 Mb environ).

Le faible nombre total de QTL identifiés pourrait s'expliquer par un caractère contrôlé par un grand nombre de régions ayant chacune un effet faible sur le phénotype. Bien que *O. glaberrima* ait moins de diversité morphologique qu'*O. sativa*, ce type d'effets pourrait être relié à des caractères morphologiques, comme suggéré dans le cas des QTL de résistance partielle de la variété *O. sativa* Azucena (Albar et al., 1998).

Nous avons comparé la localisation des QTL identifiés avec celles des QTL de résistance partielle au RYMV de la variété *O. sativa* Azucena (Albar et al., 1998; Boisnard et al., 2007). Seul le QTL₂ pourrait colocaliser avec un QTL à effet mineur défini sur une très large portion du chromosome 2 chez Azucena. La très faible colocalisation des QTL que nous avons identifiés dans l'espèce *O. glaberrima* avec ceux de l'espèce *O. sativa* suggère que la résistance partielle au RYMV de l'espèce africaine est contrôlée par des gènes différents de celle de l'espèce asiatique.

Si les cinq régions faiblement associées ne doivent pas être négligées, nous nous sommes toutefois focalisés sur les QTL des chromosomes 6 et 11, qui présentent des SNP significatifs avec une grande partie des variables testées. La région du chromosome 11 en particulier est toujours la plus fortement associée.

Le QTL du chromosome 6 (QTL₆) est proche de deux gènes ayant une fonction annotée. L'un d'eux code pour une enzyme "1,4-alpha-glucane-branching". Un gène homologue de celui-ci avait précédemment été décrit comme étant fortement induit

lors de l'inoculation avec le RYMV de la variété *O. sativa* Azucena (Ventelon-Debout et al., 2003) et colocalise avec un QTL à effet mineur identifié dans la population IR64 x Azucena (Boisnard et al., 2007). Toutefois, la région codante de ce gène n'est pas polymorphe dans la collection. Au niveau protéique, la séquence de la copie du gène colocalisant avec le QTL identifié sur le chromosome 6 n'est pas polymorphe dans la collection d'accessions *O. glaberrima* mais la résistance partielle pourrait provenir d'une différence dans les séquences régulatrices du gène, ce qui serait cohérent avec une induction du gène.

Le QTL₁₁ est situé dans une région où le DL est relativement important. Dans la région définie par le DL, de nombreux gènes sont annotés dont beaucoup ont des fonctions pouvant être impliquées dans des mécanismes de défense. Il s'agit notamment d'une région très riche en gènes de type NB-LRR (Monosi et al., 2004) qui sont impliqués dans des résistances élevées mais peuvent également l'être dans des résistances partielles (Broglie and Butler, 2008; Hayashi et al., 2010). Ce QTL colocalise notamment avec *RYMV3*, suggérant que la résistance partielle et la résistance majeure pourraient être conférées par le même gène. Le phénotype de résistance partielle observé ne semble cependant pas associé à un haplotype particulier à *RYMV3*. Par ailleurs, le pic de significativité sur ce QTL étant large, il pourrait y avoir plusieurs gènes de la région impliqués dans un gain de résistance partielle.

4.3 Conclusions et perspectives

Nous avons détecté deux QTL principaux sur les chromosomes 6 et 11. Ces QTL ne colocalisent pas avec les QTL identifiés précédemment dans l'espèce *O. sativa*, indiquant que le contrôle de la résistance partielle au RYMV dans les deux espèces est différent.

Les QTL identifiés doivent maintenant être validés. Cela peut notamment être fait en complétant l'analyse GWAS par une analyse QTL dans une population bi-parentale entre une accession très sensible et une des accessions présentant les SNP significatifs identifiés. L'utilisation complémentaire de ces deux méthodes est en effet souvent utile pour détecter des gènes de résistance partielle (Brachi et al., 2010; Huang et al., 2012a; Zhao et al., 2007), les avantages des deux approches étant complémentaires. L'approche GWAS permet notamment d'identifier les parents adaptés à l'étude du QTL. En effet, étant donné la faible occurrence de l'allèle favorable au QTL₁₁ dans la collection IRD, il est peu probable que nous ayons identifié l'accession adaptée à son étude sans l'étude GWAS. L'étude QTL pourrait notamment permettre de dissocier ou de confirmer une épistasie entre les QTL 6 et 11. L'étude des polymorphismes des accessions résistantes aux QTL et de leurs effets potentiels pourrait également permettre d'établir une liste de gènes candidats.

Par ailleurs, la résistance partielle est souvent très liée à des interactions environnementales (George and Cavanagh, 2015). Il est donc nécessaire de valider la résistance partielle des accessions identifiées dans d'autres conditions, notamment dans les condi-

tions de culture.

Enfin, du fait de la meilleure annotation du génome de Nipponbare, du pourcentage de *mapping* élevé des données de séquences de la collection IRIGIN sur celui-ci et de la disponibilité des séquences, l'analyse GWAS a été réalisée sur les données de polymorphismes par rapport au génome de Nipponbare. Toutefois, certaines régions sont spécifiques de l'espèce *O. glaberrima* et peuvent contenir des QTL que nous n'aurions pu identifier à cause du choix réalisé. Ainsi, l'utilisation de la séquence de la sous-espèce *Aus* dans une étude GWAS a permis de détecter des QTL non identifiés avec la séquence de Nipponbare (Price et al., 2016). Par ailleurs, le QTL₆ ne tenant qu'à un seul SNP en faible DL avec sa région génomique chez Nipponbare, est en très fort DL avec le QTL₁₁. Cela pourrait être dû à un effet d'épistasie entre les deux QTL, pour la résistance ou un autre caractère, ou bien le SNP du chromosome pourrait avoir été *mappé* au mauvais endroit sur le génome ou correspondre à une région génomique présente sur le chromosome 11 dans l'espèce *O. glaberrima*.

Il serait donc intéressant de refaire l'analyse GWAS sur les polymorphismes par rapport à la séquence de CG14 lorsque ces données seront disponibles.

CHAPITRE

6

Discussion générale et perspectives

Chapitre 6 : Discussion générale et perspectives

Le travail présenté dans cette thèse avait pour objectif de caractériser les bases génétiques de la résistance élevée et de la résistance partielle au RYMV chez les riz africains *O. glaberrima* et *O. barthii*. Nous avons exploité pour cela une collection de 278 accessions *O. glaberrima* et 103 accessions *O. barthii* dont respectivement 163 et 84 ont été séquencées à grande profondeur.

Nous avons utilisé une approche de cartographie génétique pour identifier le gène majeur *RYMV3*, contrôlant la seule résistance dominante au RYMV identifiée à ce jour. Nous avons ensuite analysé la variabilité allélique aux trois gènes de résistance identifiés afin de rechercher de nouveaux allèles de résistance. Enfin, nous avons réalisé une étude du déterminisme de la résistance partielle dans l'espèce *O. glaberrima* par une approche de génétique d'association réalisée sur 125 accessions.

Ce travail a permis l'identification d'un gène majeur, de deux QTL et de nouveaux allèles de résistance qui contribuent à une meilleure compréhension des interactions riz/RYMV et sont utilisables en sélection pour améliorer la durabilité des variétés résistantes.

1 Une grande diversité de résistances au RYMV

1.1 De la diversité des déterminants de résistance

Les bases génétiques de la résistance au RYMV ont été étudiées chez les deux espèces de riz cultivé *O. sativa* et *O. glaberrima*. Trois gènes majeurs y ont été identifiés.

RYMV1 présente quatre allèles de résistance (Albar et al., 2006; Thiémélé et al., 2010) : l'allèle *rymv1-2* a été identifié dans l'espèce *O. sativa* et a été validé fonctionnellement par complémentation d'une lignée résistante avec l'allèle de sensibilité, alors que les allèles *rymv1-3*, *rymv1-4* et *rymv1-5* ont été identifiés dans des accessions *O. glaberrima* et validés par des tests d'allélisme. L'étude de séquence que nous avons menée a permis d'identifier un nouvel allèle qui pourrait être responsable de la résistance d'une accession de l'espèce sauvage *O. barthii*, apparentée à l'espèce *O. glaberrima*.

Les résistances conférées par les gènes *RYMV2* et *RYMV3* ont été identifiées uniquement dans l'espèce *O. glaberrima*. Leur cartographie fine a permis l'identification de candidats (Orjuela et al., 2013; Chapitre 3) mais, à ce jour, ceux-ci n'ont pas été validés fonctionnellement. Toutefois, le rôle de *CPR5-1*, candidat pour *RYMV2*, a été confirmé par la résistance de deux mutants d'insertion T-DNA indépendants dans ce gène (Albar, communication personnelle) et par l'existence de six nouveaux allèles correspondant à une protéine tronquée et associés à la résistance, identifiés dans les espèces *O. glaberrima* et *O. barthii* (Chapitre 4). Si ces allèles doivent encore faire l'objet de validation *via* des tests d'allélisme, l'absence d'un autre déterminisme de

résistance connu dans les accessions qui les portent plaide en faveur de leur implication dans la résistance.

Au gène *RYMV3*, nous avons confirmé l'association d'un SNP non-synonyme dans le domaine LRR d'un NB-LRR avec la résistance dans la collection IRD de riz africains (Chapitres 3 et 4). La validation fonctionnelle du gène est en cours par transformation stable d'une variété *O. sativa* sensible. À ce gène, deux haplotypes différents de celui de Tog5307 sont également portés par des accessions résistantes mais leur implication dans la résistance doit encore être testée dans des descendance avec des accessions sensibles.

La résistance élevée au RYMV dans le riz présente donc une diversité de déterminants importante. L'espèce *O. glaberrima* semble abriter un nombre de résistances bien supérieur à l'espèce *O. sativa* mais, comme discuté dans le chapitre 4, les variétés traditionnelles africaines pourraient receler de nouvelles sources de résistance. Par ailleurs, nous avons aussi souligné l'intérêt d'analyser attentivement la variabilité de gènes candidats dans les données génomiques disponibles pour de nombreuses accessions des riz sauvages et cultivés qui pourraient présenter de nouvelles sources de résistance.

La résistance partielle au RYMV a également été étudiée dans les deux espèces de riz cultivé. Les QTL identifiés ne colocalisent pas (Chapitre 5), suggérant que les déterminismes de résistance partielle des deux espèces sont probablement différents. Toutefois certains d'entre eux colocalisent avec des gènes de résistance majeurs.

Dans l'espèce *O. sativa*, la majorité des *japonica* pluviaux sont partiellement résistants au RYMV. La résistance partielle y a été étudiée dans un croisement bi-parental entre la variété *japonica* Azucena partiellement résistante et la variété *indica* IR64 sensible. Plusieurs QTL ont été identifiés, en particulier trois QTL à effet majeur sur les chromosomes 1, 2 et 12 (Albar et al., 1998; Boissard et al., 2007). Une cartographie fine du QTL₁ a permis de le réduire à un intervalle de 150 kb, colocalisant avec *RYMV2* (Orjuela et al., 2013), suggérant que des allèles différents à ce gène pourraient être impliqués soit dans la résistance partielle, soit dans la résistance élevée. En revanche, le QTL₁₂ n'a pas pu être cartographié précisément car la région recombine très peu dans le croisement utilisé et le déterminant de cette résistance est inconnu à cette date (Boissard, 2007). De la même manière, le QTL₂ est très étendu et aucun candidat ne se détache particulièrement.

L'analyse de la résistance partielle menée sur 125 accessions *O. glaberrima* non hautement résistantes a permis de confirmer la présence de sources de résistance partielle dans l'espèce, comme déjà montrée par Thiémélé et al. (2010). Deux QTL principaux ont été identifiés sur les chromosomes 6 et 11. Le QTL₁₁ est situé dans une région où le DL est relativement important et la région associée à la résistance est d'environ 2 Mb. Il colocalise avec *RYMV3* mais comprend également de nombreux gènes probablement associés à des mécanismes de défense. La colocalisation avec *RYMV3* peut donc n'être qu'une coïncidence. De plus, on ne peut exclure que plusieurs déterminants de résistance partielle soient présents dans cette même région du génome. Le QTL identifié sur le chromosome 6, est moins fortement associé à la résistance partielle et nous avons émis

l'hypothèse (Chapitre 5) que sa détection pourrait n'être qu'un artefact résultant d'un fort DL entre le SNP significatif à ce QTL et les SNP localisés dans la région du QTL du chromosome 11.

Ces deux études de résistance partielle ont été réalisées avec deux stratégies différentes, cartographie dans un croisement bi-parental et génétique d'association dans la diversité naturelle, présentant des avantages et des inconvénients distincts (Zhu et al., 2008 ; Chapitre 1 partie 2). Toutefois ces approches peuvent également être complémentaires. Ainsi, comme discuté dans le chapitre 5, l'utilisation de la cartographie dans des populations bi-parentales peut permettre de préciser la position des QTL identifiés en GWAS.

Par ailleurs, les populations multi-parentales comme les MAGIC ou les NAM tentent de combiner les avantages des deux types de populations mais nécessitent un gros investissement pour leur création (Cavanagh et al., 2008; Fauquet et al., 2005; Huang et al., 2015; Rakshit et al., 2012). Plusieurs populations de ce type ont récemment été développées à partir de variétés *O. sativa* puis génotypées par GBS (Bandillo et al., 2013; Frago et al., 2016). Ces populations constituent des ressources intéressantes qui pourraient permettre d'améliorer la précision de détection des QTL et de valider leurs effets dans différents fonds génétiques ou d'identifier de nouveaux QTL.

1.2 De la diversité des mécanismes de résistance

La grande diversité de déterminants de la résistance au RYMV est assortie d'une diversité de mécanismes, discutée au chapitre 4 et résumée ici. Parmi les trois gènes identifiés, deux sont des gènes de résistance fréquemment décrits dans la résistance aux virus (Boualem et al., 2016; Sanfaçon, 2015) :

RYMV1 code pour un facteur d'initiation de la traduction (Albar et al., 2006) et confère une résistance récessive. Il s'agit d'un facteur de sensibilité de la plante, détourné par le virus pour sa multiplication. Les allèles de résistance correspondent à une rupture de l'interaction entre RYMV et le virus (Hébrard et al., 2010) (Figure 6.1). Par ailleurs, et bien qu'on ne puisse pas exclure que d'autres gènes majeurs soient encore à découvrir dans ce pathosystème, il semble que *RYMV1* soit le seul facteur d'initiation de la traduction impliqué dans la résistance au RYMV (Boisnard, 2007). Ceci est cohérent avec la spécificité d'interaction observée dans d'autres pathosystèmes puisqu'à ce jour, aucun virus n'a été décrit comme interagissant avec plusieurs facteurs d'initiation de la traduction dans une même plante (Sanfaçon, 2015).

RYMV3 code très certainement pour un gène de type CC-NBS-LRR. Il interagirait donc, directement ou non, avec une protéine du virus, déclenchant une cascade de signalisation aboutissant au phénomène de HR ou de résistance extrême (Collier and Moffett, 2009) (Figure 6.2).

Le troisième gène, *RYMV2*, n'a pas d'homologue décrit dans d'autres pathosystèmes, en dehors de mutants de laboratoire d'*A. thaliana*. Les allèles de résistance à *RYMV2*,

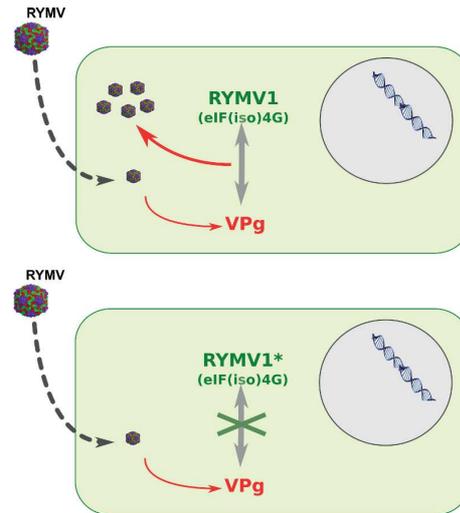


FIGURE 6.1 – Mécanisme probable de la résistance conférée par *RYMV1*. La figure du haut matérialise l'interaction entre la VPg du virus et la protéine eIF(iso)4G1, permettant au virus d'accomplir son cycle infectieux dans la cellule. La figure du bas représente la rupture de cette interaction, menant à la résistance.

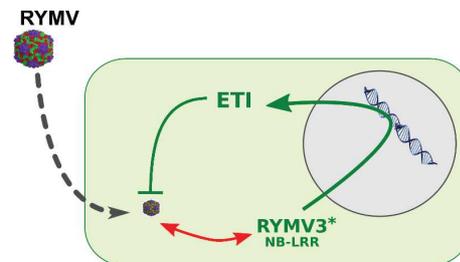


FIGURE 6.2 – Mécanisme probable de la résistance conférée par *RYMV3*. Dans la cellule résistante, le virus est reconnu par RYMV3, conduisant au déclenchement d'une cascade de signalisation conduisant à l'ETI.

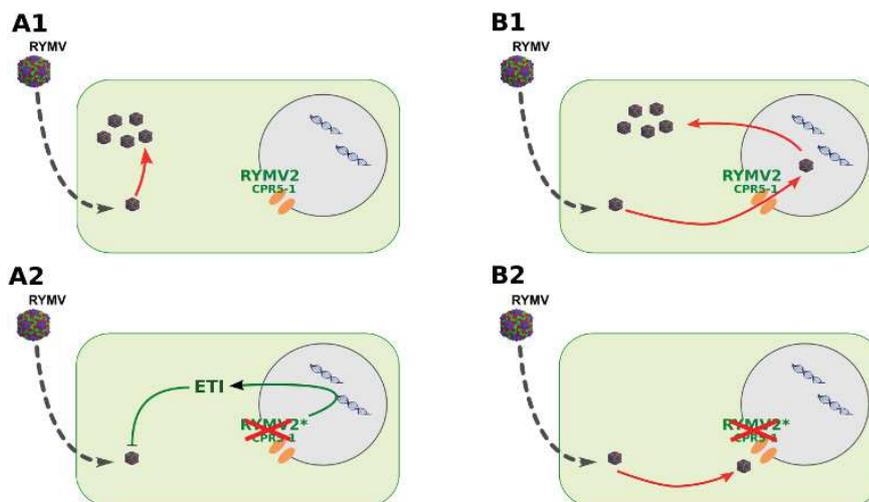


FIGURE 6.3 – Hypothèses de mécanismes de la résistance conférée par *RYMV2*. **A** : Mécanisme impliquant une ETI constitutive (A1 dans une plante sensible, A2 dans une plante avec un allèle de résistance à *RYMV2*). L'allèle nul à *RYMV2* entraîne une résistance constitutive. **B** : Hypothèse d'interaction entre CPR5-1 et une protéine virale. Dans une plante sensible (A1) le virus interagit avec CPR5-1 pour accomplir son cycle d'infection virale. En l'absence d'une protéine CPR5-1 fonctionnelle (A2), l'interaction n'a pas lieu et la plante est résistante.

confirmés ou putatifs, correspondent à des allèles nuls du gène, codant pour une protéine tronquée et conférant une résistance récessive. Les allèles de sensibilité codent pour une protéine homologue à la protéine CPR5 d'*A. thaliana*, impliquée dans le transfert des signaux de résistance au noyau lors de l'établissement de la résistance (Gu et al., 2016). Au vu des mécanismes décrits chez *A. thaliana*, nous avons proposé deux hypothèses de mécanisme de résistance chez *O. glaberrima* (Figure 6.3, Chapitre 4). La première possibilité, conforme au phénomène observé chez *A. thaliana*, est qu'un allèle nul de *CPR5-1* entraînerait une ETI constitutive et en conséquence la résistance. La deuxième hypothèse est que RYMV2 soit, comme RYMV1, un facteur de sensibilité avec lequel le virus doit interagir. Le phénotype des plantes est plutôt en faveur de la deuxième hypothèse mais toutes deux devront être testées en étudiant les changements moléculaires se produisant en l'absence d'une protéine CPR5-1 fonctionnelle. L'interaction directe entre CPR5-1 et les protéines virales pourrait également être testée en système hétérologue bien que la complexité du pore nucléaire rende incertain ce type d'approche. De plus, comme évoqué précédemment, le QTL₁ d'Azucena colocalise avec *RYMV2* et l'allèle de cette variété à *CPR5-1* pourrait donc conférer la résistance partielle. Si le modèle que nous avons proposé pour ce gène est le bon, le QTL pourrait alors impliquer une interaction affaiblie avec le virus ou bien CPR5-1 pourrait être faiblement exprimé et peu disponible pour l'interaction avec le virus. Par ailleurs, le gène *CPR5-2*, homologue de *CPR5-1*, pourrait également jouer un rôle dans le mécanisme de résistance. En effet, des lignées quasi-isogénique d'IR64 ayant l'allèle de Tog7291 à *CPR5-1* ne présentent qu'une résistance partielle, soulignant l'importance du fond génétique. Une étude préliminaire suggère qu'une région de plusieurs mégabases contenant *CPR5-2* pourrait également être impliquée dans la résistance élevée (Albar, communication personnelle). Dans certains fonds génétiques, il est donc possible que CPR5-2 puisse en partie compléter CPR5-1 pour son rôle dans l'infection par le RYMV.

L'absence d'identification d'homologues de *CPR5* comme gènes de résistance dans d'autres pathosystèmes pourrait s'expliquer par le fait qu'il s'agisse d'un mécanisme spécifique du pathosystème riz/RYMV ou bien par le rôle important de la protéine dans la cellule végétale. En effet, la mutation perte de fonction de *CPR5* chez *A. thaliana* est délétère (Bowling et al., 1997). *A. thaliana* comme beaucoup d'autres plantes, ne possède qu'une copie du gène, qui doit donc subir une pression de sélection importante empêchant l'apparition d'allèle nul. En revanche, le riz, comme la plupart des céréales, en possède deux homologues et il est possible qu'une complémentation fonctionnelle de CPR5-1 par CPR5-2 permette l'apparition de la mutation perte de fonction de *CPR5-1* sans baisse de fitness importante. Ainsi, la relative rareté d'une deuxième copie de *CPR5*, pourrait avoir limité la détection de tels gènes de résistance.

D'autres mécanismes de résistance au RYMV peuvent exister, en particulier au niveau des QTL de résistance partielle.

Ainsi, le QTL₆ identifié dans l'espèce *O. glaberrima* colocalise notamment avec un gène annoté comme codant pour une enzyme "1,4-alpha-glucane-branching", impliquée

dans le métabolisme et dont un homologue, colocalisant avec un QTL d'Azucena a été identifié comme fortement surexprimé lors de l'infection par le RYMV (Ventelon-Debout et al., 2003). Ces deux QTL pourraient donc mettre en jeu le même mécanisme tout en utilisant deux gènes homologues différents. Il faut toutefois noter que la précision du QTL d'Azucena est faible et que le QTL₆ d'*O. glaberrima* pourrait n'être qu'un artefact.

Le QTL₁₂ d'Azucena a également un effet fort (Albar et al., 1998; Boissard et al., 2007). Bien que colocalisant avec un gène apparenté aux facteurs d'initiation de la traduction, l'implication de ce dernier n'a pu être confirmée (Boissard, 2007) et le QTL pourrait mettre en jeu un mécanisme original dans le pathosystème. L'intervalle dans lequel il a été cartographié contient notamment un gène "jacalin-like", appartenant à la famille dont font partie les gènes de résistance d'*A. thaliana* aux potyvirus *RTM1* et *JAX1* (Chisholm et al., 2000; Yamaji et al., 2012). Il est à ce jour considéré comme le meilleur gène candidat au QTL.

Par ailleurs, du fait des différences importantes de caractères morphologiques dans le croisement de cartographie *indica* x *japonica*, il avait été suggéré qu'un grand nombre de QTL identifiés dans la variété *O. sativa* Azucena pourrait être associé au développement de la plante (Albar et al., 1998; Boissard et al., 2007). Cela pourrait également être le cas pour une partie des QTL identifiés dans l'espèce *O. glaberrima*. Or, si les caractères phénologiques d'une variété peuvent jouer un rôle dans la résistance partielle, ils ne sont toutefois pas facilement utilisables dans le cadre de la création variétale car ils peuvent représenter des caractères non désirés par les agriculteurs, voire affecter le rendement.

La diversité de gènes et de mécanismes identifiés dans le pathosystème riz/RYMV est très grande et couvre une large gamme des mécanismes de résistance aux virus. Le RYMV est également très bien caractérisé et est facilement manipulable. Ces éléments font donc du pathosystème riz/RYMV un modèle très adapté aux études en laboratoire et l'identification de la diversité des déterminants de résistance partielle et élevée.

1.3 Quelles raisons à la diversité de résistance chez *O. glaberrima* ?

La résistance au RYMV est plus diverse dans l'espèce *O. glaberrima* que celle identifiée dans l'espèce *O. sativa*. Quelles raisons peut-on avancer à cette différence ?

Le centre de diversité du RYMV est en Afrique de l'Est (Fargette et al., 2008b; Pinel-Galzi et al., 2015; Trovão et al., 2015), et son apparition n'est datée que d'environ 200 ans. A partir de là, le virus aurait migré relativement rapidement, vers la fin du XIX^{ème} ou le début du XX^{ème} siècle en Afrique de l'Ouest. La pression de sélection exercée par le RYMV sur les deux espèces de riz cultivé : *O. glaberrima*, domestiqué il y a environ 3500 ans, et *O. sativa*, introduit en Afrique de l'Est dès le X^{ème} siècle et en Afrique de l'Ouest au XVI^{ème} (Murray, 2004; Portères, 1962), ne devrait donc pas être très différente. En outre, une période de 100 à 200 ans paraît trop courte

pour l'évolution d'une si grande diversité de déterminants de résistance dans l'espèce *O. glaberrima*.

Une hypothèse est que la pression ait été exercée par un autre virus, proche du RYMV. En effet, il a été montré dans le cas des potyvirus, qu'un même gène peut contrôler la résistance à plusieurs virus proches. Ainsi, un allèle du facteur eIF4E *pvr2* chez le piment contrôle une résistance au PVY et au TEV (Charron et al., 2008) et la résistance conférée par les gènes *RTM* contrôle la résistance d'au moins trois potyvirus chez *A. thaliana* (Decroocq et al., 2006; Mahajan et al., 1998; Revers et al., 2003). Plusieurs Sobemovirus ont été identifiés en Afrique (Séréme et al., 2008; Thottappilly et al., 1992), preuve de la diversité de cette famille sur le continent. Un ou plusieurs des virus identifiés ou un autre Sobemovirus, aujourd'hui éteint ou ayant une faible incidence, pourraient ainsi avoir fait évoluer les différents allèles de résistance au RYMV dans les riz africains.

Cette hypothèse pourrait également expliquer la structuration géographique observée entre les allèles de résistance à *RYMV1* et à *RYMV2* : deux virus différents, présents de part et d'autre de l'axe séparant les allèles aux deux gènes, auraient pu imprimer des pressions de sélection différentes sur des facteurs hôtes nécessaires à leur cycle.

2 Vers une résistance durable au RYMV au champ ?

Des variétés comportant l'allèle de résistance *rymv1-2* sont en cours d'homologation (Bouet et al., 2013; Ndjiondjop et al., 2013). Toutefois, du fait de l'existence de contournement de la résistance, l'usage d'un seul gène de résistance n'est pas suffisant pour assurer une résistance durable et d'autres gènes doivent être exploités. Les autres déterminants de résistance contre le RYMV identifiés soit l'ont été dans l'espèce *O. glaberrima* (Orjuela et al. (2013); chapitre 3), qui a de faibles rendements et n'est plus cultivée que localement, soit sont des QTL (Albar et al., 1998; Boissard et al., 2007; Chapitre 5). Ces résistances sont-elles exploitables en sélection ? Et quelles stratégies déployer afin d'obtenir une résistance durable ?

2.1 Exploitation des résistances de l'espèce *O. glaberrima*

Malgré son adaptation aux stress biotiques et abiotiques en Afrique, l'espèce *O. glaberrima* est actuellement très peu cultivée à cause notamment de ses faibles rendements. L'utilisation des locus d'intérêt identifiés chez cette l'espèce passe donc par leur transfert dans l'espèce *O. sativa*. La combinaison des avantages des deux espèces est notamment l'objectif des lignées hybrides NERICA (*NEw RICes for Africa*) développées dans les années 90 (Jones et al., 1997; Ndjiondjop et al., 2008; Semagn et al., 2007). Toutefois, la barrière de stérilité entre les deux espèces est très importante, notamment en raison du locus *S1*, gène majeur de stérilité interspécifique qui subit une forte distorsion de ségrégation vers l'allèle *O. glaberrima* dans les descendance *O. sativa* x *O. glaberrima*

(Garavito et al., 2010; Lorieux et al., 2000). Le croisement initial entre les deux espèces est donc difficile et, afin de restaurer la fertilité et donc le rendement, plusieurs cycles de rétrocroisements et d'autofécondations sont nécessaires. La présence de l'allèle d'*O. glaberrima* au gène *S1* peut néanmoins se maintenir malgré un nombre important de rétrocroisements et être source de stérilité lors de nouveaux croisements avec une variété *O. sativa*.

Malgré ces difficultés, les allèles de résistance *rymv1-3*, *rymv2-2* et l'allèle de Tog5307 au gène *RYMV3* ont été introgressés dans la variété *O. sativa* spp *indica* IR64. Les lignées obtenues, présentant un génotype et un phénotype essentiellement de type IR64, peuvent être introduites dans des programmes de sélection, bien qu'elles portent toujours l'allèle *O. glaberrima* au gène *S1*. Par ailleurs, des lignées quasi isogéniques de IR64 portant deux gènes de résistance sont également en cours de développement. Cumulant *rymv2-2* et *rymv1-2* ou *rymv1-3*, elles devraient permettre d'obtenir une résistance plus durable.

Toutefois, comme il a été mentionné précédemment, les lignées introgressées avec l'allèle *rymv2-2* ne présentent qu'une résistance partielle. Ce résultat suggère un effet important du fond génétique sur la résistance, comme relevé dans plusieurs autres pathosystèmes (Banerjee et al., 2001; Sun et al., 2004) et matérialise la nécessité de mieux comprendre le mécanisme de résistance afin de pouvoir l'exploiter plus efficacement. Cette résistance partielle peut toutefois se révéler efficace au champ, où l'infection virale peut être moins importante que celle réalisée dans nos conditions expérimentales.

Outre les problèmes de stérilité interspécifique, certains auteurs ont évoqué la possibilité que des variétés issues de croisements éloignés soient plus sujettes à l'émergence de lignées présentant un syndrome de dédomestication (Qiu et al., 2014). Ce phénomène pourrait expliquer l'instabilité qui a ponctuellement été observée chez certaines lignées NERICA. Afin de contourner les problèmes rencontrés par les croisements interspécifiques, il est également possible d'utiliser le génie génétique afin de modifier directement les variétés d'intérêt. Les nouvelles approches d'éditions de génome telles que CRISPR-Cas9 ou les TALEN peuvent ainsi être des moyens efficaces et rapides pour cela (Wei et al., 2013), notamment dans le cadre des résistances aux pathogènes comme cela a été fait pour améliorer la résistance du riz contre *Xanthomonas oryzae* (Blanvillain-Baufumé et al., 2016). Si la génération d'allèles spécifiques à *RYMV1* ou *RYMV3* est sans doute difficile, l'inactivation du gène *CPR5-1* chez *O. sativa* pourrait être facilement envisagée par une telle approche. Toutefois, le génie génétique fait face à une méfiance sociétale rendant incertaine l'adoption de variétés ainsi modifiées.

2.2 Exploitation des QTL de résistance

Bien que la résistance conférée par les QTL soit moins efficace que les résistances majeures, elle peut toutefois permettre d'obtenir une résistance efficace au champ en

diminuant la pression virale (Kou and Wang, 2010). L'association d'un gène majeur et d'une résistance partielle a également pour effet d'augmenter la durabilité de la première (Quenouille et al., 2013, 2015). Bien que trois gènes de résistance majeurs au RYMV aient été identifiés, l'utilisation des QTL pourrait ainsi présenter un intérêt en sélection. Les QTL sont cependant des caractères hautement quantitatifs et leur effet est souvent influencé par l'environnement. Les principaux QTL de résistance partielle identifiés chez *O. sativa* ont été validés au champ (Albar et al., 1998), mais ce n'est cependant pas le cas de ceux mis en évidence chez *O. glaberrima*. Les conditions d'évaluation en serre étant relativement éloignées de celle du champ, leur efficacité en condition de culture doit être vérifiée.

Par ailleurs, les colocalisations entre gènes majeurs et QTL sont susceptibles de limiter l'intérêt de ces derniers. Par exemple, si, comme supposé, le QTL₁ identifié chez Azucena correspond à un allèle de *CPR5-1*, il est probable que l'allèle conférant la résistance élevée aura plus d'intérêt au champ. Par ailleurs, même si les deux types de résistance n'étaient pas conférés par le même gène, comme cela est possible sur le chromosome 11, le lien physique entre les deux gènes pourrait être difficile à rompre pour cumuler leurs effets. De plus, certains QTL identifiés pourraient être liés à des caractères non souhaités dans les variétés améliorées, notamment des caractères morphologiques, comme cela a été envisagé pour certains QTL identifiés chez Azucena. L'arbitrage entre résistance et morphologie n'étant pas forcément en faveur de la résistance, leur intérêt en sélection serait donc faible.

Quel que soit leur intérêt direct en sélection, la caractérisation des QTL permettra cependant une meilleure compréhension des interactions en jeu dans le pathosystème, pouvant ainsi permettre le développement et la meilleure exploitation des résistances.

2.3 Le contournement des résistances : quelle stratégie employer pour une résistance durable ?

De nombreux gènes de résistance contre des agents pathogènes déployés au champ ont été contournés en quelques années. On parle de *boom-and-bust* des gènes de résistance (Mundt, 2014). L'apparition d'un variant virulent de l'agent pathogène et son augmentation en fréquence dans la population pathogène peut ainsi rendre un gène de résistance inutile en quelques années.

L'une des méthodes les plus fréquemment envisagées pour ralentir le contournement est l'association, ou *pyramiding* de gènes de résistance. Dans le cas de l'interaction riz/RYMV, les trois gènes de résistance identifiés mettent en jeu des mécanismes de contournement distincts et leur association dans un même fond génétique est supposée limiter considérablement les possibilités d'émergence d'un variant contournant. L'accession Tog5672, qui possède à la fois l'allèle *rymv1-4* et l'allèle de résistance à RYMV3, est par exemple contournée très difficilement (Hébrard et Albar, communication personnelle).

Une meilleure compréhension des mécanismes de contournement peut également permettre de mieux gérer le déploiement des gènes de résistance et d'optimiser leur durabilité. Ainsi, des phénomènes de contournement ont été observés pour les trois gènes de résistance et les variants virulents sur différentes accessions résistantes ont été séquencés, mettant en évidence des mutations responsables du contournement pour chacun des gènes.

Dans le cas des allèles *RYMV1*, les mutations de contournement sont situées dans la VPg et diffèrent en fonction des allèles de résistance (Pinel-Galzi et al., 2007; Traoré et al., 2010). Les principales mutations responsables du contournement affectent les acides aminés 41, 48 et 52 de la VPg. Ces mutations permettent le rétablissement de l'interaction entre la VPg et le facteur eIF(iso)4G1 et l'accomplissement du cycle du virus (Hébrard et al., 2010). D'autres mutations dans la région de chevauchement des ORF2a et ORF2b ont également été détectées comme permettant le contournement de l'allèle *rymv1-2* (Poulicard et al., 2014) mais leur rôle dans l'interaction plante/virus est inconnue.

Le contournement de la résistance de Tog7291, conférée par *RYMV2*, implique des mutations dans le domaine d'ancrage à la membrane de la polyprotéine 2a (Pinel-Galzi et al., sub). Ce domaine est supposé permettre l'ancrage de la protéine dans les membranes cellulaires et serait impliqué dans la synthèse du brin négatif du virus (Sõmera et al., 2015). Si *RYMV2* est un facteur de sensibilité de la plante, il est possible que ce domaine interagisse avec la protéine CPR5-1 dans les plantes sensibles. Dans cette hypothèse, en présence d'une protéine CPR5-1 tronquée et non fonctionnelle, les mutations de contournement pourraient alors permettre l'établissement d'une interaction avec CPR5-2.

Enfin, les mutations de contournement de la résistance de l'accession Tog5307 sont situées dans la protéine de capsid, suggérant que celle-ci serait la protéine d'avirulence reconnue dans le mécanisme de la résistance conférée par *RYMV3* (Albar, communication personnelle). Les protéines de capsid des virus sont des protéines d'avirulence communes dans le cas des résistances aux virus conférées par des gènes de types NB-LRR (Boualem et al., 2016), ce qui est donc cohérent avec le gène candidat identifié.

Comme cela a été décrit dans de nombreuses interactions plantes/virus (García-Arenal and Fraile, 2013), il semble que l'acquisition des mutations impliquées dans le contournement entraînerait une baisse de fitness des variants virulents, quel que soit le gène de résistance considéré (Albar, communication personnelle). Ainsi, les variants virulents sur les accessions résistantes Gigante et Tog7291 sont contre-sélectionnés au profit des isolats sauvages chez la variété sensible IR64 (Pinel-Galzi et al., sub; Poulicard et al., 2010). Un phénomène similaire est également suspecté pour les variants contournant Tog5307 (Albar, communication personnelle). Cette baisse de fitness pourrait limiter la fixation des mutations de contournement au champ et permettre ainsi une meilleure durabilité des gènes de résistance lors de leur déploiement (Sorel et al., 2014).

Par ailleurs, la capacité du RYMV à acquérir les mutations de contournement et la fréquence de contournement observée est très dépendante des isolats viraux. Le polymorphisme acide glutamique (E) / thréonine (T) en position 49 de la VPg du RYMV a ainsi été identifié comme un déterminant majeur de la capacité d'un isolat à contourner la résistance de Gigante (*rymv1-2*) et celle de Tog5681 (*rymv1-3*). (Pinel-Galzi et al., 2007; Traoré et al., 2010). Les isolats T49 contournent fréquemment Tog5681 et rarement Gigante, tandis que des résultats opposés sont observés avec les isolats E49. De même, les résistances des accessions *O. glaberrima* Tog7291 (*RYMV2*) et Tog5307 (*RYMV3*) sont plus fréquemment contournées par les isolats T49 que les isolats E49 (Pinel-Galzi et al., soumis; Chapitre 3). Les isolats T49 ont également une meilleure fitness que les isolats E49 dans les accessions *O. glaberrima* sensibles (Poulicard et al., 2012). L'ensemble de ces éléments suggère donc une meilleure adaptation des isolats T49 à l'espèce *O. glaberrima*.

Cette variabilité dans la capacité de contournement permet d'envisager un déploiement géographique raisonné des différents gènes de résistance en fonction des isolats présents. En effet, les isolats E49 et T49 ont une répartition géographique différente : les isolats E49 sont présents uniquement en Afrique de l'Est, alors que les deux types d'isolats sont présents en Afrique de l'Ouest. En Afrique de l'Est, les accessions *O. glaberrima* résistantes sont donc supposées avoir une résistance plus durable qu'une accession *O. sativa* porteuse de l'allèle de résistance *rymv1-2*. Cependant, l'introgression des gènes de résistance d'*O. glaberrima* dans un fond génétique *O. sativa*, auxquels les isolats E49 sont plus adaptés, pourrait impacter la durabilité des gènes. Il est donc essentiel de comprendre le déterminisme génétique de l'adaptation entre isolats viraux et espèces de riz pour optimiser la durabilité des résistances.

Dans ce contexte, il est intéressant de noter que toutes les accessions *O. sativa* étudiées jusqu'à maintenant ont une alanine (A) en position 303 de la protéine eIF(iso)4G1, dans le domaine MIF4G, tandis que les accessions *O. glaberrima* sont porteuses d'un acide aspartique (D). Or le domaine MIF4G interagit avec la VPg du RYMV (Hébrard et al., 2010). L'adaptation des isolats T49 à l'espèce *O. glaberrima* pourrait donc provenir d'une meilleure interaction avec un domaine MIF4G porteur d'un acide aspartique en position 303. Au cours de cette étude, nous avons identifié trois accessions *O. barthii* porteuses d'une alanine en position 303. Une étude de la compétition entre isolats T49 et E49 sur ces accessions, comparée à celle qui s'opère dans d'autres accessions *O. barthii*, pourrait donc permettre de valider ou d'infirmer cette hypothèse.

3 Conclusion générale

Le travail présenté ici a permis l'identification de nouveaux déterminants de résistance dans les espèces *O. glaberrima* et *O. barthii*. À l'issue de cette étude, nous avons probablement élucidé la totalité des déterminants de résistance élevée au RYMV présents dans la collection IRD d'accessions *O. glaberrima*. Par ailleurs, l'étude de la distribution

phylogénétique et géographique des déterminants de résistance contribue à une meilleure compréhension des phénomènes évolutifs ayant mené à l'établissement de leur grande diversité. Les nouveaux déterminants identifiés doivent encore être validés et leur exploitation en sélection pose plusieurs problèmes techniques qui devront être résolus pour permettre leur déploiement. Toutefois ces gènes et allèles constituent des ressources de choix pour la sélection de variétés résistantes au RYMV et leur association et l'optimisation de leur déploiement devraient pouvoir conférer une résistance durable à la panachure jaune.

Ce travail est une nouvelle confirmation que le modèle riz/RYMV est un modèle particulièrement intéressant du fait des nombreuses ressources génétiques et génomiques chez le riz et de la manipulation facile du virus. Les études, menées à l'IRD, en collaboration avec différents partenaires, simultanément sur le virus et sur ses hôtes permettent de progresser à la fois sur l'identification des mécanismes moléculaires de leurs interactions, la diversité des facteurs impliqués et les contraintes évolutives façonnant cette diversité et constituent un pas de plus vers une meilleure compréhension globale des interactions plante/virus.

Références bibliographiques

- Abo, M. E., Alegbejo, M. D., Sy, A. A., and Misari, S. M. (2000). An Overview of the Mode of Transmission, Host Plants and Methods of Detection of *Rice yellow mottle virus*. *Journal of Sustainable Agriculture*, 17 :19–36.
- Abubakar, Z. (2003). Phylogeography of *Rice yellow mottle virus* in Africa. *Journal of General Virology*, 84 :733–743.
- Albar, L. (1998). *Etude des bases génétiques de la résistance partielle du riz au virus de la panachure jaune du riz (RYMV). Cartographie moléculaire et relation avec la diversité génétique*. PhD thesis, Thèse de l'Université Paris XI.
- Albar, L., Bangratz-Reyser, M., Hébrard, E., Ndjiondjop, M. N., Jones, M., and Ghesquière, A. (2006). Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to *Rice yellow mottle virus*. *Plant Journal*, 47 :417–426.
- Albar, L., Lorieux, M., Ahmadi, N., Rimbault, I., Pinel, A., Sy, A. A., Fargette, D., and Ghesquière, A. (1998). Genetic basis and mapping of the resistance to *Rice yellow mottle virus*. I. QTLs identification and relationship between resistance and plant morphology. *TAG Theoretical and Applied Genetics*, 97 :1145–1154.
- Alexandrov, N., Tai, S., Wang, W., Mansueto, L., Palis, K., Fuentes, R. R., Ulat, V. J., Chebotarov, D., Zhang, G., Li, Z., Mauleon, R., Hamilton, R. S., and McNally, K. L. (2015). SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Research*, 43(D1) :D1023–D1027.
- Allarangaye, M. D., Traoré, O., Traoré, E. V. S., Millogo, R. J., and Konaté, G. (2006). Evidence of non-transmission of *Rice yellow mottle virus* through seeds of wild host species. *Journal of Plant Pathology*, 88(3) :309–315.
- Ashikari, M. and Matsuoka, M. (2006). Identification, isolation and pyramiding of quantitative trait loci for rice breeding. *Trends in Plant Science*, 11(7) :344–350.
- Awoderu, V. (1991). The *Rice yellow mottle virus* situation in west africa. *Journal of basic microbiology*, 31(2) :91–99.
- Bakker, W. (1974). *Characterization and ecological aspects of Rice yellow mottle virus in Kenya*.

- Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature reviews. Genetics*, 7 :781–91.
- Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., Tung, C.-W., McCouch, S., Thomson, M., Mauleon, R., et al. (2013). Multi-parent advanced generation inter-cross (magic) populations in rice : progress and potential for genetics research and breeding. *Rice*, 6(1) :1.
- Banerjee, D., Zhang, X., and Bent, A. F. (2001). The leucine-rich repeat domain can determine effective interaction between rps2 and other host factors in arabidopsis rps2-mediated disease resistance. *Genetics*, 158(1) :439–450.
- Bendahmane, a., Kanyuka, K., and Baulcombe, D. C. (1999). The Rx gene from potato controls separate virus resistance and cell death responses. *The Plant cell*, 11(May) :781–792.
- Bhullar, N. K., Zhang, Z., Wicker, T., and Keller, B. (2010). Wheat gene bank accessions as a source of new alleles of the powdery mildew resistance gene Pm3 : a large scale allele mining project. *BMC Plant Biology*, 10(1) :88.
- Blanvillain-Baufumé, S., Reschke, M., Solé, M., Auguy, F., Doucoure, H., Szurek, B., Meynard, D., Portefaix, M., Cunnac, S., Guiderdoni, E., et al. (2016). Targeted promoter editing for rice resistance to xanthomonas oryzae pv. oryzae reveals differential activities for sweet14-inducing tal effectors. *Plant Biotechnology Journal*.
- Boch, J., Verbsky, M. L., Robertson, T. L., Larkin, J. C., and Kunkel, B. N. (1998). Analysis of resistance gene-mediated defense responses in arabidopsis thaliana plants carrying a mutation in cpr5. *Molecular Plant-Microbe Interactions*, 11(12) :1196–1206.
- Boisnard, A. (2007). *Caractérisation des QTL de résistance au Rice yellow mottle virus chez le riz et relation avec les facteurs d'initiation de la traduction de type 4E et 4G*. PhD thesis, Montpellier 2.
- Boisnard, A., Albar, L., Thiémélé, D. E., Rondeau, M., and Ghesquière, A. (2007). Evaluation of genes from eIF4E and eIF4G multigenic families as potential candidates for partial resistance QTLs to *Rice yellow mottle virus* in rice. *TAG. Theoretical and applied genetics.*, 116 :53–62.
- Bokros, C. L., Hugdahl, J. D., Kim, H. H., Hanesworth, V. R., van Heerden, A., Browning, K. S., and Morejohn, L. C. (1995). Function of the p86 subunit of eukaryotic initiation factor (iso)4F as a microtubule-associated protein in plant cells. *Proceedings of the National Academy of Sciences*, 92 :7120–7124.
- Bonneau, C., Brugidou, C., Chen, L., Beachy, R. N., and Fauquet, C. (1998). Expression of the *Rice yellow mottle virus* P1 protein in vitro and in vivo and its involvement in virus spread. *Virology*, 244 :79–86.
- Boualem, A., Dogimont, C., and Bendahmane, A. (2016). The battle for survival between viruses and their host plants. *Current Opinion in Virology*, 17 :32–38.
- Bouet, A., Amancho, A. N., Kouassi, N., and Anguete, K. (2013). Comportement de nouvelles lignées isogéniques de riz irrigué dotées du gène de résistance (rymv1) au rymv en afrique de l'ouest : situation en côte d'ivoire. *International Journal of Biological and Chemical Sciences*, 7(3) :1221–1233.
- Bougerol, B. and Pham, J.-L. (1989). Influence of the *Oryza sativa* genotype on the fertility and quantitative traits of F 1 hybrids between the two cultivated rice species *O . sativa* L. and *O . glaberrima* Steud. *Genome*, 32 :810–815.

- Bowling, S. a., Clarke, J. D., Liu, Y., Klessig, D. F., and Dong, X. (1997). The *cpr5* mutant of *Arabidopsis* expresses both NPR1-dependent and NPR1-independent resistance. *The Plant cell*, 9 :1573–84.
- Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J., and Roux, F. (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet*, 6(5) :e1000940.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). Tassel : software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19) :2633–2635.
- Brininstool, G., Kasili, R., Simmons, L. A., Kirik, V., Hülkamp, M., and Larkin, J. C. (2008). Constitutive Expressor Of Pathogenesis-related Genes5 affects cell wall biogenesis and trichome development. *BMC plant biology*, 8 :58.
- Brogie, K. E. and Butler, K. H. (2008). Polynucleotides and methods for making plants resistant to fungal pathogens.
- Bryan, G. T., Wu, K. S., Farrall, L., Jia, Y., Hershey, H. P., McAdams, S. A., Faulk, K. N., Donaldson, G. K., Tarchini, R., and Valent, B. (2000). A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. *The Plant cell*, 12(11) :2033–46.
- Burguán, J. and Havelda, Z. (2011). Viral suppressors of RNA silencing. *Trends in Plant Science*, 16 :265–272.
- Bush, M. S., Hutchins, A. P., Jones, A. M., Naldrett, M. J., Jarmolowski, A., Lloyd, C. W., and Doonan, J. H. (2009). Selective recruitment of proteins to 5' cap complexes during the growth cycle in *Arabidopsis*. *The Plant Journal*, 59 :400–412.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). Blast+ : architecture and applications. *BMC bioinformatics*, 10(1) :1.
- Caplan, J. L., Mamillapalli, P., Burch-Smith, T. M., Czymbek, K., and Dinesh-Kumar, S. P. (2008). Chloroplastic Protein NRIP1 Mediates Innate Immune Receptor Recognition of a Viral Effector. *Cell*, 132 :449–462.
- Carr, J. P., Lewsey, M. G., and Palukaitis, P. (2010). *Signaling in induced resistance.*, volume 76. Elsevier Inc., 1 edition.
- Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From mutations to MAGIC : resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology*, 11 :215–221.
- Chandra-Shekara, A., Gupte, M., Navarre, D., Raina, S., Raina, R., Klessig, D., and Kachroo, P. (2006). Light-dependent hypersensitive response and resistance signaling against turnip crinkle virus in *Arabidopsis*. *The Plant Journal*, 45(3) :320–334.
- Charron, C., Nicolai, M., Gallois, J.-L., Robaglia, C., Moury, B., Palloix, A., and Caranta, C. (2008). Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. *The Plant journal : for cell and molecular biology*, 54 :56–68.
- Chisholm, S. T., Mahajan, S. K., Whitham, S. A., Yamamoto, M. L., and Carrington, J. C. (2000). Cloning of the *Arabidopsis* RTM1 gene, which controls restriction of long-distance movement of tobacco etch virus. *Proceedings of the National Academy of Sciences*, 97 :489–494.

- Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, 138(3) :963–971.
- Civáň, P., Craig, H., Cox, C. J., and Brown, T. A. (2015). Three geographically separate domestications of Asian rice. *Nature Plants*, 1(11) :1–5.
- Collier, S. M. and Moffett, P. (2009). NB-LRRs work a "bait and switch" on pathogens. *Trends in Plant Science*, 14(August) :521–529.
- Collmer, C. W., Marston, M. F., Taylor, J. C., and Jahn, M. (2000). The I gene of bean : a dosage-dependent allele conferring extreme resistance, hypersensitive resistance, or spreading vascular necrosis in response to the potyvirus Bean common mosaic virus. *Molecular plant-microbe interactions : MPMI*, 13(11) :1266–1270.
- Cosson, P., Schurdi-Levraud, V., Le, Q. H., Sicard, O., Caballero, M., Roux, F., Le Gall, O., Candresse, T., and Revers, F. (2012). The RTM resistance to potyviruses in *Arabidopsis thaliana* : natural variation of the RTM genes and evidence for the implication of additional genes. *PLoS one*, 7 :e39169.
- Cosson, P., Sofer, L., Hien Le, Q., Leger, V., Schurdi-Levraud, V., Whitham, S. a., Yamamoto, M. L., Gopalan, S., Le Gall, O., Candresse, T., Carrington, J. C., and Revers, F. (2010a). RTM3, Which Controls Long-Distance Movement of Potyviruses, Is a Member of a New Plant Gene Family Encoding a Meprin and TRAF Homology Domain-Containing Protein. *Plant Physiology*, 154(September) :222–232.
- Cosson, P., Sofer, L., Schurdi-Levraud, V., and Revers, F. (2010b). A member of a new plant gene family encoding a meprin and TRAF homology (MATH) domain-containing protein is involved in restriction of long distance movement of plant viruses. *Plant signaling & behavior*, 5(October) :1321–1323.
- Cronn, R., Knaus, B. J., Liston, A., Maughan, P. J., Parks, M., Syring, J. V., and Udall, J. (2012). Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany*, 99(2) :291–311.
- Crowell, S., Korniliev, P., Falcão, A., Ismail, A., Gregorio, G., Mezey, J., and McCouch, S. (2016). Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. *Nature Communications*, 7 :10527.
- Dawkins, R. and Krebs, J. R. (1979). Arms Races between and within Species. *Proceedings of the Royal Society B : Biological Sciences*, 205 :489–511.
- de Ronde, D., Butterbach, P., and Kormelink, R. (2014). Dominant resistance against plant viruses. *Frontiers in plant science*, 5 :307.
- Decroocq, V., Sicard, O., Alamillo, J., Lansac, M., Eyquard, J., Garcia, J., Candresse, T., Le Gall, O., and Revers, F. (2006). Multiple resistance traits control plum pox virus infection in *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions*, 19(5) :541–549.
- Djian-Caporalino, C., Palloix, A., Fazari, A., Marteu, N., Barbary, A., Abad, P., Sage-Palloix, A.-M., 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) :1.
- Dogimont, C., Chovelon, V., Pauquet, J., Boualem, A., and Bendahmane, A. (2014). The Vat locus encodes for a CC-NBS-LRR protein that confers resistance to *Aphis gossypii* infestation and *A. gossypii*-mediated virus resistance. *The Plant journal : for cell and molecular biology*, pages 1–12.

- Dolejsi, E., Bodenstorfer, B., and Frommlet, F. (2014). Analyzing genome-wide association studies with an fdr controlling modification of the bayesian information criterion. *PloS one*, 9(7) :e103322.
- Doc, G., Perin, C., Fromentin, S., and Larmande, P. (2009). Orygenesdb 2008 update : database interoperability for functional genomics of rice. *Nucleic acids research*, 37(suppl 1) :D992–D995.
- Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic acids research*, 19(6) :1349.
- Eitas, T. K. and Dangl, J. L. (2010). NB-LRR proteins : Pairs, pieces, perception, partners, and pathways. *Current Opinion in Plant Biology*, 13(4) :472–477.
- Ellis, J. G. (1999). Identification of Regions in Alleles of the Flax Rust Resistance Gene L That Determine Differences in Gene-for-Gene Specificity. *THE PLANT CELL ONLINE*, 11 :495–506.
- Fargette, D., Konaté, G., Fauquet, C., Muller, E., Peterschmitt, M., and Thresh, J. M. (2006). Molecular ecology and emergence of tropical plant viruses. *Annual review of phytopathology*, 44 :235–60.
- Fargette, D., Pinel, A., Abubakar, Z., Traoré, O., Brugidou, C., Fatogoma, S., Hébrard, E., Choisy, M., Séré, Y., Fauquet, C., and Konaté, G. (2004). Inferring the evolutionary history of *Rice yellow mottle virus* from genomic, phylogenetic, and phylogeographic studies. *Journal of virology*, 78 :3252–61.
- Fargette, D., Pinel, A., Halimi, H., Brugidou, C., Fauquet, C., and Van, R. M. (2002a). Comparison of molecular and immunological typing of isolates of *Rice yellow mottle virus*. *Archives of virology*, 147 :583–96.
- Fargette, D., Pinel, A., Rakotomalala, M., Sangu, E., Traoré, O., Sérémé, D., Sorho, F., Issaka, S., Hébrard, E., Séré, Y., Kanyeka, Z., and Konaté, G. (2008a). *Rice yellow mottle virus*, an RNA plant virus, evolves as rapidly as most RNA animal viruses. *Journal of virology*, 82 :3584–9.
- Fargette, D., Pinel, A., and Traoré, O. (2002b). Emergence of resistance-breaking isolates of *Rice yellow mottle virus* during serial inoculations. *European journal of . . .*, 108 :585–591.
- Fargette, D., Pinel-Galzi, A., Sérémé, D., Lacombe, S., Hébrard, E., Traoré, O., and Konaté, G. (2008b). Diversification of *Rice yellow mottle virus* and related viruses spans the history of agriculture from the neolithic to the present. *PLoS pathogens*, 4 :e1000125.
- Farnham, G. and Baulcombe, D. C. (2006). Artificial evolution extends the spectrum of viruses that are targeted by a disease-resistance gene from potato. *Proceedings of the National Academy of Sciences of the United States of America*, 103 :18828–18833.
- Fauquet, C., Mayo, M., Maniloff, J., Desselberger, U., and Ball, L. (2005). The Positive Sense Single Stranded RNA Viruses. In *Virus Taxonomy*, pages 739–1128. Elsevier.
- Fauquet, C. and Thouvenel, J.-C. (1977). Isolation of the *Rice yellow mottle virus* in Ivory Coast. *Plant disease reporter*, 61(June) :443–446.
- Flint-Garcia, S., Thornsberry, J., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annual review of plant biology*, 54.
- Fragoso, C., Aguirre, A., Heffelfinger, C., Moreno, M., Franco, N., Arbelaez, L., Dellaporta, S., and Lorieux, M. (2016). Genome structure of a rice Nested Association Mapping population, and trait mapping. In *ISRFG2016*, Montpellier.

- Frichot, E. and François, O. (2015). *Lea* : an R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8) :925–929.
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4) :973–983.
- Fujisaki, K., Iwahashi, F., Kaido, M., Okuno, T., and Mise, K. (2009). Genetic analysis of a host determination mechanism of bromoviruses in *Arabidopsis thaliana*. *Virus research*, 140 :103–11.
- Fujita, D., Kohli, A., and Horgan, F. G. (2013). Rice Resistance to Planthoppers and Leafhoppers. *Critical Reviews in Plant Sciences*, 32 :162–191.
- Gao, X., Starmer, J., and Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic epidemiology*, 32(4) :361–369.
- Garavito, A., Guyot, R., Lozano, J., Gavory, F., Samain, S., Panaud, O., Tohme, J., Ghesquière, A., and Lorieux, M. (2010). A genetic model for the female sterility barrier between Asian and African cultivated rice species. *Genetics*, 185 :1425–1440.
- García-Arenal, F. and Fraile, A. (2013). Trade-offs in host range evolution of plant viruses. *Plant Pathology*, 62(S1) :2–9.
- Garris, a. J. (2005). Genetic Structure and Diversity in *Oryza sativa* L. *Genetics*, 169(March) :1631–1638.
- George, A. W. and Cavanagh, C. (2015). Genome-wide association mapping in plants. *Theoretical and Applied Genetics*.
- Ghesquiere, A., Séquier, J., Second, G., and Lorieux, M. (1997). First steps towards a rational use of African rice, *Oryza glaberrima*, in rice breeding through a 'contig line' concept. *Euphytica*, 96 :31–39.
- Giannakopoulou, A., Steele, J. F., Segretin, M. E., Bozkurt, T. O., Zhou, J., Robatzek, S., Banfield, M. J., Pais, M., and Kamoun, S. (2015). Tomato i2 immune receptor can be engineered to confer partial resistance to the oomycete phytophthora infestans in addition to the fungus fusarium oxysporum. *Molecular Plant-Microbe Interactions*, 28(12) :1316–1329.
- Gibbs, A. J., Ohshima, K., Phillips, M. J., and Gibbs, M. J. (2008). The prehistory of potyviruses : their initial radiation was during the dawn of agriculture. *PLoS One*, 3(6) :e2523.
- Gillet, F.-X., Cattoni, D. I., Petiot-Bécard, S., Delalande, F., Poignavent, V., Brizard, J.-P., Bessin, Y., Dorselaer, A. V., Declerck, N., Sanglier-Cianféroni, S., Brugidou, C., and Vignols, F. (2013). The RYMV-encoded viral suppressor of RNA silencing P1 is a zinc-binding protein with redox-dependent flexibility. *Journal of molecular biology*, 425 :2423–35.
- Glaszmann, J. C. (1987). Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics*, 74 :21–30.
- Goff, S. A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.-l., Chen, L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavitgian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano,

- T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science (New York, N.Y.)*, 296 :92–100.
- Gómez, P., Rodríguez-Hernández, A., Moury, B., and Aranda, M. A. (2009). Genetic resistance for the sustainable control of plant virus diseases : breeding, mechanisms and durability. *European Journal of Plant Pathology*, 125 :1–22.
- Gu, Y., Zebell, S. G., Liang, Z., Wang, S., Kang, B.-h., and Dong, X. (2016). Nuclear Pore Permeabilization Is a Convergent Signaling Event in Effector-Triggered Immunity. *Cell*, 166 :1526–1538.
- Gupta, P. K., Rustgi, S., and Kulwal, P. L. (2005). Linkage disequilibrium and association studies in higher plants : present status and future prospects. *Plant molecular biology*, 57(4) :461–85.
- Haley, C. S., Knott, S. A., et al. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69(4) :315–324.
- Hamada, H., Tomita, R., Iwadate, Y., Kobayashi, K., Munemura, I., Takeuchi, S., Hikichi, Y., and Suzuki, K. (2007). Cooperative effect of two amino acid mutations in the coat protein of Pepper mild mottle virus overcomes L3-mediated resistance in Capsicum plants. *Virus Genes*, 34 :205–214.
- Hannon, G. J. (2002). RNA interference. *Nature*, 418(6894) :244–251.
- Hayashi, N., Inoue, H., Kato, T., Funao, T., Shiota, M., Shimizu, T., Kanamori, H., Yamane, H., Hayano-Saito, Y., Matsumoto, T., Yano, M., and Takatsuji, H. (2010). Durable panicle blast-resistance gene Pb1 encodes an atypical CC-NBS-LRR protein and was generated by acquiring a promoter through local genome duplication. *The Plant Journal*, 64 :498–510.
- Hébrard, E., Pinel-Galzi, A., Bersoult, A., Siré, C., and Fargette, D. (2006). Emergence of a resistance-breaking isolate of *Rice yellow mottle virus* during serial inoculations is due to a single substitution in the genome-linked viral protein VPg. *The Journal of general virology*, 87(Pt 5) :1369–73.
- Hébrard, E., Pinel-Galzi, A., and Fargette, D. (2008). Virulence domain of the RYMV genome-linked viral protein VPg towards rice rymv1-2-mediated resistance. *Archives of virology*, 153(6) :1161–4.
- Hébrard, E., Poulicard, N., Gérard, C., Traoré, O., Wu, H.-C., Albar, L., Fargette, D., Bessin, Y., and Vignols, F. (2010). Direct interaction between the *Rice yellow mottle virus* (RYMV) VPg and the central domain of the rice eIF(iso)4G1 factor correlates with rice susceptibility and RYMV virulence. *Molecular plant-microbe interactions : MPMI*, 23(11) :1506–13.
- Hill, W. and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, 38(6) :226–231.
- Huang, B. E., Verbyla, K. L., Verbyla, A. P., Raghavan, C., Singh, V. K., Gaur, P., Leung, H., Varshney, R. K., and Cavanagh, C. R. (2015). Magic populations in crops : current status and future prospects. *Theoretical and Applied Genetics*, 128(6) :999–1017.
- Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., Guo, Y., Lu, Y., Zhou, C., Fan, D., Weng, Q., Zhu, C., Huang, T., Zhang, L., Wang, Y., Feng, L., Furuumi, H., Kubo, T., Miyabayashi, T., Yuan, X., Xu, Q., Dong, G., Zhan, Q., Li, C., Fujiyama, A., Toyoda, A., Lu, T., Feng, Q., Qian, Q., Li, J., and Han, B. (2012a). A map of rice genome variation reveals the origin of cultivated rice. *Nature*, 490(7421) :497–501.

- Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C., Fan, D., Lu, Y., Weng, Q., Liu, K., Zhou, T., Jing, Y., Si, L., Dong, G., Huang, T., Lu, T., Feng, Q., Qian, Q., Li, J., and Han, B. (2012b). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature genetics*, 44(1) :32–9.
- Hubert, J., Luzi-kihupi, A., Hébrard, E., and Lyimo, H. J. F. (2016). Farmers' Knowledge and Perceptions of *Rice yellow mottle virus* in Selected Rice Growing Areas in Tanzania. *International Journal of Science and Research*, 5(2) :549–559.
- Hull, S. and Fargette, D. (2005). Sobemovirus. In *Virus Taxonomy : Eighth report of the international committee on taxonomy of viruses*, pages 885–890. Elsevier.
- Hutin, M., Sabot, F., Ghesquière, A., Koebnik, R., and Szurek, B. (2015). A knowledge-based molecular screen uncovers a broad-spectrum *ossweet14* resistance allele to bacterial blight from wild rice. *The Plant Journal*, 84(4) :694–703.
- Ikeda, R., Khush, G., and Tabien, R. (1990). A new resistance gene to bacterial blight derived from *O. longistaminata*. *Jpn J Breed*, 40(Suppl 1) :280–281.
- Ingle, R. a., Carstens, M., and Denby, K. J. (2006). PAMP recognition and the plant-pathogen arms race. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 28(9) :880–9.
- Ingvarsson, P. K. and Street, N. R. (2011). Association genetics of complex traits in plants. *The New phytologist*, 189(4) :909–22.
- International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. *Nature*, 436 :793–800.
- Ioannidou, D., Pinel, A., Brugidou, C., Albar, L., Ahmadi, N., Ghesquière, A., Nicole, M., and Fargette, D. (2003). Characterisation of the effects of a major qtl of the partial resistance to *Rice yellow mottle virus* using a near-isogenic-line approach. *Physiological and Molecular Plant Pathology*, 63(4) :213–221.
- Ishibashi, K. and Ishikawa, M. (2013). The resistance protein Tm-1 inhibits formation of a Tomato mosaic virus replication protein-host membrane protein complex. *Journal of virology*, 87(14) :7933–9.
- Ishibashi, K., Masuda, K., Naito, S., Meshi, T., and Ishikawa, M. (2007). An inhibitor of viral RNA replication is encoded by a plant resistance gene. *PNAS*, 104(34) :13833–13838.
- Ishibashi, K., Miyashita, S., Katoh, E., and Ishikawa, M. (2012). Host membrane proteins involved in the replication of tobamovirus RNA. *Current Opinion in Virology*, 2(6) :693–698.
- Issaka, S., Basso, A., Sorho, F., Onasanya, A., Haougui, A., Sido, A. Y., Aké, S., Fargette, D., and Séré, Y. (2012). Diagnosis and importance of rice yellow mottle disease epidemics in Niger republic . Mali Niger Burkina Faso Nigeria. *Journal of Applied Biosciences*, 5902 :3501–3511.
- Jeon, J.-S., Lee, S., Jung, K.-H., Jun, S.-H., Jeong, D.-H., Lee, J., Kim, C., Jang, S., Lee, S., Yang, K., et al. (2000). T-dna insertional mutagenesis for functional genomics in rice. *The Plant Journal*, 22(6) :561–570.
- Jeong, D.-H., An, S., Park, S., Kang, H.-G., Park, G.-G., Kim, S.-R., Sim, J., Kim, Y.-O., Kim, M.-K., Kim, S.-R., et al. (2006). Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. *The Plant Journal*, 45(1) :123–132.

- Jing, H.-C., Anderson, L., Sturre, M. J. G., Hille, J., and Dijkwel, P. P. (2007). Arabidopsis CPR5 is a senescence-regulatory gene with pleiotropic functions as predicted by the evolutionary theory of senescence. *Journal of experimental botany*, 58(14) :3885–94.
- Jing, H.-C., Sturre, M. J. G., Hille, J., and Dijkwel, P. P. (2002). Arabidopsis onset of leaf death mutants identify a regulatory pathway controlling leaf senescence. *The Plant journal : for cell and molecular biology*, 32(1) :51–63.
- Johannes, F., Porcher, E., Teixeira, F. K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuissou, J., Heredia, F., Audigier, P., et al. (2009). Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet*, 5(6) :e1000530.
- Jones, J. D. G. and Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117) :323–9.
- Jones, M., Dingkuhn, M., Aluko snm, G., and Semon, M. (1997). Interspecific *Oryza Sativa* L. X *O. Glaberrima* Steud. progenies in upland rice improvement. *Euphytica*, 94 :237–246.
- Kam, H., Laing, M., and Ouoba, J. (2013). Rice traits preferred by farmers and their perceptions of *Rice yellow mottle virus* (RYMV) disease in Cascades Region of Burkina Faso. *African Journal of Agricultural . . .*, 8(22) :2703–2712.
- Kang, H. M., Zaitlen, N. a., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics*, 178(3) :1709–23.
- Kato, M., Ishibashi, K., Kobayashi, C., Ishikawa, M., and Katoh, E. (2013). Expression, purification, and functional characterization of an N-terminal fragment of the tomato mosaic virus resistance protein Tm-1. *Protein Expression and Purification*, 89 :1–6.
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., Schwartz, D. C., Tanaka, T., Wu, J., Zhou, S., Childs, K. L., Davidson, R. M., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S. S., Kim, J., Numa, H., Itoh, T., Buell, C. R., and Matsumoto, T. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (New York, N.Y.)*, 6 :4.
- Khan, M. A., Miyoshi, H., Gallie, D. R., and Goss, D. J. (2008). Potyvirus Genome-linked Protein, VPg, Directly Affects Wheat Germ in Vitro Translation : INTERACTIONS WITH TRANSLATION INITIATION FACTORS eIF4F AND eIFiso4F. *Journal of Biological Chemistry*, 283 :1340–1349.
- Kirik, V., Bouyer, D., Schöbinger, U., Bechtold, N., Herzog, M., Bonneville, J.-M., and Hülskamp, M. (2001). CPR5 is involved in cell proliferation and cell death control and encodes a novel transmembrane protein. *Current Biology*, 11(23) :1891–1895.
- Kobayashi, K., Sekine, K.-T., and Nishiguchi, M. (2014). Breakdown of plant virus resistance : can we predict and extend the durability of virus resistance? *Journal of General Plant Pathology*, 80(4) :327–336.
- Komatsu, K., Hashimoto, M., Ozeki, J., Yamaji, Y., Maejima, K., Senshu, H., Himeno, M., Okano, Y., Kagiwada, S., and Namba, S. (2010). Viral-Induced Systemic Necrosis in Plants Involves Both Programmed Cell Death and the Inhibition of Viral Multiplication, Which Are Regulated by Independent Pathways. *Molecular Plant-Microbe Interactions*, 23 :283–293.
- Konaté, G., Traoré, O., and Coulibaly, M. M. (1997). Characterization of *Rice yellow mottle virus* isolates in Sudano-Sahelian areas. *Archives of virology*, 142(6) :1117–24.

- Korte, A. and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS : a review. *Plant methods*, 9(1) :29.
- Kou, Y. and Wang, S. (2010). Broad-spectrum and durability : understanding of quantitative disease resistance. *Current Opinion in Plant Biology*, 13(2) :181–185.
- Kou, Y. and Wang, S. (2012). Toward an understanding of the molecular basis of quantitative disease resistance in rice. *Journal of biotechnology*, 159(4) :283–90.
- Kouassi, N. K., N’Guessan, P., Albar, L., Fauquet, C., and Brugidou, C. (2005). Distribution and Characterization of *Rice yellow mottle virus* : A Threat to African Farmers. *Plant Disease*, 89(2) :124–133.
- Koudamilaro, A., Nwilene, F. E., Togola, A., and Akogbeto, M. (2015). Insect Vectors of *Rice yellow mottle virus*. *Journal of Insects*, 2015 :1–12.
- Lacombe, S., Bangratz, M., Vignols, F., and Brugidou, C. (2010). The *Rice yellow mottle virus* P1 protein exhibits dual functions to suppress and activate gene silencing. *The Plant journal : for cell and molecular biology*, 61(3) :371–82.
- Lapin, D. and Van den Ackerveken, G. (2013). Susceptibility to plant disease : more than a failure of host immunity. *Trends in plant science*, 18(10) :546–554.
- Lecoq, H., Moury, B., Desbiez, C., Palloix, A., and Pitrat, M. (2004). Durable virus resistance in plants through conventional approaches : a challenge. *Virus research*, 100(1) :31–39.
- Lee, J.-H., Muhsin, M., Atienza, G. A., Kwak, D.-Y., Kim, S.-M., De Leon, T. B., Angeles, E. R., Coloquio, E., Kondoh, H., Satoh, K., et al. (2010). Single nucleotide polymorphisms in a gene for translation initiation factor (eif4g) of rice (*oryza sativa*) associated with resistance to rice tungro spherical virus. *Molecular plant-microbe interactions*, 23(1) :29–38.
- Leonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M. G., and Laliberte, J.-F. (2000). Complex Formation between Potyvirus VPg and Translation Eukaryotic Initiation Factor 4E Correlates with Virus Infectivity. *Journal of Virology*, 74 :7730–7737.
- Lewontin, R. (1964). The interaction of selection and linkage. i. general considerations ; heterotic models. *Genetics*, 49(1) :49–67.
- Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, 25(14) :1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., et al. (2009). The sequence alignment/map format and samtools. *Bioinformatics*, 25(16) :2078–2079.
- Li, J.-Y., Wang, J., and Zeigler, R. S. (2014). The 3,000 rice genomes project : new opportunities and challenges for future rice research. *GigaScience*, 3(1) :8.
- Linares, O. F. (2002). African rice (*Oryza glaberrima*) : history and future potential. *Proceedings of the National Academy of Sciences of the United States of America*, 99 :16360–16365.
- Ling, R., Pate, A. E., Carr, J. P., and Firth, A. E. (2013). An essential fifth coding ORF in the sobemoviruses. *Virology*, 446 :397–408.

- Lipka, A. E., Kandianis, C. B., Hudson, M. E., Yu, J., Drnevich, J., Bradbury, P. J., and Gore, M. a. (2015). From association to prediction : statistical methods for the dissection and selection of complex traits in plants. *Current Opinion in Plant Biology*, 24 :110–118.
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. a., Buckler, E. S., and Zhang, Z. (2012). GAPIT : genome association and prediction integrated tool. *Bioinformatics (Oxford, England)*, 28(18) :2397–9.
- Liu, X., Lin, F., Wang, L., and Pan, Q. (2007). The in Silico Map-Based Cloning of Pi36, a Rice Coiled-Coil Nucleotide-Binding Site Leucine-Rich Repeat Gene That Confers Race-Specific Resistance to the Blast Fungus. *Genetics*, 176(4) :2541–2549.
- Liu, Y.-G. and Chen, Y. (2007). High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *BioTechniques*, 43(5) :649–656.
- Loebenstein, G. (2009). *Local Lesions and Induced Resistance*, volume 75. Elsevier.
- Lorieux, M., Ndjondjop M-N, M. N., and Ghesquière, A. (2000). A first interspecific *Oryza sativa* × *Oryza glaberrima* microsatellite-based genetic linkage map. *Theoretical and Applied Genetics*, 100 :593–601.
- Love, A. J., Laval, V., Geri, C., Laird, J., Tomos, A. D., Hooks, M. A., and Milner, J. J. (2007). Components of Arabidopsis defense- and ethylene-signaling pathways regulate susceptibility to Cauliflower mosaic virus by restricting long-distance movement. *Molecular plant-microbe interactions : MPMI*, 20(6) :659–70.
- Mahajan, S. K., Chisholm, S. T., Whitham, S. A., and Carrington, J. C. (1998). Identification and characterization of a locus (*rtm1*) that restricts long-distance movement of tobacco etch virus in *Arabidopsis thaliana*. *The Plant Journal*, 14(2) :177–186.
- Mahesh, H. B., Shirke, M. D., Singh, S., Rajamani, A., Hittalmani, S., Wang, G.-L., and Gowda, M. (2016). Indica rice genome assembly, annotation and mining of blast disease resistance genes. *BMC Genomics*, 17(1) :1–12.
- Mandadi, K. K. and Scholthof, K.-B. G. (2013). Plant immune responses against viruses : how does a virus cause disease? *The Plant cell*, 25(5) :1489–505.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1) :10.
- Maule, A. J., Caranta, C., and Boulton, M. I. (2007). Sources of natural resistance to plant viruses : Status and prospects : Review. *Molecular Plant Pathology*, 8(2) :223–231.
- McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000). Targeted screening for induced mutations. *Nature biotechnology*, 18(4) :455–457.
- McDowell, J. M. and Simon, S. A. (2006). Recent insights into r gene evolution. *Molecular Plant Pathology*, 7(5) :437–448.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit : a mapreduce framework for analyzing next-generation dna sequencing data. *Genome research*, 20(9) :1297–1303.
- Michel, Z., Hilaire, K. T., Mongomaké, K., and Souley, I. (2008). Screening rice (*Oryza sativa* L.) varieties for resistance to *Rice yellow mottle virus*. *Scientific Research and Essays*, 3(9) :416–424.

- Mitchell-Olds, T. (2010). Complex-trait analysis in plants. *Genome biology*, 11(4) :113.
- Moffett, P. (2009). Mechanisms of recognition in dominant R gene mediated resistance. *Advances in virus research*, 75 :1–33.
- Monat, C., Tranchant-Dubreuil, C., Kougbéadjó, A., Farcy, C., Ortega-Abboud, E., Amanzougarene, S., Ravel, S., Agbessi, M., Orjuela-Bouniol, J., Summo, M., et al. (2015). Toggle : toolbox for generic ngs analyses. *BMC bioinformatics*, 16(1) :1.
- Monosi, B., Wisser, R. J., Pennill, L., and Hulbert, S. H. (2004). Full-genome analysis of resistance gene homologues in rice. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 109(7) :1434–47.
- Montarry, J., Cartier, E., Jacquemond, M., Palloix, A., and Moury, B. (2012). Virus adaptation to quantitative plant resistance : erosion or breakdown? *Journal of evolutionary biology*, 25(11) :2242–2252.
- Moskvina, V. and Schmidt, K. M. (2008). On multiple-testing correction in genome-wide association studies. *Genetic epidemiology*, 32(6) :567–573.
- Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A., and Jacquemond, M. (2004). Mutations in Potato virus Y Genome-Linked Protein Determine Virulence Toward Recessive Resistances in *Capsicum annuum* and *Lycopersicon hirsutum*. *Molecular Plant-Microbe Interactions*, 17 :322–329.
- Mundt, C. C. (2014). Durable resistance : A key to sustainable management of pathogens and pests. *Infection, Genetics and Evolution*, 27 :446–455.
- Murray, S. S. (2004). Searching for the origins of African rice domestication. *Antiquity*, 78(300).
- Mysore, K. S. and Ryu, C.-M. (2004). Nonhost resistance : how much do we know? *Trends in Plant Science*, 9(2) :97–104.
- Nabholz, B., Sarah, G., Sabot, F., Ruiz, M., Adam, H., Nidelet, S., Ghesquière, A., Santoni, S., David, J., and Glémin, S. (2014). Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Molecular ecology*, 23(9) :2210–27.
- Nagy, P. D., Pogany, J., and Lin, J.-Y. (2014). How yeast can be used as a genetic platform to explore virus-host interactions : from 'omics' to functional studies. *Trends in microbiology*, 22(6) :309–316.
- Nair, S. and Savithri, H. (2010). Processing of SeMV polyproteins revisited. *Virology*, 396 :106–117.
- Ndjiondjop, M., Semagn, K., Sie, M., Cissoko, M., Fatondji, B., and Jones, M. (2008). Molecular profiling of interspecific lowland rice populations derived from IR64 (*Oryza sativa*) and Tog5681 (*Oryza glaberrima*). *African Journal of Biotechnology*, 7(23) :4219–4229.
- Ndjiondjop, M.-N., Albar, L., Fargette, D., Fauquet, C., and Ghesquière, A. (1999). The Genetic Basis of High Resistance to *Rice yellow mottle virus* (RYMV) in Cultivars of Two Cultivated Rice Species. *Plant Disease*, 83(10) :931–935.
- Ndjiondjop, M.-N., Albar, L., Sow, M., Yao, N., Djedatin, G., Thiemélé, D., and Ghesquière, A. (2013). Integration of molecular markers in rice improvement : A case study on resistance to *Rice yellow mottle virus*. *Realizing Africa's Rice Promise*, page 161.

- Ndjiondjop, M.-N., Brugidou, C., Zang, S., Fargette, D., Ghesquière, A., and Fauquet, C. (2001). High resistance to *Rice yellow mottle virus* in two cultivated rice cultivars is correlated with failure of cell to cell movement. *Physiological and Molecular Plant Pathology*, 59(6) :309–316.
- Neff, M. M., Turk, E., and Kalishman, M. (2002). Web-based primer design for single nucleotide polymorphism analysis. *TRENDS in Genetics*, 18(12) :613–615.
- N'guessan, P., Pinel, A., Caruana, M., Frutos, R., Sy, A., Ghesquiere, A., and Fargette, D. (2000). Evidence of the presence of two serotypes of rice yellow mottle sobemovirus in côte d'ivoire. *European Journal of Plant Pathology*, 106(2) :167–178.
- N'Guessan, P., Pinel, A., Sy, A. A., Ghesquière, A., and Fargette, D. (2001). Distribution, Pathogenicity, and Interactions of Two Strains of *Rice yellow mottle virus* in Forested and Savanna Zones of West Africa. *Plant Disease*, 85(1) :59–64.
- Nicaise, V. (2003). The Eukaryotic Translation Initiation Factor 4E Controls Lettuce Susceptibility to the Potyvirus Lettuce mosaic virus. *PLANT PHYSIOLOGY*, 132 :1272–1282.
- Nicaise, V. (2014). Crop immunity against viruses : outcomes and future challenges. *Frontiers in Plant Science*, 5(November) :1–18.
- Ochola, D. and Tusiime, G. (2011). Survey on Incidences and Severity of *Rice yellow mottle virus* Disease in Eastern Uganda. *International Journal of Plant Pathology*, 2(1) :15–25.
- Opalka, N., Brugidou, C., Bonneau, C., Nicole, M., Beachy, R. N., Yeager, M., and Fauquet, C. (1998). Movement of *Rice yellow mottle virus* between xylem cells through pit membranes. *Proceedings of the National Academy of Sciences*, 95(6) :3323–3328.
- Opalka, N., Tihova, M., Brugidou, C., Kumar, a., Beachy, R. N., Fauquet, C., and Yeager, M. (2000). Structure of native and expanded sobemoviruses by electron cryo-microscopy and image reconstruction. *Journal of molecular biology*, 303(2) :197–211.
- Orjuela, J., Deless, E. F. T., Kolade, O., Chéron, S., Ghesquière, A., and Albar, L. (2013). A Recessive Resistance to *Rice yellow mottle virus* Is Associated with a Rice Homolog of the CPR5 Gene, a Regulator of Active Defense Mechanisms. *Molecular plant-microbe interactions : MPMI*, 26(12) :1455–63.
- Orjuela, J., Sabot, F., Chéron, S., Vigouroux, Y., Adam, H., Chrestin, H., Sanni, K., Lorieux, M., and Ghesquière, A. (2014). An extensive analysis of the African rice genetic diversity through a global genotyping. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*.
- Ouibrahim, L., Mazier, M., Estevan, J., Pagny, G., Decroocq, V., Desbiez, C., Moretti, A., Gallois, J.-L., and Caranta, C. (2014). Cloning of the Arabidopsis rwm1 gene for resistance to Watermelon mosaic virus points to a new function for natural virus resistance genes. *The Plant journal : for cell and molecular biology*, 79(5) :705–16.
- Pajeroska-Mukhtar, K. M., Mukhtar, M. S., Guex, N., Halim, V. a., Rosahl, S., Somssich, I. E., and Gebhardt, C. (2008). Natural variation of potato allene oxide synthase 2 causes differential levels of jasmonates and pathogen resistance in Arabidopsis. *Planta*, 228 :293–306.
- Palloix, A., Ayme, V., and Moury, B. (2009). Durability of plant major resistance genes to pathogens depends on the genetic background, experimental evidence and consequences for breeding strategies. *New Phytologist*, 183 :190–199.

- Peart, J. R., Mestre, P., Lu, R., Malcuit, I., and Baulcombe, D. C. (2005). NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Current Biology*, 15(10) :968–973.
- Pinel, A., Abubakar, Z., Traoré, O., Konaté, G., and Fargette, D. (2003). Molecular epidemiology of the RNA satellite of *Rice yellow mottle virus* in Africa. *Archives of virology*, 148(9) :1721–33.
- Pinel, A., N’Guessan, P., Bousalem, M., and Fargette, D. (2000). Molecular variability of geographically distinct isolates of *Rice yellow mottle virus* in Africa. *Archives of virology*, 145(8) :1621–38.
- Pinel-Galzi, A., Christine, D.-T., Eugénie, H., Cédric, M., Alain, G., and Laurence, A. (sub.). Mutations in *Rice yellow mottle virus* polyprotein p2a involved in rymv2 gene resistance breakdown.
- Pinel-Galzi, A., Mpunami, a., Sangu, E., Rakotomalala, M., Traoré, O., Sérémé, D., Sorho, F., Séré, Y., Kanyeka, Z., Konaté, G., and Fargette, D. (2009). Recombination, selection and clock-like evolution of *Rice yellow mottle virus*. *Virology*, 394(1) :164–72.
- Pinel-Galzi, A., Rakotomalala, M., Sangu, E., Sorho, F., Kanyeka, Z., Traoré, O., Sérémé, D., Poulicard, N., Rabenantoandro, Y., Séré, Y., Konaté, G., Ghesquière, A., Hébrard, E., and Fargette, D. (2007). Theme and variations in the evolutionary pathways to virulence of an RNA plant virus species. *PLoS pathogens*, 3(11) :e180.
- Pinel-Galzi, A., Traoré, O., Séré, Y., Hébrard, E., and Fargette, D. (2015). The biogeography of viral emergence : *Rice yellow mottle virus* as a case study. *Current Opinion in Virology*, 10 :7–13.
- Plante, D., Viel, C., Léonard, S., Tampo, H., Laliberté, J.-F., and Fortin, M. G. (2004). Turnip mosaic virus VPg does not disrupt the translation initiation complex but interferes with cap binding. *Physiological and Molecular Plant Pathology*, 64 :219–226.
- Poland, J. a., Balint-Kurti, P. J., Wisser, R. J., Pratt, R. C., and Nelson, R. J. (2009). Shades of gray : the world of quantitative disease resistance. *Trends in Plant Science*, 14(December) :21–29.
- Poque, S., Pagny, G., Ouibrahim, L., Chague, a., Eyquard, J.-P., Caballero, M., Candresse, T., Caranta, C., Mariette, S., and Decroocq, V. (2015). Allelic variation at the rpv1 locus controls partial resistance to Plum pox virus infection in *Arabidopsis thaliana*. *BMC Plant Biology*, 15 :159.
- Portères, R. (1957). Compagnonnage agraire et génétique biogéographique chez les riz cultivés. *CR Soc. Biogéogr*, 298 :68–99.
- Portères, R. (1962). Berceaux Agricoles Primaires Sur le Continent Africain. *Journal of African History*, 3(2) :195–210.
- Poulicard, N., Pinel-Galzi, A., Fargette, D., and Hébrard, E. (2014). Alternative mutational pathways, outside the VPg, of *Rice yellow mottle virus* to overcome eIF(iso)4G-mediated rice resistance under strong genetic constraints. *The Journal of general virology*, 95(Pt 1) :219–24.
- Poulicard, N., Pinel-Galzi, A., Hébrard, E., and Fargette, D. (2010). Why *Rice yellow mottle virus*, a rapidly evolving RNA plant virus, is not efficient at breaking rymv1-2 resistance. *Molecular plant pathology*, 11(1) :145–54.
- Poulicard, N., Pinel-Galzi, A., Traoré, O., Vignols, F., Ghesquière, A., Konaté, G., Hébrard, E., and Fargette, D. (2012). Historical contingencies modulate the adaptability of *Rice yellow mottle virus*. *PLoS pathogens*, 8(1) :e1002482.

- Price, A., Travis, T., Alves, E. D. P., Douglas, A., McNally, K., Sumon, M., Islam, R., and Norton, G. (2016). The Bengal and Assam Aus Panel for genetic diversity and GWA studies. In *ISRFG2016*, Montpellier.
- Price, A. L., Zaitlen, N. A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11(7) :459–463.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2) :945–59.
- Pumplin, N. and Voinnet, O. (2013). RNA silencing suppression by plant pathogens : defence, counter-defence and counter-counter-defence. *Nature reviews. Microbiology*, 11(11) :745–60.
- Qiu, J., Zhu, J., Fu, F., Ye, C.-Y., Wang, W., Mao, L., Lin, Z., Chen, L., Zhang, H., Guo, L., et al. (2014). Genome re-sequencing suggested a weedy rice origin from domesticated indica-japonica hybridization : a case study from southern china. *Planta*, 240(6) :1353–1363.
- Qu, C., Liljas, L., Opalka, N., Brugidou, C., Yeager, M., Beachy, R. N., Fauquet, C., Johnson, J. E., and Lin, T. (2000). 3D domain swapping modulates the stability of members of an icosahedral virus group. *Structure (London, England : 1993)*, 8(10) :1095–103.
- Qu, F., Ye, X., and Morris, T. J. (2008). Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38) :14732–7.
- Quenouille, J., Montarry, J., Palloix, A., and Moury, B. (2013). Farther, slower, stronger : how the plant genetic background protects a major resistance gene from breakdown. *Molecular Plant Pathology*, 14 :109–118.
- Quenouille, J., Saint-Felix, L., Moury, B., and Palloix, A. (2015). Diversity of genetic backgrounds modulating the durability of a major resistance gene. Analysis of a core collection of pepper landraces resistant to *Potato virus Y*. *Molecular Plant Pathology*, pages n/a–n/a.
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current opinion in plant biology*, 5(2) :94–100.
- Rafalski, A. and Morgante, M. (2004). Corn and humans : recombination and linkage disequilibrium in two genomes of similar size. *TRENDS in Genetics*, 20(2) :103–111.
- Rakotomalala, M., Pinel-Galzi, A., Albar, L., Ghesquière, A., Rabenantoandro, Y., Ramavovololona, P., and Fargette, D. (2008). Resistance to *Rice yellow mottle virus* in rice germplasm in madagascar. *European Journal of Plant Pathology*, 122(2) :277–286.
- Rakshit, S., Rakshit, A., and Patil, J. (2012). Multiparent intercross populations in analysis of quantitative traits. *Journal of genetics*, 91(1) :111–117.
- Ravensdale, M., Bernoux, M., Ve, T., Kobe, B., Thrall, P. H., Ellis, J. G., and Dodds, P. N. (2012). Intramolecular Interaction Influences Binding of the Flax L5 and L6 Resistance Proteins to their AvrL567 Ligands. *PLoS Pathogens*, 8 :e1003004.
- Revers, F., Guiraud, T., Houvenaghel, M.-C., Mauduit, T., Le Gall, O., and Candresse, T. (2003). Multiple resistance phenotypes to lettuce mosaic virus among arabidopsis thaliana accessions. *Molecular plant-microbe interactions*, 16(7) :608–616.

- Robaglia, C. and Caranta, C. (2006). Translation initiation factors : a weak link in plant RNA virus infection. *Trends in plant science*, 11(1) :40–5.
- Ruffel, S., Dussault, M.-H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C., and Caranta, C. (2002). A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). *The Plant journal : for cell and molecular biology*, 32(6) :1067–75.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., and Barrell, B. (2000). Artemis : sequence visualization and annotation. *Bioinformatics*, 16(10) :944–945.
- Rybicki, E. P. (2015). A Top Ten list for economically important plant viruses. *Archives of Virology*, 160 :17–20.
- Sacco, M. A., Mansoor, S., and Moffett, P. (2007). A rangap protein physically interacts with the nb-lrr protein rx, and is required for rx-mediated viral resistance. *The Plant Journal*, 52(1) :82–93.
- Sakai, H., Kanamori, H., Arai-Kichise, Y., Shibata-Hatta, M., Ebana, K., Oono, Y., Kurita, K., Fujisawa, H., Katagiri, S., Mukai, Y., et al. (2014). Construction of pseudomolecule sequences of the aus rice cultivar kasalath for comparative genomics of asian cultivated rice. *DNA research*, page dsu006.
- Sanfaçon, H. (2015). Plant Translation Factors and Virus Resistance. *Viruses*, 7 :3392–3419.
- Sano, Y. (1985). Interspecific cytoplasm substitutions of an Indica strain of *Oryza sativa* L. and *O. Glaberrima* Steud. *Euphytica*, 34(3) :587–592.
- Sano, Y. (1990). The genic nature of gamete eliminator in rice. *Genetics*, 125 :183–91.
- Sarla, N. and Swamy, B. M. (2005). *Oryza glaberrima* : a source for the improvement of *oryza sativa*. *CURRENT SCIENCE-BANGALORE*, 89(6) :955.
- Sarra, S., Oevering, P., Guindo, S., and Peters, D. (2004). Wind-mediated spread of *Rice yellow mottle virus* (RYMV) in irrigated rice crops. *Plant Pathology*, 53(2) :148–153.
- Satheshkumar, P., Lokesh, G., and Savithri, H. (2004). Polyprotein processing : cis and trans proteolytic activities of *Sesbania mosaic virus* serine protease. *Virology*, 318 :429–438.
- Schatz, M. C., Maron, L. G., Stein, J. C., Wences, A., Gurtowski, J., Biggers, E., Lee, H., Kramer, M., Antoniou, E., Ghiban, E., Wright, M. H., Chia, J.-m., Ware, D., McCouch, S. R., and McCombie, W. (2014). Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biology*, 15(11) :506.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2) :461–464.
- Seeholzer, S., Tsuchimatsu, T., Jordan, T., Bieri, S., Pajonk, S., Yang, W., Jahoor, A., Shimizu, K. K., Keller, B., and Schulze-Lefert, P. (2010). Diversity at the Mla Powdery Mildew Resistance Locus from Cultivated Barley Reveals Sites of Positive Selection. *Molecular Plant-Microbe Interactions*, 23 :497–509.
- Sekine, K.-T., Tomita, R., Takeuchi, S., Atsumi, G., Saitoh, H., Mizumoto, H., Kiba, A., Yamaoka, N., Nishiguchi, M., Hikichi, Y., et al. (2012). Functional differentiation in the leucine-rich repeat domains of closely related plant virus-resistance proteins that recognize common avr proteins. *Molecular Plant-Microbe Interactions*, 25(9) :1219–1229.

- Semagn, K., Ndjiondjop M-N, M. N., Lorieux, M., Cissoko, M., Jones, M., and McCouch SR, S. R. (2007). Molecular profiling of an interspecific rice population derived from a cross between WAB 56-104 (*Oryza sativa*) and CG 14 (*Oryza glaberrima*). *African Journal of Biotechnology*, 6(September 2007) :2014–2022.
- Séréme, D., Lacombe, S., Konaté, M., Pinel-Galzi, A., Traoré, V. S. E., Hébrard, E., Traoré, O., Brugidou, C., Fargette, D., and Konaté, G. (2008). Biological and molecular characterization of a putative new sobemovirus infecting *imperata cylindrica* and maize in africa. *Archives of virology*, 153(10) :1813–1820.
- Sharma, N., Sahu, P. P., Puranik, S., and Prasad, M. (2013). Recent advances in plant-virus interaction with emphasis on small interfering RNAs (siRNAs). *Molecular Biotechnology*, 55(1) :63–77.
- Shi, A., Chen, P., Li, D., Zheng, C., Zhang, B., and Hou, A. (2009). Pyramiding multiple genes for resistance to soybean mosaic virus in soybean using molecular markers. *Molecular Breeding*, 23 :113–124.
- Shin, J.-H., Blay, S., McNeney, B., Graham, J., et al. (2006). Ldheatmap : an r function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *Journal of Statistical Software*, 16(3) :1–10.
- Siré, C., Bangratz-Reyser, M., Fargette, D., and Brugidou, C. (2008). Genetic diversity and silencing suppression effects of *Rice yellow mottle virus* and the P1 protein. *Virology journal*, 5(55) :1–12.
- Siré, C. and Brugidou, C. (2002). Les virus du riz : état des lieux, impact et méthodes de lutte. *Virologie*, 6(6) :431–443.
- Sõmera, M., Sarmiento, C., and Truve, E. (2015). Overview on Sobemoviruses and a Proposal for the Creation of the Family Sobemoviridae. *Viruses*, 7 :3076–3115.
- Sorel, M., Svanella-Dumas, L., Candresse, T., Acelin, G., Pitarch, A., Houvenaghel, M., and German-Retana, S. (2014). Key mutations in the cylindrical inclusion involved in lettuce mosaic virus adaptation to eif4e-mediated resistance in lettuce. *Molecular Plant-Microbe Interactions*, 27(9) :1014–1024.
- Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Zuloaga, F. O., Judziewicz, E. J., Filgueiras, T. S., Davis, J. I., and Morrone, O. (2015). A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, 53 :117–137.
- Sow, M., Seck, P. A., Maiga, I. M., Laing, M., Ortiz, R., and Ndjiondjop, M.-N. (2015). Farmers' rice knowledge and adoption of new cultivars in the Tillabéry region of western Niger. *Agriculture & Food Security*, 4 :5.
- Spielman, R. S., McGinnis, R. E., and Ewens, W. J. (1993). Transmission test for linkage disequilibrium : the insulin gene region and insulin-dependent diabetes mellitus (iddm). *American journal of human genetics*, 52(3) :506.
- Stahl, E., Dwyer, G., Mauricio, R., Kreitman, M., and Bergelson, J. (1999). Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. *Nature*, 400(6745) :667–71.
- Sun, X., Cao, Y., Yang, Z., Xu, C., Li, X., Wang, S., and Zhang, Q. (2004). Xa26, a gene conferring resistance to *xanthomonas oryzae* pv. *oryzae* in rice, encodes an lrr receptor kinase-like protein. *The Plant Journal*, 37(4) :517–527.

- Sweeney, M. and McCouch, S. (2007). The Complex History of the Domestication of Rice. *Annals of Botany*, 100 :951–957.
- Takahashi, A., Casais, C., Ichimura, K., and Shirasu, K. (2003). Hsp90 interacts with rar1 and sgt1 and is essential for rps2-mediated disease resistance in arabidopsis. *Proceedings of the National Academy of Sciences*, 100(20) :11777–11782.
- Takken, F. L. W. and Goverse, A. (2012). How to build a pathogen detector : Structural basis of NB-LRR function. *Current Opinion in Plant Biology*, 15(4) :375–384.
- Tameling, W. I., Elzinga, S. D., Darmin, P. S., Vossen, J. H., Takken, F. L., Haring, M. A., and Cornelissen, B. J. (2002). The tomato r gene products i-2 and mi-1 are functional atp binding proteins with atpase activity. *The Plant Cell*, 14(11) :2929–2939.
- Thakur, S., Singh, P. K., Das, A., Rathour, R., Variar, M., Prashanthi, S. K., Singh, A. K., Singh, U. D., Chand, D., Singh, N. K., and Sharma, T. R. (2015). Extensive sequence variation in rice blast resistance gene Pi54 makes it broad spectrum in nature. *Frontiers in Plant Science*, 6 :345.
- Thiémélé, D. E. (2012). *Diversité des gènes de résistance au virus de la panachure jaune du riz (RYMV) chez l'espèce africaine de riz cultivé, Oryza glaberrima Steud : intérêt pour le développement d'une résistance durable à la maladie*. PhD thesis, UFR Biosciences d'Abidjan.
- Thiémélé, D. E., Boisnard, A., Ndjiondjop, M.-N., Chéron, S., Séré, Y., Aké, S., Ghesquière, A., and Albar, L. (2010). Identification of a second major resistance gene to *Rice yellow mottle virus*, RYMV2, in the African cultivated rice species, *O. glaberrima*. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 121(1) :169–79.
- Thoday, J. (1961). Location of polygenes. *Nature*, 191 :368–370.
- Thottappilly, G., LENT, J., Rossel, H., and Sehgal, O. (1992). Rottboellia yellow mottle virus, a new sobemovirus affecting rottboellia cochinchinensis (itch grass) in nigeria1. *Annals of applied biology*, 120(3) :405–415.
- Thottappilly, G. and Rossel, H. (1993). Evaluation of resistance to *Rice yellow mottle virus* in oryza species. *Indian Journal of Virology*, 9(1) :65–73.
- Tomita, R., Sekine, K.-T., Mizumoto, H., Sakamoto, M., Murai, J., Kiba, A., Hikichi, Y., Suzuki, K., and Kobayashi, K. (2011). Genetic basis for the hierarchical interaction between tobamovirus spp. and l resistance gene alleles from different pepper species. *Molecular plant-microbe interactions*, 24(1) :108–117.
- Traoré, M. D., Traoré, V. S. E., Galzi-Pinel, A., Fargette, D., Konaté, G., and Traoré, O. (2008). Abiotic transmission of *Rice yellow mottle virus* through soil and contact between plants. *Pakistan journal of biological sciences : PJBS*, 11(6) :900–4.
- Traoré, O., Pinel-Galzi, A., Issaka, S., Poulicard, N., Aribi, J., Aké, S., Ghesquière, A., Séré, Y., Konaté, G., Hébrard, E., and Fargette, D. (2010). The adaptation of *Rice yellow mottle virus* to the eIF(iso)4G-mediated rice resistance. *Virology*, 408(1) :103–8.
- Traoré, O., Pinel-Galzi, A., Sorho, F., Sarra, S., Rakotomalala, M., Sangu, E., Kanyeka, Z., Séré, Y., Konaté, G., and Fargette, D. (2009). A reassessment of the epidemiology of *Rice yellow mottle virus* following recent advances in field and molecular studies. *Virus research*, 141(2) :258–67.

- Traoré, O., Sorho, F., Pinel, A., Abubakar, Z., Banwo, O., Maley, J., Hébrard, E., Winter, S., Sere, Y., Konaté, G., and Fargette, D. (2005). Processes of diversification and dispersion of *Rice yellow mottle virus* inferred from large-scale and high-resolution phylogeographical studies. *Molecular ecology*, 14(7) :2097–110.
- Trovão, N. S., Baele, G., Vrancken, B., Bielejec, F., Suchard, M. A., Fargette, D., and Lemey, P. (2015). Host ecology determines the dispersal patterns of a plant virus. *Virus Evolution*, 1(1) :vev016.
- Tsuda, S., Kirita, M., and Watanabe, Y. (1998). Characterization of a Pepper Mild Mottle Tobamovirus Strain Capable of Overcoming the L 3 Gene-Mediated Resistance, Distinct from the Resistance-Breaking Italian Isolate. *Molecular Plant-Microbe Interactions*, 11 :327–331.
- Turner, S. D. (2014). qqman : an r package for visualizing gwas results using q-q and manhattan plots. *bioRxiv*.
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. (2007). Primer3plus, an enhanced web interface to primer3. *Nucleic acids research*, 35(suppl 2) :W71–W74.
- van der Hoorn, R. and Kamoun, S. (2008). From Guard to Decoy : A New Model for Perception of Plant Pathogen Effectors. *the Plant Cell Online*, 20(August 2008) :2009–2017.
- VanderPlank, J. (1968). Academic Press, New York.
- VanRaden, P. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science*, 91(11) :4414–4423.
- Vasudevan, K., Gruissem, W., and Bhullar, N. K. (2015). Identification of novel alleles of the rice blast resistance gene Pi54. *Scientific Reports*, 5 :15678.
- Vaughan, D. a., Lu, B.-R., and Tomooka, N. (2008). The evolving story of rice evolution. *Plant Science*, 174(4) :394–408.
- Vaughan, D. a., Morishima, H., and Kadowaki, K. (2003). Diversity in the *Oryza* genus. *Current Opinion in Plant Biology*, 6(2) :139–146.
- Ventelon-Debout, M., Nguyen, T.-T. H., Wissocq, A., Berger, C., Laudie, M., Piégu, B., Cooke, R., Ghesquière, A., Delseny, M., and Brugidou, C. (2003). Analysis of the transcriptional response to *Rice yellow mottle virus* infection in *oryza sativa indica* and *japonica* cultivars. *Molecular Genetics and Genomics*, 270(3) :253–262.
- Verlaan, M. G., Hutton, S. F., Ibrahim, R. M., Kormelink, R., Visser, R. G. F., Scott, J. W., Edwards, J. D., and Bai, Y. (2013). The Tomato Yellow Leaf Curl Virus Resistance Genes Ty-1 and Ty-3 Are Allelic and Code for DFDGD-Class RNA-Dependent RNA Polymerases. *PLoS Genetics*, 9(3) :e1003399.
- Vos, P., Simons, G., Jesse, T., Wijbrandi, J., Heinen, L., Hogers, R., Frijters, A., Groenendijk, J., Diergaarde, P., Reijans, M., Fierens-Onstenk, J., de Both, M., Peleman, J., Liharska, T., Hontelez, J., and Zabeau, M. (1998). The tomato Mi-1 gene confers resistance to both root-knot nematodes and potato aphids. *Nature Biotechnology*, 16 :1365–1369.
- Wang, A. and Krishnaswamy, S. (2012). Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. *Molecular plant pathology*, 13(7) :795–803.

- Wang, C., Zhang, X., Fan, Y., Gao, Y., Zhu, Q., Zheng, C., Qin, T., Li, Y., Che, J., Zhang, M., et al. (2015). Xa23 is an executor r protein and confers broad-spectrum disease resistance in rice. *Molecular plant*, 8(2) :290–302.
- Wang, C.-H., Zheng, X.-M., Xu, Q., Yuan, X.-P., Huang, L., Zhou, H.-F., Wei, X.-H., and Ge, S. (2014a). Genetic diversity and classification of *Oryza sativa* with emphasis on Chinese rice germplasm. *Heredity*, 112(October 2013) :489–496.
- Wang, G.-L., Ruan, D.-L., Song, W.-Y., Sideris, S., Chen, L., Pi, L.-Y., Zhang, S., Zhang, Z., Fauquet, C., Gaut, B. S., Whalen, M. C., and Ronald, P. C. (1998). Xa21D Encodes a Receptor-Like Molecule with a Leucine-Rich Repeat Domain That Determines Race-Specific Recognition and Is Subject to Adaptive Evolution. *The Plant Cell*, 10(5) :765.
- Wang, K.-D., Empleo, R., Nguyen, T. T. V., Moffett, P., and Sacco, M. A. (2014b). Elicitation of Hypersensitive Responses in *Nicotiana glutinosa* by the Suppressor of RNA Silencing Protein P0 from Poleroviruses. *Molecular plant pathology*, pages 1–43.
- Wang, M., Yu, Y., Haberer, G., Marri, P. R., Fan, C., Goicoechea, J. L., Zuccolo, A., Song, X., Kudrna, D., Ammiraju, J. S. S., Cossu, R. M., Maldonado, C., Chen, J., Lee, S., Sisneros, N., de Baynast, K., Golser, W., Wissotski, M., Kim, W., Sanchez, P., Ndjiondjop, M.-N., Sanni, K., Long, M., Carney, J., Panaud, O., Wicker, T., Machado, C. a., Chen, M., Mayer, K. F. X., Rounsley, S., and Wing, R. a. (2014c). The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nature Genetics*, advance on(9) :982–988.
- Wang, M.-B., Masuta, C., Smith, N. a., and Shimura, H. (2012). RNA silencing and plant viral diseases. *Molecular plant-microbe interactions : MPMI*, 25(10) :1275–85.
- Wang, Q. (2014). STV11 encodes a sulphotransferase and confers durable resistance to rice stripe virus - ncomms5768.pdf.
- Wang, S., Gu, Y., Zebell, S. G., Anderson, L. K., Wang, W., Mohan, R., and Dong, X. (2014d). A Noncanonical Role for the CKI-RB-E2F Cell-Cycle Signaling Pathway in Plant Effector-Triggered Immunity. *Cell Host & Microbe*, 16 :787–794.
- Wang, W., Barnaby, J. Y., Tada, Y., Li, H., Tör, M., Caldelari, D., Lee, D.-u., Fu, X.-D., and Dong, X. (2011). Timing of plant immune responses by a central circadian regulator. *Nature*, 470(7332) :110–114.
- Wei, C., Liu, J., Yu, Z., Zhang, B., Gao, G., and Jiao, R. (2013). Talen or cas9—rapid, efficient and specific choices for genome modifications. *Journal of Genetics and Genomics*, 40(6) :281–289.
- Werner, K., Friedt, W., and Ordon, F. (2005). Strategies for Pyramiding Resistance Genes Against the Barley Yellow Mosaic Virus Complex (BaMMV, BaYMV, BaYMV-2). *Molecular Breeding*, 16 :45–55.
- Whitham, S. A., Anderberg, R. J., Chisholm, S. T., and Carrington, J. C. (2000). Arabidopsis RTM2 Gene Is Necessary for Specific Restriction of Tobacco Etch Virus and Encodes an Unusual Small Heat Shock-Like Protein. *The Plant Cell*, 12 :569.
- Wiermer, M., Feys, B. J., and Parker, J. E. (2005). Plant immunity : the eds1 regulatory node. *Current opinion in plant biology*, 8(4) :383–389.
- Xiao, S., Ellwood, S., Calis, O., Patrick, E., Li, T., Coleman, M., and Turner, J. G. (2001). Broad-spectrum mildew resistance in arabidopsis thaliana mediated by rpw8. *Science*, 291(5501) :118–120.

- Xu, X., Lv, Q., Shang, J., Pang, Z., Zhou, Z., Wang, J., Jiang, G., Tao, Y., Xu, Q., Li, X., Zhao, X., Li, S., Xu, J., and Zhu, L. (2014). Excavation of Pid3 Orthologs with Differential Resistance Spectra to Magnaporthe oryzae in Rice Resource. *PLoS ONE*, 9 :e93275.
- Yamaji, Y., Maejima, K., Ozeki, J., Komatsu, K., Shiraishi, T., Okano, Y., Himeno, M., Sugawara, K., Neriya, Y., Minato, N., Miura, C., Hashimoto, M., and Namba, S. (2012). Lectin-mediated resistance impairs plant virus infection at the cellular level. *The Plant cell*, 24(2) :778–93.
- Yang, P., Lüpken, T., Habekuss, A., Hensel, G., Steuernagel, B., Kilian, B., Ariyadasa, R., Himmelbach, A., Kumlehn, J., Scholz, U., Ordon, F., and Stein, N. (2014). PROTEIN DISULFIDE ISOMERASE LIKE 5-1 is a susceptibility factor to plant viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 111(6) :2104–9.
- Yoav Benjamini, Y. H. (1995). Controlling the false discovery rate : A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1) :289–300.
- Yoshida, S., Ito, M., Nishida, I., and Watanabe, A. (2002). Identification of a novel gene HYS1/CPR5 that has a repressive role in the induction of leaf senescence and pathogen-defence responses in Arabidopsis thaliana. *The Plant Journal*, 29(4) :427–437.
- Yu, J., Holland, J. B., McMullen, M. D., and Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics*, 178(1) :539–51.
- Yu, J., Hu, S., Wang, J., Wong, G. K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., and Yang, H. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science (New York, N.Y.)*, 296 :79–92.
- Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D., Holland, J. B., Kresovich, S., and Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics*, 38(2) :203–8.
- Zeng, Z.-B. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences*, 90(23) :10972–10976.
- Zhang, Z., Ersoz, E. S., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., and Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature genetics*, 42(4) :355–60.
- Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., et al. (2007). An arabidopsis example of association mapping in structured samples. *PLoS Genet*, 3(1) :e4.
- Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008). Status and Prospects of Association Mapping in Plants. *The Plant Genome Journal*, 1(1) :5.

- Zhu, Q. and Ge, S. (2005). Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytologist*, 167(1) :249–265.
- Zhu, Y., Qian, W., and Hua, J. (2010). Temperature modulates plant defense responses through nb-lrr proteins. *PLoS Pathog*, 6(4) :e1000844.
- Zuriaga, E., Soriano, J. M., Zhebentyayeva, T., Romero, C., Dardick, C., Cañizares, J., and Badenes, M. L. (2013). Genomic analysis reveals MATH gene(s) as candidate(s) for Plum pox virus (PPV) resistance in apricot (*Prunus armeniaca* L.). *Molecular Plant Pathology*, 14 :663–677.

ANNEXE

A

Tableaux annexes

TABLE A.1 – Accessions de l'espèce *O. glaberrima* de la collection IRD. Les accessions dont le nom est souligné ont été phénotypées pour la résistance partielle. L'origine des graines utilisées pour le test de résistance partielle est précisée dans la dernière colonne du tableau. Chaque accession est associée à son pays d'origine et à ses noms synonymes dans les bases IRRI et IITA. L'allèle de résistance porté par les accessions hautement résistantes est également indiqué.

Nom IRD	Origine géographique	nom IRRI (IRGC)	nom IITA (Tog)	Allèle de résistance	Collection IRIGIN ?	Origine des graines
Og2	Côte d'Ivoire	96720	5283			
Og3	Côte d'Ivoire	96721	5284		oui	
Og5	Nigeria	96725	5293		oui	
Og6	Nigeria	96726	5307	<i>RYMV3</i>		
<u>Og7</u>	Nigeria		5314		oui	IRD
Og8	Nigeria		5321	<i>rymv1-3</i>	oui	
Og9	Nigeria		5324			
<u>Og10</u>	Nigeria		5326		oui	IRD
<u>Og11</u>	Mali		RAM68			
<u>Og12</u>	Nigeria	96740	5390		oui	IRD
Og13	Nigeria		5418	<i>rymv1-3</i>	oui	
Og14	Nigeria	96744	5420			
<u>Og15</u>	Nigeria		5424		oui	IRD
Og16	Nigeria		5428			
<u>Og17</u>	Nigeria		5429		oui	IRD
<u>Og19</u>	Nigeria		5437			
Og20	Nigeria	96751	5438	<i>rymv1-4</i>	oui	
Og21	Nigeria		5440			
<u>Og22</u>	Nigeria		5453		oui	IRD
Og23	Nigeria		5457			
Og24	Nigeria		5463	<i>rymv1-4</i>		
Og25	Nigeria		5467			
Og26	Nigeria		5474	<i>RYMV3</i>		
Og27	Nigeria		5479			
Og28	Nigeria	86752	5486	<i>rymv1-3</i>	oui	
Og29	Nigeria		5491			
<u>Og30</u>	Nigeria		5494		oui	IRD
<u>Og32</u>	Nigeria		5500		oui	IRD
Og33	Nigeria		5503			
Og34	Nigeria		5522			
Og35	Nigeria		5523			
Og36	Nigeria		5556	<i>rymv1-3</i>	oui	
Og38	Ghana		5601			
Og39	Ghana	86762	5602			
Og40	Ghana	86763	5603			
<u>Og41</u>	Ghana	86764	5620		oui	IRD
<u>Og43</u>	Nigeria		5666		oui	IRD
Og44	Nigeria		5672	<i>rymv1-4 + RYMV3</i>	oui	
Og46	Nigeria	96790	5674		oui	
Og47	Nigeria		5675			
Og49	Nigeria	96793	5681	<i>rymv1-3</i>		
Og50	Liberia		5725			
Og51	Liberia	103968, 112565	5747	<i>RYMV3</i>		
<u>Og56</u>	Nigeria		5814		oui	IRD
<u>Og57</u>	Liberia	86785	5815		oui	IRD
Og60	Nigeria	96809	5882			
<u>Og61</u>	Liberia	112568	5885		oui	IRD
<u>Og62</u>	Liberia	86789	5887		oui	IRD
<u>Og63</u>	Liberia	86790	5919		oui	IRD
<u>Og64</u>	Liberia	86791	5923		oui	IRD
<u>Og65</u>	Nigeria		5953		oui	IRD
<u>Og66</u>	Nigeria		5969		oui	IRD

Annexes

Nom IRD	Origine géographique	nom IRRI (IRGC)	nom IITA (Tog)	Allèle de résistance	Collection IRIGIN ?	Origine des graines
Og67	Nigeria		5976			
Og68	Nigeria	96816	5980			
Og69	Nigeria		5984			
Og73	Nigeria		6142			
Og75	Nigeria	112575	6165			
Og76	Guinea	96837	6195			
Og77	Guinea	96839	6197			
Og78	Guinea	96840	6198			
Og79	Guinea	86805	6202			
<u>Og80</u>	Guinea	86806	6203		oui	IRD
<u>Og81</u>	Guinea		6205		oui	IRD
<u>Og82</u>	Zimbabwe	96841	6206		oui	IRD
<u>Og83</u>	Zimbabwe		6207		oui	IRD
Og84	Guinea	96842	6208			
Og85	Sierra Leone	96843	6210			
<u>Og86</u>	Nigeria	96844	6211		oui	Burkina-Faso
<u>Og87</u>	Burkina Faso	112577	6220	<i>rymv2-2</i>	oui	
Og88	Burkina Faso	96849	6221			
<u>Og89</u>	Mali	96854	6231		oui	IRD
Og93	Liberia	86812	6238			
<u>Og94</u>	Liberia		6308		oui	IRD
Og95	Liberia		6334			
Og96	Liberia		6343			
<u>Og97</u>	Liberia	115598	6356		oui	IRD
Og98	Liberia		6549			
Og99	Liberia		6584			
<u>Og100</u>	Liberia		6603		oui	IRD
Og101	Liberia		6664			
<u>Og102</u>	Liberia		6688		oui	Burkina-Faso
Og103	Liberia		6698	<i>rymv2-2</i>	oui	
Og104	Sierra Leone		6790			
<u>Og105</u>	Sierra Leone		6943		oui	IRD
<u>Og106</u>	Sierra Leone		6951		oui	IRD
<u>Og107</u>	Sierra Leone	96888	7020		oui	Burkina-Faso
<u>Og108</u>	Sierra Leone		7047		oui	IRD
<u>Og109</u>	Mali		7106		oui	IRD
<u>Og110</u>	Mali		7108		oui	IRD
<u>Og111</u>	Nigeria		5286	<i>RYMV3</i>	oui	IRD
<u>Og112</u>	Nigeria	96741	5400		oui	IRD
<u>Og113</u>	Nigeria		5439		oui	IRD
<u>Og114</u>	Mali	103512	7115		oui	Burkina-Faso
<u>Og115</u>	Nigeria		5464		oui	IRD
<u>Og116</u>	Nigeria		5533		oui	IRD
<u>Og117</u>	Nigeria		5566		oui	Burkina-Faso
<u>Og118</u>	Ghana		5591		oui	IRD
<u>Og119</u>	Nigeria		5639		oui	Burkina-Faso
<u>Og121</u>	Senegal	103439	7131		oui	IRD
<u>Og122</u>	Senegal		7132		oui	Burkina-Faso
<u>Og124</u>	Senegal		7134		oui	IRD
Og126	Liberia		5747	<i>RYMV3</i>	oui	
<u>Og127</u>	Liberia		5775		oui	IRD
Og128	Liberia		5832			
<u>Og129</u>	Nigeria		5997		oui	IRD
<u>Og130</u>	Sierra Leone		7420		oui	Burkina-Faso
Og131	Côte d'Ivoire		10449			
<u>Og132</u>	Mali	103544	103544		oui	IRD
<u>Og133</u>	Mali		RAM 131	<i>rymv2-4</i>	oui	IRD
<u>Og134</u>	Mali		RAM 137		oui	Burkina-Faso
<u>Og135</u>	Mali		RAM 24		oui	Burkina-Faso
<u>Og136</u>	Mali		RAM 48		oui	IRD

Nom IRD	Origine géographique	nom IRRI (IRGC)	nom IITA (Tog)	Allèle de résistance	Collection IRIGIN ?	Origine des graines
<u>Og137</u>	Mali		RAM 55		oui	Burkina-Faso
<u>Og138</u>	Mali		RAM 77		oui	Burkina-Faso
Og139	Mali		RAM 90			
Og140	Mali		RAM111			
Og142	Senegal	96717	CG 14			
Og143	Senegal	86741	CG17			
Og146	Sierra Leone		Saliforeh			
Og147	Côte d'Ivoire					
Og150	Côte d'Ivoire	103490	7206	<i>rymv2-2</i>	oui	
Og151	Guinea		14367	<i>rymv2-3</i>	oui	
Og152	Guinea		13943	<i>rymv2-3</i>		
Og153	Guinea		13709	<i>rymv2-3</i>		
<u>Og154</u>	Guinea	104194	12444		oui	IRD
<u>Og155</u>	Mali	103538	7213		oui	IRD
<u>Og156</u>	Mali	103539	7214		oui	IRD
<u>Og158</u>	Senegal	103461 96893	7172 7173		oui	IRD
<u>Og159</u>	Mali		7219		oui	IRD
Og160	Mali	103549	7235	<i>rymv2-2</i>	oui	
Og161	Côte d'Ivoire		10434	<i>rymv2-2</i>	oui	
<u>Og162</u>	Chad		7255		oui	IRD
Og163	Nigeria		12086	<i>RYMV3</i>	oui	
Og164	Nigeria		12160	<i>RYMV3</i>	oui	
Og165	Nigeria		12188	<i>rymv1-4</i>	oui	
Og166	Nigeria		12249	<i>rymv1-3</i>	oui	
<u>Og168</u>	Cameroon		7273		oui	IRD
<u>Og169</u>	Cameroon		7274		oui	IRD
Og170	Nigeria		7280			
Og172	Burkina Faso	104589 103469	7291	<i>rymv2-2</i>	oui	
Og177	Ghana		7401			
<u>Og178</u>	Ghana	86826	7402		oui	IRD
<u>Og179</u>	Ghana		7406		oui	IRD
Og180	Burkina Faso		7449			
<u>Og181</u>	Burkina Faso		7451		oui	IRD
<u>Og182</u>	Burkina Faso		7455		oui	IRD
Og183	Burkina Faso		7456	<i>autre rymv2</i>	oui	
<u>Og184</u>	Côte d'Ivoire		7554		oui	IRD
<u>Og185</u>	Nigeria		7993		oui	IRD
Og186	Nigeria		8049	<i>rymv1-3</i>	oui	
<u>Og187</u>	Gambie		8527		oui	IRD
<u>Og188</u>	Gambie		8537		oui	Burkina-Faso
Og189	Gambie		8544			
<u>Og190</u>	Gambie		8545		oui	Burkina-Faso
<u>Og191</u>	Côte d'Ivoire		9524		oui	IRD
Og198	Senegal		12352			
<u>Og200</u>	Côte d'Ivoire		12358		oui	IRD
<u>Og205</u>	Guinea Bissau		12366		oui	IRD
<u>Og206</u>	Ghana		12372		oui	IRD
Og207	Tanzania		12386	<i>rymv1-3</i>		
Og208	Tanzania		12387	<i>rymv1-3</i>	oui	
<u>Og209</u>	Cameroon		12388		oui	Burkina-Faso
Og210	Cameroon		12390			
Og211	Guinea		12397			
<u>Og212</u>	Guinea		12399		oui	IRD
<u>Og213</u>	Guinea		12401	<i>rymv2-3</i>	oui	
Og214	Guinea		12405			
Og215	Guinea		12407			
<u>Og216</u>	Guinea		12411		oui	Burkina-Faso
<u>Og217</u>	Guinea		12414		oui	IRD
Og218	Guinea		12426			
Og219	Guinea		12428			

Annexes

Nom IRD	Origine géographique	nom IRRI (IRGC)	nom IITA (Tog)	Allèle de résistance	Collection IRIGIN ?	Origine des graines
Og220	Guinea		12430			
Og222	Guinea	104188	12437			
<u>Og225</u>	Guinea	104190	12439		oui	Burkina-Faso
Og230	Chad	104042 104294	12446			
Og231	Guinea		13645			
<u>Og234</u>	Guinea		13708		oui	IRD
<u>Og238</u>	Guinea		14093		oui	Burkina-Faso
<u>Og239</u>	Liberia		14116		oui	IRD
<u>Og243</u>	Zimbabwe		14184		oui	Burkina-Faso
<u>Og253</u>	Guinea	105021 115592	14359		oui	IRD
<u>Og256</u>	Guinea		14361	<i>autre rymv2</i>	oui	IRD
Og258	Guinea	105052	14364		oui	oui
<u>Og264</u>	Guinea	105043	14373		oui	IRD
Og273	Guinea	105004	14601			
<u>Og274</u>	Guinea	105007	14602		oui	IRD
<u>Og276</u>	Guinea	105051	14606		oui	IRD
<u>Og280</u>	Guinea	105028	14610		oui	IRD
<u>Og297</u>	Côte d'Ivoire	103475 104573	7190 7191		oui	IRD
<u>Og400</u>	Côte d'Ivoire	103475 104573	7190		oui	IRD
<u>Og401</u>	Côte d'Ivoire	103476	7300		oui	Burkina-Faso
<u>Og402</u>	Côte d'Ivoire	103478	7193		oui	IRD
<u>Og403</u>	Côte d'Ivoire	103479	7194		oui	IRD
<u>Og404</u>	Côte d'Ivoire	103480	7195		oui	IRD
<u>Og405</u>	Côte d'Ivoire	103481	7196		oui	Burkina-Faso
Og406	Côte d'Ivoire	103482	7197	<i>rymv2-3</i>		
<u>Og407</u>	Côte d'Ivoire	103483	7198		oui	IRD
<u>Og408</u>	Côte d'Ivoire	103484	7199		oui	IRD
Og409	Côte d'Ivoire	103485	7200			
<u>Og410</u>	Côte d'Ivoire	103486	7201		oui	IRD
Og411	Côte d'Ivoire	103487	7202	<i>rymv2-2</i>	oui	
<u>Og412</u>	Côte d'Ivoire	103488	7203		oui	Burkina-Faso
Og414	Côte d'Ivoire	103491	7208			
<u>Og415</u>	Côte d'Ivoire	103492	7210		oui	IRD
<u>Og416</u>	Côte d'Ivoire	103493	7211		oui	IRD
Og417	Côte d'Ivoire	104034	12360			
<u>Og418</u>	Côte d'Ivoire	104035	12361		oui	IRD
Og419	Côte d'Ivoire	104036	12363			
Og423	Tanzania	104018		<i>rymv1-4</i>	oui	
Og424	Tanzania	104019				
<u>Og425</u>	Tanzania	104020			oui	IRD
<u>Og426</u>	Cameroon	103593			oui	Burkina-Faso
<u>Og427</u>	Cameroon	103594	7266		oui	Burkina-Faso
Og428	Cameroon	103596	7268		oui	oui
<u>Og429</u>	Cameroon	103598 104582 104583	7273 7274		oui	IRD
<u>Og430</u>	Cameroon	104048			oui	Burkina-Faso
<u>Og431</u>	Cameroon	104049			oui	IRD
Og432	Cameroon	104051				
Og433	Cameroon	103599 104584	7275 7276			
Og434	Cameroon	103600 104585	7278 7279			
Og435	Cameroon	103601	7282			
Og436	Cameroon	103602 104586	7283 7284			
Og437	Mali	103495				
Og439	Mali	103502				
Og440	Mali	103505	7109			
Og443	Mali					
Og445	Mali					
Og446	Mali					
Og447	Mali			<i>rymv2-4</i>		
<u>Og448</u>	Mali				oui	IRD
<u>Og449</u>	Mali					

Nom IRD	Origine géographique	nom IRRI (IRGC)	nom IITA (Tog)	Allèle de résistance	Collection IRIGIN ?	Origine des graines
Og450	Mali	103530	7183			
Og451	Mali					
Og452	Mali			<i>rymv2-4</i>		
Og456	Mali					
Og457	Mali				oui	IRD
Og458	Chad	104039				
Og459	Chad	104041			oui	Burkina-Faso
Og461	Chad	103587 104579 104580	7260 7261		oui	IRD
Og463	Chad	104045			oui	IRD
Og475	Senegal				oui	Burkina-Faso
Og477	Senegal					
Og478	Senegal				oui	IRD
Og485	Senegal					
Og486	Senegal				oui	IRD
Og487	Senegal	103459	7167 7168		oui	IRD
Og488	Senegal				oui	IRD
Og489	Senegal				oui	IRD
Og490	Senegal	103462	7175 7176			
Og491	Senegal			<i>autre rymv2</i>	oui	IRD
Og492	Senegal				oui	IRD
Og495	Senegal	103570			oui	Burkina-Faso
Og496	Senegal				oui	IRD
Og498	Senegal			<i>rymv2-3</i>	oui	
Og500	Guinea					
Og502	Guinea	104174			oui	IRD
Og504	Guinea	104182				
Og505	Guinea	104185			oui	IRD
Og507	Guinea	104189	12438			
Og508	Guinea	104193	12443			
Og438_2	Mali	103499			oui	IRD
Og443_2	Mali	103523	7124		oui	
Og454_2	Mali	103547				
Og455_2	Mali	103552				
Og456_2	Mali				oui	
Og460_1	Chad	103586 104578 104588	7259			
Og462_1	Chad	104044				
Og467_2	Senegal	103326				
Og468_2	Senegal	104565 104566	7127 7128			
Og470_2	Senegal	96718 103330				
Og472_1	Senegal	103339				
Og473_2	Senegal	103341 103446	7141			
Og474_1	Senegal				oui	
Og474_2	Senegal	103448	7147			
Og483_2	Senegal					
Og484_2	Senegal	103458	7166			
Og506_2	Guinea	104187	12436		oui	IRD

TABLE A.2 – Accessions de l'espèce *O. barthii* de la collection IRD. Chaque accession est associée à son pays d'origine et à son nom synonyme dans la base IRRI. L'allèle de résistance porté par les accessions hautement résistantes est également indiqué.

Nom IRD	Origine géographique	nom IRRI (IRGC)	Allèle de résistance	Collection IRIGIN ?
Ob511W1	Cameroon			oui
Ob512A1	Cameroon			
Ob513A1	Cameroon	103590		
Ob513W1	Cameroon	103590		oui
Ob514G1	Mali			oui
Ob515A1	Mali			
Ob516W1	Mali			oui
Ob517A1	Mali			
Ob518G1	Mali			oui
Ob519W1	Mali			oui
Ob520W1	Mali		<i>RYMV3</i>	oui
Ob521A1	Mali			oui
Ob521W1	Mali			
Ob522W1	Mali			oui
Ob523W1	Mali			oui
Ob524W1	Mali			oui
Ob525W1	Mali		<i>autre rymv2</i>	oui
Ob526W1	Mali			oui
Ob527W1	Mali			oui
Ob528A1	Mali			
Ob529W1	Mali		<i>RYMV3</i>	oui
Ob530A1	Mali			
Ob530W1	Mali			oui
Ob531A1	Mali			
Ob532A1	Mali			
Ob532W1	Mali			oui
Ob533W1	Mali			oui
Ob534A1	Mali			oui
Ob535A1	Mali			
Ob535W1	Mali			oui
Ob536A1	Mali			
Ob537W1	Lake Chad			oui
Ob538W1	Lake Chad			oui
Ob539W1	Lake Chad			oui
Ob540W1	Lake Chad			oui
Ob541W1	Lake Chad			oui
Ob542W1	Lake Chad			oui
Ob543W1	Lake Chad			oui
Ob544W1	Lake Chad			oui
Ob545W1	Lake Chad			oui
Ob546W1	Lake Chad			oui
Ob547W1	Lake Chad			oui
Ob548W1	Lake Chad			oui
Ob549W1	Lake Chad			oui
Ob550W1	Lake Chad			oui
Ob551W1	Lake Chad			oui
Ob552W1	Lake Chad			oui
Ob553W1	Lake Chad			oui
Ob554W1	Lake Chad			oui
Ob555W1	Lake Chad			oui
Ob556W1	Botswana			oui
Ob557A1	Guinea Bissau			
Ob557W1	Guinea Bissau			oui
Ob558W1	Senegal			
Ob559A1	Guinea			
Ob560W	Lake Chad	83799		

Nom IRD	Origine géographique	nom IRRI (IRGC)	Allèle de résistance	Collection IRIGIN ?
Ob561W	Zambia	86481		
Ob562W	Lake Chad	86524		oui
Ob563W	Cameroon	86525		oui
Ob564W	Mali	103534		oui
Ob565W	Lake Chad	103578		oui
Ob566W	Lake Chad	103579		oui
Ob567W	Lake Chad	103580		oui
Ob568W	Lake Chad	103581		oui
Ob569W	Lake Chad	103582		oui
Ob570W	Tanzania	103909		oui
Ob571W	Nigeria	104061		oui
Ob572W	Lake Chad	104066		oui
Ob573W	Lake Chad	104069		oui
Ob574W	Cameroon	104072		
Ob575W	Nigeria	104076		oui
Ob576W	Lake Chad	104102		oui
Ob577W	Lake Chad	104104		oui
Ob578W	Lake Chad	104107		oui
Ob579W	Lake Chad	104114		oui
Ob580W	Lake Chad	104115		oui
Ob581W	Lake Chad	104116		oui
Ob582W	Lake Chad	104117		oui
Ob583W	Lake Chad	104120		oui
Ob584W	Lake Chad	104121		oui
Ob585W	Lake Chad	104122		oui
Ob586W	Lake Chad	104124		oui
Ob587W	Cameroon	104136		oui
Ob588W	Cameroon	104138		oui
Ob589W	Cameroon	104141	<i>rymv1-3</i>	oui
Ob590W	Cameroon	104142		
Ob591W	Cameroon	104143		oui
Ob592W	Mali	104284		oui
Ob593W	Mali	104287		oui
Ob594W	Cameroon	104296		oui
Ob595W	Lake Chad	105571		oui
Ob596W	Lake Chad	105606		oui
Ob597W	Nigeria	105609		oui
Ob598W	Lake Chad	105611		oui
Ob599W	Botswana	105613		oui
Ob600W	Cameroon	105983		oui
Ob601W	Cameroon	105985		oui
Ob602W	Cameroon	105987		oui
Ob603W	Cameroon	105988		oui
Ob604W	Cameroon	105989		
Ob605W	Lake Chad	106294		oui
Ob606W	Lake Chad	106295		oui
Ob608W	Lake Chad	106302		oui

TABLE A.3 – Gènes annotés au QTL₆

Gène	Annotation
Os06g26234	1 4-alpha-glucan-branching enzyme
Os06g26270	prefoldin

TABLE A.4 – Gènes annotés au QTL₁₁

Gène	Annotation
Os11g42940	CASP-like protein
Os11g42960	CASP-like protein
Os11g42970	membrane associated DUF588 domain containing protein
Os11g42989	exo70 exocyst complex subunit
Os11g43000	ATEXO70G1
Os11g43030	von Willebrand factor type A domain containing protein
Os11g43049	exo70 exocyst complex subunit
Os11g43200	tropinone reductase 2
Os11g43250	Leucine Rich Repeat family protein
Os11g43320	NBS-LRR type disease resistance protein
Os11g43360	tropinone reductase 2
Os11g43410	Leucine Rich Repeat family protein
Os11g43420	LZ-NBS-LRR class RGA
Os11g43480	Leucine Rich Repeat family protein
Os11g43500	NBS-LRR type disease resistance protein
Os11g43510	lipase precursor
Os11g43520	OsGrx_C17 - glutaredoxin subgroup III
Os11g43530	OsGrx_C13 - glutaredoxin subgroup III
Os11g43550	OsGrx_C14 - glutaredoxin subgroup III
Os11g43580	OsGrx_C16 - glutaredoxin subgroup III
Os11g43600	peptide chain release factor protein
Os11g43610	oxidoreductase 2OG-Fe oxygenase family protein
Os11g43620	LSM domain containing protein
Os11g43650	ompA MotB
Os11g43700	RYMV3
Os11g43740	OsMADS68 - MADS-box family gene with MIKC* type-box
Os11g43750	polygalacturonase
Os11g43760	lipase class 3 family protein
Os11g43770	Leucine Rich Repeat family protein
Os11g43790	DUF581 domain containing protein
Os11g43800	PPR repeat domain containing protein
Os11g43820	myristoyl-acyl carrier protein thioesterase chloroplast precursor
Os11g43830	pectinesterase
Os11g43860	sodium calcium exchanger protein
Os11g43890	WD domain G-beta repeat domain containing protein
Os11g43900	translationally-controlled tumor protein
Os11g43934	pentatricopeptide
Os11g43950	dnaJ domain containing protein
Os11g43960	mitochondrial carrier protein

Gène	Annotation
Os11g43970	AAA-type ATPase family protein
Os11g43980	peroxidase precursor
Os11g44170	calmodulin binding protein
Os11g44250	protein kinase
Os11g44310	calmodulin binding protein
Os11g44430	protein kinase
Os11g44500	protein kinase domain containing protein
Os11g44550	protein kinase family protein
Os11g44560	protein kinase domain containing protein
Os11g44580	go35 NBS-LRR
Os11g44660	protein kinase
Os11g44680	calmodulin binding protein
Os11g44690	cysteine-rich receptor-like protein kinase 19 precursor
Os11g44700	calmodulin binding protein
Os11g44790	EAP30 Vps36 family domain containing protein
Os11g44810	auxin-repressed protein
Os11g44860	cysteine-rich receptor-like protein kinase 28 precursor
Os11g44880	kinesin-4
Os11g44930	pentatricopeptide repeat domain containing protein
Os11g44940	seed maturation protein PM27
Os11g44950	glycosyl hydrolase family 3 protein
Os11g44960	NBS-LRR disease resistance protein
Os11g44970	NBS-LRR disease resistance protein
Os11g44990	NB-ARC domain containing protein
Os11g45050	NBS-LRR disease resistance protein
Os11g45060	NB-ARC domain containing protein
Os11g45090	NB-ARC domain containing protein
Os11g45130	pollen signalling protein with adenylyl cyclase activity
Os11g45160	disease resistance protein
Os11g45180	NBS-LRR disease resistance protein
Os11g45190	NB-ARC domain containing protein
Os11g45220	IWS1 homolog A
Os11g45250	ribonucleases P MRP protein subunit POP1 containing protein
Os11g45330	disease resistance protein RPM1
Os11g45360	DEFL15 - Defensin and Defensin-like DEFL family
Os11g45540	DUF26 kinase
Os11g45560	Tramtrack Broad Complex BTB domain with MATH domain
Os11g45580	embryogenesis transmembrane protein
Os11g45600	rust resistance protein Rp1-dp8
Os11g45620	rust-resistance protein Lr21
Os11g45710.5	Beta-glucosidase-like SFR2 chloroplastic
Os11g45720	pectinesterase
Os11g45730	pectinesterase
Os11g45740	MYB family transcription factor
Os11g45780	RGH2B
Os11g45790	NB-ARC domain containing protein
Os11g45850	WRKY61
Os11g45930	NBS-LRR type disease resistance protein

Gène	Annotation
Os11g45950	NAC domain-containing protein 90
Os11g45970	NBS-LRR disease resistance protein
Os11g45980	NBS-LRR type disease resistance protein
Os11g46000	von Willebrand factor type A domain containing protein

ANNEXE

B

Figures annexes

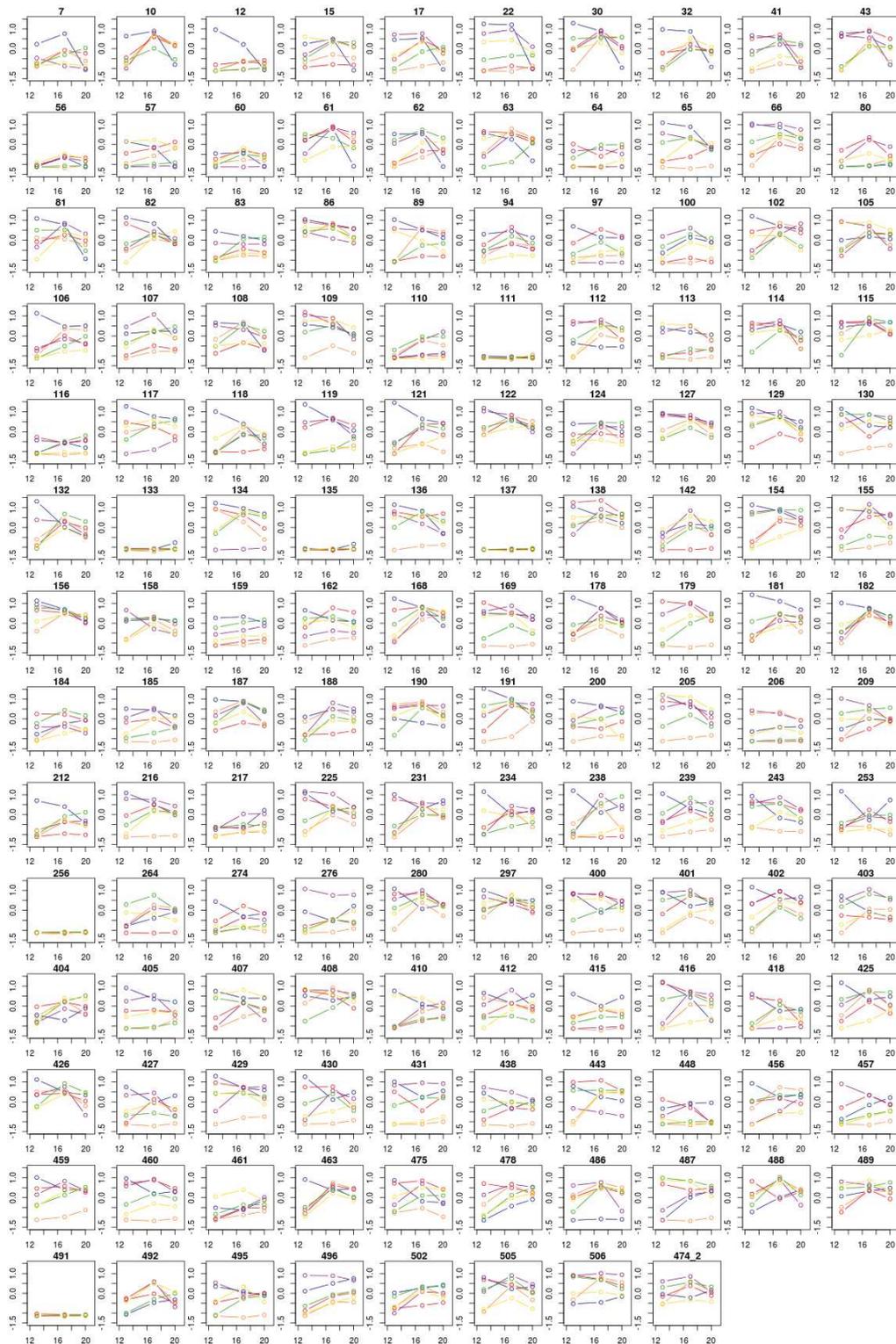


FIGURE B.1 – Accumulation virale mesurée dans les accessions du test de phénotypage aux trois dates et six répétitions. La répétition 1 est en bleu, la répétition 2 en rouge, la répétition 3 en violet, la répétition 4 en orange, la répétition 5 en vert, et la répétition 6 en jaune.

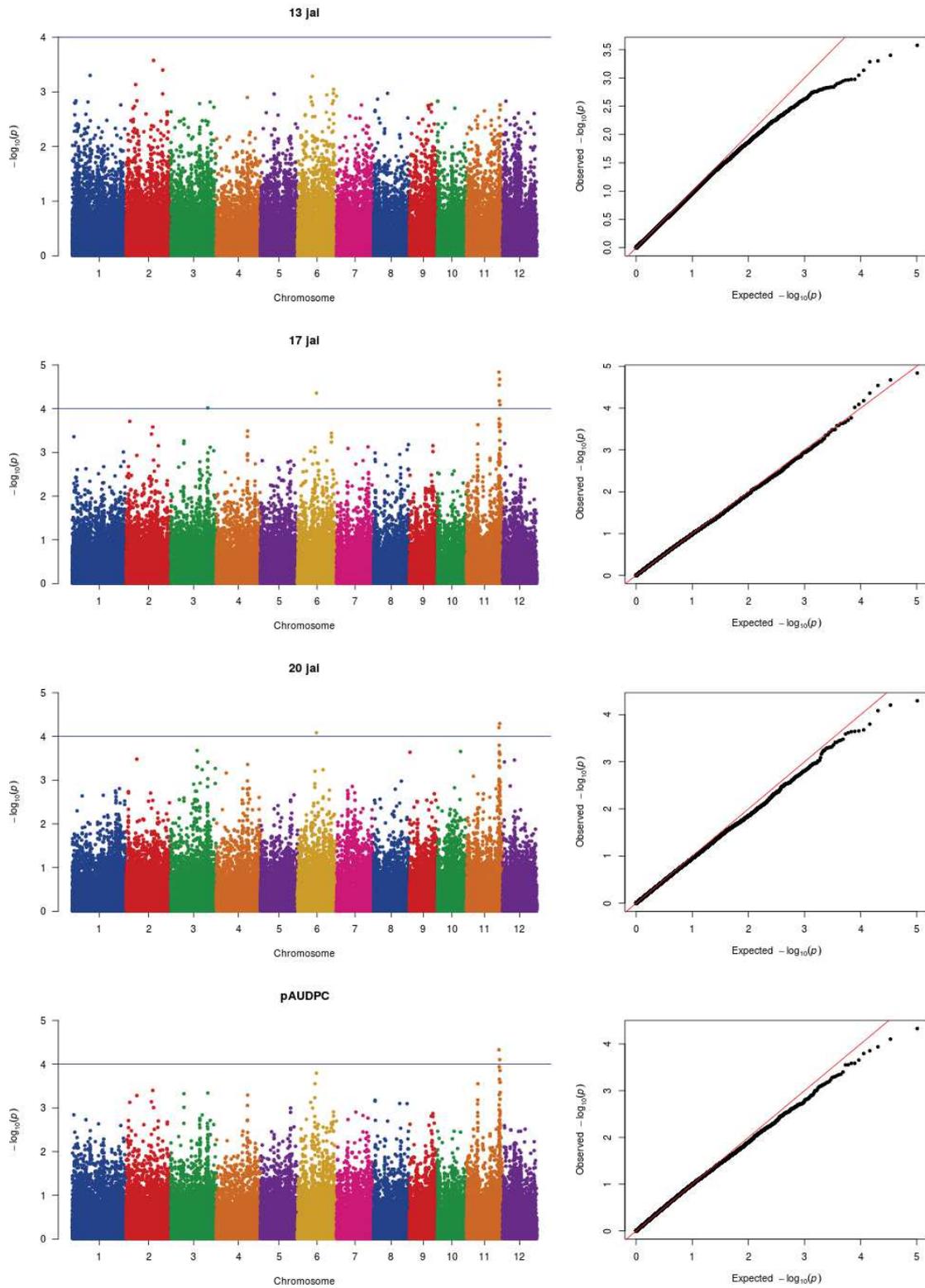


FIGURE B.2 – *Manhattan plots* et *Q-Q plots* obtenus pour les quatre variables de résistance partielle avec le modèle CMLM sans structure de population.

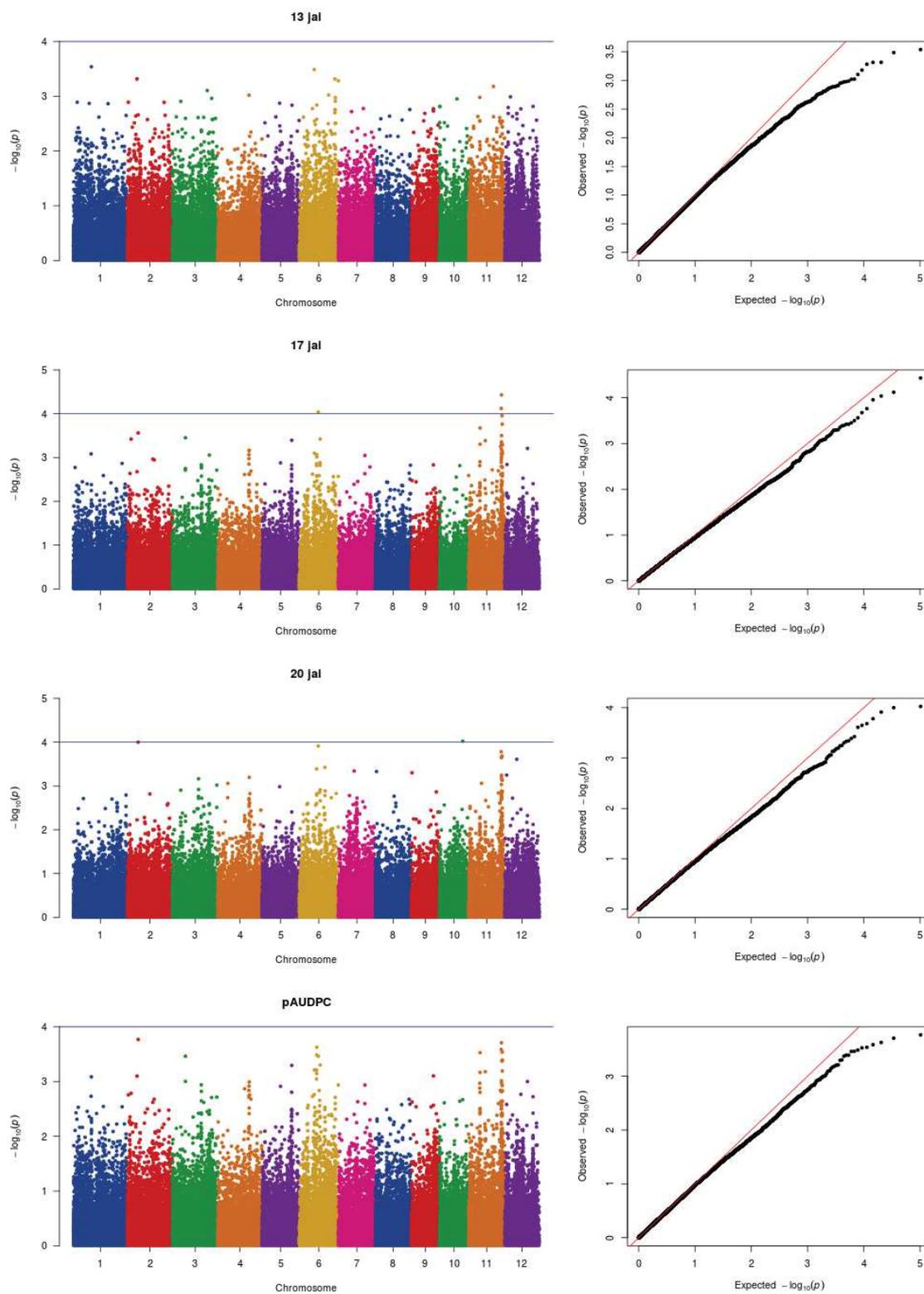


FIGURE B.3 – *Manhattan plots* et *Q-Q plots* obtenus pour les quatre variables de résistance partielle calculés sans la répétition 4 avec le modèle CMLM sans structure de population.

Résumé :

Le virus de la panachure jaune (RYMV) est une contrainte majeure pour la riziculture en Afrique. Deux gènes contrôlant des résistances récessives ont précédemment été décrits : *RYMV1*, qui code pour eIF(iso)4G1, un facteur d'initiation de la traduction et *RYMV2*, qui code pour CPR5-1, un probable composant du pore nucléaire impliqué dans la régulation des mécanismes de défense. Cependant, la capacité du virus à contourner ces résistances justifie la caractérisation de sources de résistance originales, présentes dans les espèces de riz africain *Oryza glaberrima* et *Oryza barthii*. Trois approches complémentaires ont été mises en œuvre afin d'identifier les facteurs génétiques contrôlant ces résistances.

Une approche de cartographie génétique dans des populations bi-parentales a permis l'identification du gène *RYMV3*, contrôlant la résistance de l'accession Tog5307 et sans doute également de l'accession Tog5672. Il s'agit de la première résistance dominante identifiée dans le pathosystème riz/RYMV. *RYMV3* a été cartographié dans un intervalle de 15 kb où deux gènes sont annotés, dont un gène NB-LRR. Des comparaisons de séquences entre accessions résistantes et accessions sensibles suggèrent que le polymorphisme responsable de la résistance est une mutation ponctuelle dans le domaine LRR du gène NB-LRR.

Les deux autres approches ont reposé sur l'exploitation de données de séquençage Illumina de 163 accessions *O. glaberrima* et 84 accessions *O. barthii*. Les accessions *O. glaberrima* ont été phénotypées à la fois pour la résistance élevée et pour la résistance partielle au RYMV, et une partie des accessions *O. barthii* a été évaluée pour la résistance élevée. L'analyse de la variabilité allélique aux trois gènes majeurs de résistance a permis l'identification d'un probable nouvel allèle de résistance à *RYMV1* et de six à *RYMV2*. Ces allèles sont actuellement en cours de validation. D'autre part, une approche de génétique d'association réalisée sur 125 accessions *O. glaberrima* a mis en évidence deux QTL de résistance partielle sur les chromosomes 6 et 11, dont l'un colocalise, en première approche, avec le gène *RYMV3*.

Ce travail a ainsi permis l'identification d'un gène majeur, de deux QTL et de nouveaux allèles de résistance qui contribuent à une meilleure compréhension des interactions riz/RYMV et sont utilisables en sélection pour améliorer la durabilité des variétés résistantes.

Mots clés : riz - virus - gènes de résistance - génétique d'association - cartographie génétique - RYMV

Abstract :

The *Rice yellow mottle virus* (RYMV) is a major constraint for rice production in Africa. Two genes controlling recessive resistances have been previously described : *RYMV1* coding for eIF(iso)4G1, a translation initiation factor, and *RYMV2*, coding for CPR5-1, a probable component of the nuclear pore complex involved in the regulation of defense mechanisms. However, the virus' ability to overcome these resistances highlights the need to characterize new sources of resistance in the African rice species *Oryza glaberrima* and *Oryza barthii*. Three complementary approaches were carried out in order to identify the genetic factors controlling these resistances.

A genetic mapping strategy in bi-parental populations led to the identification of the *RYMV3* gene, controlling resistance in the Tog5307 accession and probably also in the Tog5672 accession. It is the first dominant resistance identified in the rice/RYMV pathosystem. *RYMV3* mapped in a 15 kb interval in which two genes annotated occur, including one NB-LRR gene.

The two other strategies used were based on the utilization of Illumina sequencing data of 163 *O. glaberrima* accessions and 84 *O. barthii* accessions. *O. glaberrima* accessions were phenotyped for both high and partial resistance to RYMV, and the high resistance of a portion of the *O. barthii* accessions was assessed. Analysis of allelic variability at the previously identified genes led to the identification of a probable new resistance allele at *RYMV1* and of six others at *RYMV2*. These alleles are currently undergoing validation. Furthermore, a genome wide association study was carried out on 125 *O. glaberrima* accessions, revealing two partial resistance QTLs on chromosomes 6 and 11, including one colocalized with *RYMV3*.

This work has thus allowed the identification of one major resistance gene, of two QTLs and of new resistance alleles, contributing to a better understanding of rice/RYMV interactions and creating new prospects for the breeding for resistant varieties.

Keywords : rice - virus - resistance genes - genome-wide association mapping - genetic mapping - RYMV