



HAL
open science

Réponse d'*Arabidopsis thaliana* au Turnip mosaic virus (TuMV) en conditions extérieures et en conditions contrôlées : phénotypage fin de traits de maladie et métaboliques et architecture génétique associée

Bernadette Rubio

► To cite this version:

Bernadette Rubio. Réponse d'*Arabidopsis thaliana* au Turnip mosaic virus (TuMV) en conditions extérieures et en conditions contrôlées : phénotypage fin de traits de maladie et métaboliques et architecture génétique associée. Biologie végétale. Université de Bordeaux, 2017. Français. NNT : 2017BORD0758 . tel-01725515

HAL Id: tel-01725515

<https://theses.hal.science/tel-01725515>

Submitted on 7 Mar 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 PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé

SPÉCIALITÉ : Biologie Végétale

Par **Bernadette Rubio**

**Réponse d'*Arabidopsis thaliana* au *Turnip mosaic virus*
(TuMV) en conditions extérieures et en conditions
contrôlées : phénotypage fin de traits de maladie et
métaboliques et architecture génétique associée**

Sous la direction de Mme Valérie Schurdi-Levraud

Soutenue le 20 Décembre 2017

Membres du jury :

M. Rolin Dominique	Professeur de l'Université de Bordeaux	Président
M. Langlade Nicolas	Directeur de Recherche INRA	Rapporteur
M. Vile Denis	Chargé de Recherche INRA	Rapporteur
Mme Albar Laurence	Chargé de Recherche IRD	Examinatrice
M. Gravot Antoine	Maître de Conférences – Univ. de Rennes 1	Examineur

‘ Seamos realistas y hagamos lo imposible ’
Ernesto Che Guevara

Remerciements

Avant de pouvoir remercier l'ensemble des personnes qui ont fait partie de cette aventure de près ou de loin je voudrais juste préciser, que pour moi, cette thèse n'est pas uniquement l'aboutissement de trois ans de travail. Elle conclut surtout une période de ma vie qui a débuté il y a 9 ans par une année de médecine... Je suis bien loin de mes objectifs du début mais finalement heureuse d'être arrivée au bout de cette étape qui ne fut pas toujours un long fleuve tranquille.

Je tiens dans un premier temps à remercier les membres du jury qui ont accepté de juger ce travail, Madame Laurence Albar ainsi que Messieurs Nicolas Langlade, Antoine Gravot, Dominique Rolin et Denis Vile.

Je remercie aussi l'ensemble des personnes qui avaient accepté de faire partie de mon comité de thèse, Madame Sylvie German-Retana ainsi que Messieurs Fabrice Roux, Yves Gibon, Jean-Luc Gallois et Antoine Gravot.

Merci également à Thierry Candresse qui m'a permis de réaliser mon travail de recherche au sein de l'équipe Virologie Végétale de l'UMR 1332 Biologie du Fruit et Pathologie à l'INRA de Bordeaux.

Un énorme merci à ma petite équipe de travail dont j'ai eu la chance de faire partie intégrante pendant ces trois années. Merci Val pour ces trois années d'échange, de partage ! Merci pour la confiance que tu m'as accordé dès le début de notre travail ensemble. Merci aussi pour avoir partagé avec moi ton goût pour l'enseignement. Tu as fait en sorte que je réussisse chaque étape de cette thèse. Tu m'as toujours poussé à me dépasser et aller chercher plus loin pour construire petit à petit mon avenir en apportant « à chaque fois une petite pierre à l'édifice ».

Un énorme Merci à Pat ! Cette thèse aurait été complètement différente sans toi ! D'un point de vue technique parce que je ne savais vraiment pas faire grand-chose au labo, on peut dire que tu m'as vraiment tout appris, tu es mon mentor ! Et d'un point de vue humain parce que derrière ton air un peu « ronchon » tu es juste quelqu'un de génial, qui me fait beaucoup rire (et pas uniquement par tes tenues vestimentaires), avec qui j'ai adoré travailler ! En bref j'ai eu beaucoup de chance de te rencontrer.

Un gros merci à Mélo ! On n'a pas toujours fait des activités ultra-passionnantes au boulot mais on a été quand même un binôme de choc : je pense qu'on est imbattable dans le nombre de plantes inoculées/heure ! Et puis on a passé de bons moments dans une activité bien plus passionnante : la course à pied ! Un petit clin d'œil au passage à toute ta petite famille, Zoé, Noa et Manou !

Merci à Amélie, qui a fait son stage de M1 pendant deux mois dans l'équipe ! C'était ma première expérience d'encadrement et je pense que j'ai eu beaucoup de chance de t'avoir. Tu m'as vraiment facilité la tâche par ta motivation, ta compréhension rapide des consignes et du sujet, tes questions et tes réflexions toujours pertinentes et intéressantes. Une très bonne expérience pour moi. Il ne me reste plus qu'à te souhaiter bonne chance et plein de réussite pour la suite !

Merci à présent à tous les autres membres de cette équipe Viro ! Trois bonnes années passées dans ce labo ! Merci donc à tous pour toutes nos conversations scientifiques ou non et un grand merci à toutes les personnes qui ont pu être impliquées de près ou de loin, avec ou contre leur volonté dans ce projet de thèse : Luc (je suis sûre que les inoculations et récoltes matinales vont te manquer), Chantal, Laurence, Timi, Véro, Sylvie, Olivier, Armelle,

Stéphane, Michel, Pascal, David, Aurélie, Marie, Patrice, Jean-Phillipe, Geneviève, Joce, Vincent S., Thierry L., Guillaume, Amandine Ba., Amandine Bo. Merci aux serristes de l'équipe Coralie en serre HPE et Thierry M. dont l'accueil en serre HC a toujours été au top. Merci à Claudine et Flo pour toutes démarches administratives. Merci aussi aux « anciens » de la Viro que j'ai pu côtoyer au cours de ces trois ans : Fred, Valie, Sébastien, Laure, Sandy et Marie-Christine. Merci à mes colocataires de bureau anciens ou permanents : Merci à Justine qui nous a mis la pression il y a deux ans avec sa soutenance et qui a ENFIN libéré les deux bureaux qu'elle occupait à elle toute seule. Merci à Vincent, on a partagé la même galère cette année, bientôt la fin ! Tu n'auras plus d'excuses pour te lever le soir et t'occuper de ton petit ! 谢谢 Yuxin et Shuo! Merci à Xavier, Cécile, Zoé et tous les stagiaires qui ont vécu au moins pendant quelques temps l'expérience virologiste. Merci à Paul (je te mets avec la Viro au final tu travailles avec des virus aussi) pour toutes nos conversations et surtout pour m'avoir aidé à réussir à utiliser cette fameuse souche de virus GFP.

Un grand merci à l'équipe « Métabo » de notre laboratoire. Une belle collaboration qui m'a beaucoup apportée. Bien évidemment un grand merci à Yves. Merci pour toutes les discussions que nous avons eu, toujours très riches et intéressantes. C'est toujours un plaisir de travailler avec des personnes passionnées par ce qu'elles font et enthousiastes de mener à bien pleins de projets différents. Merci à Olivier, c'était une rencontre pas du tout prévue mais qui a permis de faire avancer le travail sur le métabolisme de façon considérable. Merci pour toutes nos discussions biologiques et statistiques. Merci à Thierry pour nous avoir fait un peu découvrir le monde obscur de la biochimie. Merci à Pierre B. pour les consignes et conseils pour le dosage de l'ascorbate et du glutathion et sur l'utilisation du lyophilisateur. Enfin, merci aux anciens et permanent de la plateforme Hit-Me. J'ai vraiment eu beaucoup de chance de travailler avec tout ce matériel de haute technologie. Un gros merci donc à Patricia, Duyen et Camille.

Merci à Fabrice Roux au LIPM à Toulouse pour toutes les observations, tous les conseils que tu as pu nous donner. Merci pour ton optimisme et ton enthousiasme, on a l'impression que quel que soit le résultat tu arriveras toujours à le mettre en valeur et ça c'est juste génial !

Thanks to Joy Bergelson for all the wise advices concerning our different projects.

Merci aux différentes personnes que j'ai rencontré et avec qui j'ai pu travailler dans le cadre de mon monitorat à l'Université de Bordeaux. Cette mission complémentaire à mon travail de recherche m'a permis de découvrir le monde de l'enseignement supérieur. Même si au début ça m'a fait un peu bizarre de me retrouver de « l'autre côté », j'ai adoré préparer et donner mes cours, échanger avec les étudiants... bref, une super expérience ! Merci donc à ceux qui ont partagé leur goût pour l'enseignement avec moi : Didier, Sophie, Cécile, José, Laure, Pierre-François, Marie, Xavier.

Un grand merci à tous mes partenaires complètement fou de course à pied comme moi : la team Momo du CAV ! Merci à notre « gourou », Momo, qui m'as permis d'explorer mes limites et maintenant j'espère aller encore plus loin ! Merci à Véro, Fatah, Gauthier, Laurent², Gégé, Jango, Benoit, Baroudi, Mouhssine, Kamal, Thomas, Nathan, Ahcene², Eric², Pierre-H, Pierre, Kévin, Romain, Jérémy, Olivier, Bruno.

Merci à Emilie de « l'athlé » pour tous les petits moments passés à discuter ensemble ! J'espère qu'on prendra le temps de plus se voir l'année prochaine entre tous tes déplacements et voyages !

Merci au groupe « rugby » donc merci à Mathilde, Laura, Pauline, Julien, Guillaume et Aurélien pour toutes les soirées partagées !

Merci à Alice pour tous les moments que nous partageons depuis déjà 13 ans ! Nos vies ont changé mais je suis toujours contente d'avoir de tes nouvelles.

Merci à Emilie, je suis très heureuse que notre amitié de la fac est perdurée ! Et bien sûr merci à Florian. Nos petits rendez-vous en ville ou à la campagne sont toujours trop bien avec des repas toujours au top, dignes de top chef !

Merci à Gilles pour tous les moments passés ensemble à BSA, et oui tu sais quand tu avais réussi à sortir de ton sud-est pour venir dans le nord ! Merci également bien sûr à Mathilde et à votre petit Marwan !

Un gros merci à Laura, une amitié commencée au lycée avec ses hauts et ses bas au milieu et maintenant qu'on s'est retrouvée c'est juste génial de pouvoir t'avoir ! J'adore nos discussions qui peuvent durer des heures sur tout et rien ! J'espère qu'on gardera cette complicité encore longtemps, il nous reste tellement à partager ! Et bien sûr merci aussi à Flo pour ton amitié, ta générosité sans oublier tes flans coco de folie !

Gracias à Poupou (tu vois je l'écris en espagnol pour que les gens croient que tu sais le parler) et à Marine ! On adore passer du temps avec vous parce que c'est juste trop génial à chaque fois, comme quoi basques et landais peuvent bien s'entendre parfois ! En attendant tous les autres (avec plus ou moins d'organisation) : vive la team Aviron-Basco Landais !

Merci à Philippe, papa Bellée, pour tous les week-ends, vacances à Bayonne ou encore les super week-ends à Carcan !

Merci à l'ensemble de ma famille avec une petite pensée particulière pour mes grands-parents français et chiliens : mamie Solange, papi Jean, el Tata y la Mamita.

Un énorme merci, ce qui est certainement très loin d'être suffisant, à mes parents. Vous m'avez accompagné dans toutes les étapes de ma vie. Vous avez été mes soutiens inconditionnels dans les bons et surtout dans les mauvais moments. Merci pour cette vie que vous nous avez permis d'avoir. « *No somos perfectos, pero juntos somos un grand equipo* ».

Merci à ma sœur, pour tout ce qu'on partage depuis toujours. Tu es ma sœur, ma meilleure amie, ma complice ! Toujours ensemble depuis le début, tu fais partie intégrante de ma vie donc vivement la suite de nos aventures !

Et bien sûr, merci à toi, Antho pour tout... J'ai tellement de chance de t'avoir rencontré et que tu fasses à présent partie de ma vie dans le passé, présent et futur. Tu es la force tranquille de notre duo, tu es le seul à avoir réussi à canaliser le petit monstre de 1m54 que je suis. Merci pour ton soutien, ta patience, ton attention, ta force, ton amour. Merci de croire en moi, c'est ce qui me rend plus forte. Et, ce que nous réserve l'avenir ? Et bien... il nous reste qu'à le découvrir... ensemble... ♥

Réponse d'*Arabidopsis thaliana* au *Turnip mosaic virus* (TuMV) en conditions extérieures et en conditions contrôlées : phénotypage fin de traits de maladie et métaboliques et architecture génétique associée.

Résumé

Les plantes sont des organismes immobiles qui doivent répondre et s'adapter à des contraintes abiotiques et biotiques. Parmi les stress biotiques, les maladies virales, établies ou émergentes, peuvent être responsables de pertes de rendement majeures aux conséquences économiques importantes. Face aux phytovirus la lutte génétique constitue le moyen de lutte le plus efficace, le plus respectueux de l'environnement et du consommateur. Comprendre l'interaction entre les plantes et les virus reste indispensable pour rechercher de nouvelles sources de résistances. Ce travail de thèse s'intéresse à l'étude du pathosystème naturel *Arabidopsis thaliana*/*Turnip mosaic virus* (TuMV). Les essais ont été menés majoritairement en conditions extérieures permettant une analyse de l'interaction dans un environnement multistress. La réponse d'*A. thaliana* a été explorée par l'étude de traits liés à la maladie et par la variation en métabolites primaires et secondaires. Ce travail a permis **i)** de caractériser de façon fine la réponse d'*A. thaliana* au TuMV en conditions multistress en exploitant la diversité naturelle d'une population mondiale et française **ii)** de déterminer l'architecture génétique de cette interaction par des approches de génétique d'association et de QTL mapping. Plusieurs nouveaux loci potentiellement impliqués dans la réponse ont été identifiés **iii)** de montrer l'intérêt du phénotypage métabolique pour discriminer les accessions en fonction de leur sensibilité au TuMV. La multidisciplinarité des approches constitue la richesse de ce travail de thèse qui contribue à une meilleure caractérisation et compréhension de la réponse des plantes lors d'une infection virale.

Mots clés : *Arabidopsis thaliana* – Diversité naturelle - Potyvirus – *Turnip mosaic virus* (TuMV) – Multistress – Métabolisme primaire – Métabolisme secondaire – Génétique d'association – Quantitative Trait Locus (QTL)

***Arabidopsis thaliana* – *Turnip mosaic virus* (TuMV) interaction in common garden and controlled conditions experiments: disease and metabolic traits phenotyping and genetic architecture**

Abstract

Plants are immobile organisms which have to adapt to abiotic and biotic constraints. Among biotic stress, established or emerging viral diseases, may be responsible for major yield losses with significant consequences. Genetic control is the most effective, environmentally and consumer-friendly way to control viral infections. Understanding plant/virus interactions remains essential to search for new sources of resistance. This work, focuses on the study of the natural pathosystem *Arabidopsis thaliana*/*Turnip mosaic virus* (TuMV). Most of the trials were conducted in common garden conditions allowing the analysis of the interaction in a multistress environment. *A. thaliana*'s response was explored through the study of disease-related traits and the variations in primary and secondary metabolites. This work allows **i)** the fine characterization of *A. thaliana*'s response to TuMV in multistress conditions through the exploration of the natural diversity of a world and French population **ii)** to determine the genetic architecture of this interaction by genome wide associations and QTL mapping. Several new loci potentially involved in the response have been identified **iii)** to highlight the interest of metabolic phenotyping to discriminate accessions according to their susceptibility to TuMV. The multidisciplinary approaches contribute to a better characterization and understanding of plant-virus interaction.

Keywords : *Arabidopsis thaliana* – Natural Diversity – Potyvirus – *Turnip mosaic virus* (TuMV) – Multistress – Common garden experiments – Primary metabolism – Secondary metabolism – Genome wide association – Quantitative Trait Locus (QTL)

Productions Scientifiques

✓ Publications

Rubio B., Cosson P., Caballero M., Revers F., Bergelson J., Roux F., Schurdi-Levraud V. (2017) Field experiments reveal new loci involved in *Arabidopsis thaliana* and Turnip mosaic virus. *Publication en préparation.*

Rubio B., Fernandez O., Cosson P., Berton T., Caballero M., Roux F., Bergelson J., Gibon Y., Schurdi-Levraud V. Primary and secondary metabolic content discriminate *Arabidopsis thaliana* susceptibility to Turnip mosaic virus (TuMV) under field conditions. *Publication en préparation.*

Rubio B., Cosson P., Caballero M., Michon T., Schurdi-Levraud V. Identification of new candidate genes involved in the natural pathosystem interaction *Arabidopsis thaliana* and Turnip mosaic virus (TuMV). *Publication en préparation.*

✓ Communications

Rubio B., Cosson P., Caballero M., Revers F., Gibon Y., Roux F., Schurdi-Levraud V. (2015) *Arabidopsis thaliana* susceptibility to TuMV (Turnip mosaic virus) in natural environment: mapping loci involved in viral charge and preliminary results on associated metabolic profiles. 15èmes Rencontres de Virologie Végétale (RVV). Aussois – France. [Poster]

Rubio B., Cosson P., Caballero M., Revers F., Gibon Y., Roux F., Schurdi-Levraud V. (2016) Metabolic profiles and viral charge genome wide mapping in *Arabidopsis thaliana* in response to TuMV (Turnip mosaic virus) in a natural environment. 20th EUCARPIA General Congress. Zurich – Switzerland. [Poster]

Rubio B., Cosson P., Caballero M., Schurdi-Levraud V. (2017) Fine mapping of a quantitative trait locus involved in *A.thaliana* response to TuMV (Turnip mosaic virus). 16èmes Rencontres de Virologie Végétale (RVV). Aussois – France. [Poster]

Rubio B., Cosson P., Caballero M., Revers F., Gibon Y., Roux F., Schurdi-Levraud V. (2017) A phenotype and genotype study in *Arabidopsis thaliana* in response to TuMV (Turnip mosaic virus) in a natural environment. Are primary metabolites markers of plant response to viral infection? 16èmes Rencontres de Virologie Végétale (RVV). Aussois – France. [Oral]

Rubio B., Cosson P., Caballero M., Revers F., Gibon Y., Roux F., Schurdi-Levraud V. (2017) A phenotype and genotype study in *Arabidopsis thaliana* in response to TuMV (Turnip mosaic virus) in a natural environment. 17th Doctoral School of Health and Life Sciences Day. Talence – France. [Oral – Prix du jury de la meilleure communication orale]

Rubio B., Fernandez O., Cosson P., Caballero M., Gibon Y., Roux F., Schurdi-Levraud V. (2017) A phenotype and genotype study in *Arabidopsis thaliana* in response to TuMV (Turnip mosaic virus) in a natural environment. Are primary metabolites markers of plant response to viral infection? 28th International Conference on Arabidopsis Research. Saint-Louis – Missouri – USA. [Poster]

Activités Transversales complémentaires

- **Mission d'enseignement à l'Université de Bordeaux – Année scolaire 2015-2016**

Durée = 65.08 HETD (Heure Equivalent TD)

- 6 HETD de TD (Travaux Dirigés) et 7.50 HETD de cours intégrés dans le cadre de l'UE (Unité d'Enseignement) Approches expérimentales en biologie – Master 1.
- 16 HETD de TD et 18 heures en présentiel de TP (Travaux Pratiques) dans le cadre de l'UE Génétique moléculaire et fonctionnelle – Licence 2.
- 10.67 HETD de TD dans le cadre de l'UE Ecologie et génétique des populations – Licence 2.
- 7 HETD de TD dans le cadre de l'UE La plante et son environnement – Master 1.

- **Mission d'enseignement à l'Université de Bordeaux – Année scolaire 2016-2017**

Durée = 67.83 HETD

- 8 HETD de TD dans le cadre de l'UE Projets et Méthodologies en Biologie – Biostatistiques – Master 1.
- 8 HETD de TD dans la cadre de l'UE Génétique, génomique et épigénétique des plantes – Master 1.
- 48 heures en présentiel de TP dans le cadre de l'UE Génétique – Licence 2.
- 5 HETD de cours intégrés dans le cadre de l'UE Biotechnologies – Licence 3.

- **Autres Activités dans le cadre des missions d'enseignement à l'Université de Bordeaux – Années scolaires 2015-2016 & 2016-2017**

- Surveillance d'examens en licence (2015-2016 & 2016-2017) et participation au jury des oraux de stages de master 1 (2016-2017).
- Corrections de copies et rapports dans le cadre de l'UE Ecologie et génétique des populations (Licence 2 – 2015-2016), de l'UE Approches expérimentales en biologie (Master 1 – 2015-2016), de l'UE La plante et son environnement (Master 1 – 2015-2016) et de l'UE Projets et Méthodologies en Biologie – Biostatistiques (Master 1 – 2016-2017).
- Participation à l'élaboration de l'examen de statistiques dans le cadre de l'UE Projets et Méthodologies en Biologie – Biostatistiques (Master 1 – 2016-2017).

- **Encadrement d'une étudiante de Master 1 Biologie – Agrosiences** dans le cadre de son stage de fin d'année pour une période de 2 mois (Avril – Juin 2017).

- **Participation au Club Graines de Vie du Pôle Agri Sud-Ouest Innovation dans le cadre du Tremplin Jeunes Chercheurs (2015-2016)**

- Participation à trois sessions sous forme de brainstorming entre acteurs de la recherche scientifique et de l'industrie.
- Obtention d'une bourse permettant la participation à un congrès international : 20th EUCARPIA General Congress – Zurich (Switzerland).

Sommaire

Introduction Générale	1
Revue Bibliographique	9
A. Réponse des plantes aux agents pathogènes : les plantes font de la résistance	11
1 Mécanismes généraux	11
2 Mécanismes de défense contre les maladies virales	13
3 Contournement et Durabilité des résistances	24
B. Interactions plantes/agents pathogènes en conditions de stress multiples	27
1 Les plantes face aux stress multiples	27
2 Des éléments communs de réponse aux stress biotiques et abiotiques	28
3 Interactions stress abiotiques et infections virales	31
C. Elargir le phénotypage de la réponse des plantes aux virus : l'approche métabolomique	35
1 Métabolomique – Métabolites – Biomarqueurs	35
2 Rôle du métabolisme primaire dans la réponse à des stress biotiques	41
3 Métabolisme primaire et infection virales	42
4 Généralités sur le métabolisme secondaire	45
D. Etudier l'architecture génétique des interactions plantes/virus	48
1 Cartographie par analyse de liaison	48
2 Génétique d'association	50
E. Acteurs de cette étude : <i>Arabidopsis thaliana</i> et le <i>Turnip mosaic virus</i> (TuMV)	55
1 Modèle d'étude : <i>Arabidopsis thaliana</i>	55
2 Un potyvirus : le <i>Turnip mosaic virus</i> (TuMV)	59
Chapitre 1 : Architecture génétique de la réponse d'<i>A. thaliana</i> au TuMV en conditions extérieures et phénotype fin de la réponse métabolique	65
Présentation générale du Chapitre 1	67
Chapitre 1 – PARTIE A : Bases génétiques de la réponse d'<i>A. thaliana</i> au TuMV en conditions extérieures	69
A-1 Field experiments reveal new loci involved in <i>Arabidopsis thaliana</i> and <i>Turnip mosaic virus</i> (TuMV) interaction	71
Introduction	72
Material and Method	74
Results	78
Discussion	86
Supplemental data	89
References	105
A-2 Vers une validation fonctionnelle de gènes candidats par l'analyse de mutants	111
Introduction	111
1 Analyse des symptômes	112
2 Analyse de la charge virale à 7 dpi	112
Bilan	113

A-3 Réponse d'une population d'<i>A. thaliana</i> locale française vis-à-vis du TuMV	115
Introduction	115
1 Réponse phénotypique de la population française TOU-A au TuMV	116
2 Analyse comparative des sous-groupes de la population française TOU-A	118
3 Analyse d'association : recherche des déterminants génétiques en réponse au TuMV dans la population française TOU-A	120
Discussion	126
Annexes	131
Chapitre 1 – PARTIE B : Phénotypage métabolique fin de la réponse au TuMV en conditions extérieures	135
B-1 Primary and secondary metabolic content discriminate <i>Arabidopsis thaliana</i> susceptibility to <i>Turnip mosaic virus</i> under field conditions	137
Introduction	138
Material and Method	139
Results	142
Discussion	144
Supporting information	147
References	161
B-2 Cinétique de la réponse d'<i>Arabidopsis thaliana</i> au TuMV en conditions extérieures : évolution des variables maladies et métaboliques pour des génotypes contrastés	165
Introduction	165
1 « Temps zéro » : empreinte métabolique des accessions avant inoculation	166
2 Evolution de la charge virale et du contenu en métabolites primaires	167
Discussion	173
B-3 Architecture génétique de la réponse métabolique d'<i>A. thaliana</i> au TuMV en conditions extérieures	177
Introduction	177
1 GWA et métabolisme primaire chez <i>A. thaliana</i> inoculée par le TuMV	178
2 Qu'en est-il des bases génétiques de la réponse métabolique d' <i>A. thaliana</i> en absence d'infection ?	182
Discussion	184
Annexes	188
Chapitre 2 : Recherche de déterminants génétiques de l'interaction <i>A. thaliana</i>/TuMV via une approche bi-parentale en conditions contrôlées	193
Présentation générale du Chapitre 2	195
Introduction	198
Material and Method	200
Results	203
Discussion	209
Supplemental Data	212
References	216
Discussion Générale – Conclusion	221

Matériel et Méthodes	235
A. Matériel Biologique	236
1 Matériel Viral	236
2 Matériel Végétal	236
B. Description et Caractéristiques des expérimentations	237
1 Expérimentations menées en conditions extérieures	237
2 Expérimentations en conditions contrôlées	239
C. Méthodes Générales	240
1 Méthodes d'inoculation	240
2 Notation de Symptômes	241
3 Evaluation de la charge virale	241
4 Vérification des mutants T-DNA KO	244
5 Dosages ciblés de métabolites primaires	246
6 Dosages non ciblés des métabolites : Analyses UHPLC-LTQ-Orbitrap-MS	253
7 Analyses statistiques des données	257
Bibliographie	259

Liste des Figures et Tableaux

Revue Bibliographique

- Figure R.1** | Principales étapes que les agents pathogènes doivent franchir afin d'induire la maladie chez la plante : **p.11**
- Figure R.2** | Modèle en zig-zag : **p.12**
- Figure R.3** | Mécanismes de RNA silencing : **p.14**
- Figure R.4** | Exemples de mécanismes de suppression du RNA silencing : **p.15**
- Figure R.5** | Représentation schématique de l'architecture linéaire des protéines NB-LRR : **p.16**
- Figure R.6** | Modèles de reconnaissance directe-indirect des NB-LRR : **p.17**
- Figure R.7** | Modèle "bait-and-switch" d'interaction des NB-LRR : **p.18**
- Figure R.8** | Modèle des interactions plante/virus impliquant le RNA silencing et les protéines R : **p.20**
- Figure R.9** | Mécanisme de résistance récessif aux virus : **p.20**
- Figure R.10** | Complexe d'initiation de la traduction : **p.21**
- Figure R.11** | Mécanismes du contournement des résistances par les virus : **p.24**
- Figure R.12** | Effets de la combinaison de stress en comparaison à des stress uniques : **p.27**
- Figure R.13** | Matrices de contrainte : **p.28**
- Figure R.14** | Principales étapes des différentes voies de signalisation activées en réponse à la combinaison entre stress abiotique-biotique : **p.31**
- Figure R.15** | Représentation schématique du concept de triangle de maladie - "Disease triangle" : **p.31**
- Figure R.16** | Effets des interactions stress abiotique-biotique dans la sensibilité des plantes à un agent pathogène : **p.32**
- Figure R.17** | Représentation des différents niveaux de caractérisation des systèmes biologiques : **p.36**
- Figure R.18** | Pipeline pour l'identification de biomarqueurs : **p.37**
- Figure R.19** | Approches à haut-débit verticale et horizontale pour l'établissement d'empreintes ou profils métaboliques : **p.38**
- Figure R.20** | Effet de l'infection par le TRV sur les concentrations en carbohydrates chez *Arabidopsis* : **p.43**
- Figure R.21** | Voies du métabolisme primaire impliquées dans la production d'énergie lors de stress biotiques : **p.44**
- Figure R.22** | Schéma simplifié des principales voies métaboliques impliquées dans la biosynthèse de métabolites secondaires en lien avec les métabolites primaires : **p.46**
- Figure R.23** | Différentes populations utilisées pour l'identification de QTLs par cartographie de liaison : **p.49**
- Figure R.24** | Etendue du déséquilibre de liaison et densité de marqueurs moléculaires associée : **p.51**
- Figure R.25** | Organigramme des analyses GWA : **p.52**
- Figure R.26** | Hétérogénéité génétique et allélique : **p.53**
- Figure R.27** | Scénario de la colonisation postglaciaire de l'Europe par *A. thaliana* : **p.55**
- Figure R.28** | Cycle de vie des accessions d'*A. thaliana* annuelles hivernales et estivales : **p.56**
- Figure R.29** | Analyse des divers travaux publiés sur *A.thaliana* à l'aide de la base de données BIOSIS : **p.57**
- Figure R.30** | Organisation du génome d'un potyvirus et différentes fonctions des protéines virales codées : **p.60**
- Figure R.31** | Principales étapes du cycle viral des potyvirus : **p.61**
- Figure R.32** | Diversité des symptômes causés lors d'une infection par le TuMV chez *B. rapa* et *A. thaliana* : **p.62**

Chapitre 1 – Partie A

Article : *Field experiments reveal new loci involved in A. thaliana and Turnip mosaic virus (TuMV) interaction*

Fig. 1 Distribution of viral accumulations in 2013 and 2014 common garden experiments : **p.79**

Fig. 2 Slopegraph of the response to virus of *Arabidopsis thaliana* accessions between 2013 and 2014 : **p.80**

Fig. 3 Venn Diagrams of the comparison of the genome regions detected in 2013 and 2014 : **p.81**

Fig. 4 Heatmap of GO Term enrichment from the GWA results of 2013 and 2014 common garden experiments : **p.82**

Fig. S1 Experiment design of the 2013 and 2014 common garden experiments : **p.90**

Fig. S2 Symptomatology of *A. thaliana* infected by TuMV in common garden experiments : **p.90**

Fig. S3 Climate raw data for 2013 and 2014 common garden experiments and comparison between the 2 years : **p.91**

Fig. S4 Manhattan plots of the GWA analysis with the different phenotypic traits measured in 2013 and 2014 common garden experiments : **p.92-93**

Fig. S5 Linkage disequilibrium heat maps corresponding to five genome regions mapped in GWA analysis in 2013 and 2014 : **p.94**

Table 1 Phenotypic correlations between the different traits used in the two years of common garden : **p.81**

Table 2 Candidate *A. thaliana* genes resulting from the 25 genomic regions mapped in both 2013 and 2014 : **p.85**

Table S1 List of the different *A. thaliana* accessions used in the two common garden experiments : **p.95-103**

Table S2 List of the 19 resistant *A. thaliana* accessions to TuMV and their respective OD values in 2013 and 2014 : **p.103**

Table S3 Candidate *A. thaliana* genes resulting from the year specific genome regions detected from both 2013 (a) and 2014 (b) GWA analyses : **p.104**

Figure A2.1| Comparaisons des scores d'AUDP entre les mutants T-DNA testés et les témoins d'infection : **p.112**

Figure A2.2 | Quantification de la charge virale par Q-PCR à 7 dpi sur des feuilles inoculées et en systémie des mutants T-DNA et du contrôle d'infection positif Col-0 : **p.113**

Figure A3.1| Distribution des différents traits phénotypiques mesurés en réponse au TuMV pour la population française TOU-A d'*A.thaliana* en conditions extérieures : **p.117**

Figure A3.2| Tableau des corrélations (a) et graphique des variables (b) des différents traits phénotypiques mesurés en réponse au TuMV pour la population française TOU-A d'*A.thaliana* en conditions extérieures : **p.118**

Figure A3.3| Graphique des individus obtenu par l'analyse en composantes principales sur les données phénotypiques mesurées en réponse au TuMV pour la population française TOU-A d'*A.thaliana* : **p.119**

Figure A3.4| Boxplots des différents traits phénotypiques mesurés chez les accessions TOU-A1 et TOU-A6 : **p.119**

Figure A3.5| *Manhattan plot* des analyses GWA menées sur les données phénotypiques obtenues sur l'ensemble de la population TOU-A à 13 dpi (days post inoculation) avec le modèle GEMMA : **p.121**

Figure A3.6| Diagramme de Venn des zones du génome identifiées par les analyses GWA : **p.122**

Figure A3.A1 Comparaison de la hauteur des hampes et de la biomasse sèche entre échantillons inoculés (I) et non inoculés (M) par le TuMV en conditions extérieures : **p.131**

Figure A3.A2 Manhattan plot obtenus avec les analyses GWA sur les données de hauteur de hampe des échantillons mock-inoculés (a) et inoculés (b) avec le TuMV : **p.132**

Figure A3.A3 Données phénotypiques et *Manhattan plot* concernant la sous-population TOU-A1 : **p.132**

Figure A3.A4 Données phénotypiques et *Manhattan plot* concernant la sous-population TOU-A6 : p.133

Tableau A2.1 | Liste des trois gènes choisis et de leurs mutants T-DNA associés pour la validation fonctionnelle : p.111

Tableau A3.1 | Zones du génome mises en évidence par les analyses GWA sur la population TOU-A menées sur les différents traits phénotypiques mesurés : p.123

Tableau A3.2 | Zones du génome identifiées à partir des analyses GWA à partir des 4 traits phénotypiques obtenus sur la sous-population TOU-A1 : p.124

Tableau A3.3 | Zones du génome identifiées à partir des analyses GWA à partir des 4 traits phénotypiques obtenus sur la sous-population TOU-A6 : p.125

Tableau A3.A1 Liste des accessions française TOU-A1 communes entre les différents essais extérieurs (2013-2014-2016) : p.131

Chapitre 1 – Partie B

Article : *Primary and secondary metabolic content discriminate Arabidopsis thaliana susceptibility to Turnip mosaic virus under field conditions*

Figure 1. Principal component (PC) analysis of the 130 and 29 *A. thaliana* accessions inoculated by the TuMV in the 2014 and 2015 common garden experiments : p.143

Fig. S 1. Principal component (PC) analysis of the 130 mock-inoculated accessions of the 2014 common garden experiment : p.147

Fig. S 2. Principal component (PC) analysis of the 29 mock-inoculated accessions of the 2015 common garden experiment : p.147

Fig. S 3 : Distribution of the 130 *A. thaliana* accessions of the 2014 common garden experiment according to their OD value : p.148

Fig. S 4 Dry biomass of TuMV- and mock-inoculated samples in 2014 (A) and 2015 (B) common garden experiments : p.148

Fig. S 5. Comparisons of primary metabolic content in TuMV-inoculated resistant and susceptible 130 *A. thaliana* accessions of 2014 common garden experiment : p.148

Fig. S 6. Comparisons of primary metabolic content in Mock-inoculated and TuMV-inoculated resistant and susceptible 29 *A. thaliana* accessions of 2015 common garden experiment : p.149

Fig. S 7. Dry biomass on TuMV- and mock-inoculated samples of resistant and susceptible accessions in the 2015 common garden experiment : p.149

Fig. S 8. Principal component analysis performed on the 505 metabolic signatures (m/z) measured on the 29 *A. thaliana* accessions of the 2015 common garden experiment : p.150

Fig. S 9. OPLS-DA analysis and its parameters of validation for the TuMV-inoculated resistant and susceptible 29 accessions of the 2015 common garden experiment : p.150

Fig. S 10. Prediction of the viral accumulation by the metabolite matrix : p.150

Table S 1 . Spearman correlations between dry biomass and the primary metabolites measured on the 130 and 29 *A. thaliana* accessions of the 2014 and 2015 common garden experiments : p.151

Table S 2. Spearman correlations between dry biomass and viral charge and the primary metabolites content in the two common garden experiments : p.151

Figure B2.1 | Analyse en composante principale à partir des concentrations en 10 métabolites primaires : p.166

Figure B2.2 | Concentrations des 10 métabolites primaires dosées chez les 11 accessions d'*A. thaliana* : p.166

Figure B2.3 | Evolution de la biomasse fraîche et des concentrations des 10 métabolites primaires des 11 accessions d'*A.thaliana* testées à 5-7-9 et 13 dpi (days post inoculation) en conditions extérieures : p.167

Figure B2.4| Cinétique de l'accumulation de la charge virale pour les différentes accessions d'*A.thaliana* testées en conditions extérieures : **p.168**

Figure B2.5| Evolution du contenu en 10 métabolites primaires chez les échantillons mock-inoculés et chez les échantillons inoculés avec le TuMV : **p.171**

Figure B2.6| Analyses en composante principale réalisées à 5, 7, 9 et 13 dpi en utilisant les concentrations en métabolites primaires des échantillons inoculés par le TuMV : **p.172**

Figure B3.1| Distribution des concentrations en 10 métabolites primaires dosées dans 130 accessions d'*A. thaliana* en conditions extérieures : **p.178**

Figure B3.2| Coefficient de corrélation de Spearman entre les données métaboliques pouvant être considérées comme des indicateurs clés du métabolisme de l'azote et du carbone : **p.180**

Figure B3.A1| Manhattan plots correspondants aux différentes analyses GWA à partir des concentrations en métabolites primaires des échantillons inoculés avec le TuMV : **p.188-189**

Figure B3.A2 Manhattan plots correspondants aux différentes analyses GWA à partir des concentrations en métabolites primaires des échantillons mock-inoculés : **p.190-192**

Tableau B2.1| Coefficients de corrélations de Spearman calculés pour les 10 métabolites primaires par rapport à la biomasse fraîche : **p.167**

Tableau B2.2| Pourcentages d'infection calculés pour chaque accession à 13 dpi : **p.169**

Tableau B2.3| Comparaisons du contenu en métabolites primaires entre échantillons infectés (I) et non infectés (M) par le TuMV : **p.170**

Tableau B2.4 | Coefficient de corrélation de Spearman calculés entre les 10 métabolites primaires, la biomasse fraîche et la charge virale chez les échantillons inoculés par le TuMV : **p.173**

Tableau B3.1| Héritabilités estimées pour chacun des traits métaboliques : **p.178**

Tableau B3.2| Nombre de zones du génome identifiées sur chaque chromosome et pour chaque trait métabolique : **p.179**

Tableau B3.3| Régions du génome identifiées à partir des analyses GWA sur les données métaboliques, de charge virale et de biomasse sèche : **p.180**

Tableau B3.4| Nombre d'associations mises en évidence sur chaque chromosome à partir des coordonnées des individus obtenues sur 3 analyses ACP différentes : **p.181**

Tableau B3.5| Régions du génome identifiées par les analyses GWA à partir des coordonnées des individus sur les deux axes principaux des ACP regroupant les variables par voies métaboliques : **p.182**

Tableau B3.6| Héritabilités (H^2) pour chacun des traits métaboliques mesurés en absence d'infection : **p.182**

Tableau B3.7| Zones du génome identifiées sur chaque chromosome et pour chaque trait métabolique chez échantillons mock-inoculés : **p.183**

Chapitre 2

Article : *Identification of new candidate genes involved in the natural pathosystem interaction Arabidopsis thaliana and Turnip mosaic virus (TuMV)*

Fig. 1. AUDPC scores of Bay-0 and Shahdara accessions and controls Col-0 and E6 : **p.203**

Fig. 2. Evolution of viral accumulation during TuMV infection in inoculated and systemic leaves : **p.204**

Fig. 3. Observation of green fluorescence protein (GFP)-tagged signal to different parts of *A. thaliana* plants : **p.204**

Fig. 4. Linkage mapping realized on the RIL population with three different quantitative traits : **p.205**

Fig. 5. Phenotypic characterization of heterogeneous inbred HIF 68 : **p.206**

Fig. 5. Schematic representation of RILs used to fine map the locus identified : **p.206**

Figure S1 Analytical protocol of the in silico analysis of the QTL interval of interest : **p.212**

Figure S2 Phenotypic characteristics of Bay-0 and Shahdara accessions at 23 dpi : **p.213**

Figure S3 QTL analysis performed with QTLNetwork software : **p.214**

Figure S4 Comparison of the proteins models coded by the Col-0 and Bay-0 sequence of the AT1G62510 gene : **p.214**

Table 1 Characteristics of the markers developed for the fine mapping of the locus detected on chromosome 1 : **p.202**

Table 2 List of the 18 non-synonymous polymorphic genes in Bay-0 identified in the QTL interval : **p.207**

Table 3 Effect of non-synonymous mutations on protein sequence and stability through in silico modeling analysis : **p.208**

Table S1 Heterozygous regions characteristics of the HIF068 from RIL BAY-0 X Shahdara progeny on chromosome 1 : **p.212**

Table S2 Statistical inheritance analysis in the F2 Bay-0 X Shahdara population at 15 dpi : **p.213**

Table S3 Phenotypic variation for all the traits measured on the Bay-0 X Shahdara RIL population : **p.213**

Table S4 List of the 61 non-synonymous polymorphic genes of the QTL interval : **p.215**

Matériel et Méthodes

Figure M1 | Souche TuMV Uk1 clonée dans le plasmide pCambiaTunos-GFP : **p.236**

Figure M2 | Design expérimental des essais mis en place en conditions extérieures : **p.238**

Figure M3 | Prélèvements de feuilles inoculées et en systémie pour chaque mutant analysé : **p.239**

Figure M4 | Feuilles inoculées sur une rosette d'*A.thaliana* : **p.240**

Figure M5 | Echelle de notation de symptômes d'*A. thaliana* infectée par le TuMV : **p.241**

Figure M6 | Couples d'amorces (LP/RP & LBa1/RP) utilisés pour vérifier le statut des mutants T-DNA : **p.245**

Figure M7 | Détermination du statut des mutants T-DNA testés avec les couples d'amorces LP/RP et LBa1/RP : **p.246**

Figure M8 | Schéma réactionnel impliquant une déhydrogénase : **p.249**

Figure M9 | Schéma réactionnel pour le dosage du glucose - fructose - saccharose : **p.250**

Figure M10 | Schéma réactionnel de la première étape du dosage de l'amidon (dégradation de l'amidon) : **p.251**

Figure M11 | Méthode permettant d'obtenir les concentrations pour chaque métabolite des standards biologiques : **p.252**

Figure M12 | Principe pour l'obtention des concentrations en métabolites primaires des standards sans gamme de dilution en prenant l'exemple GFS : **p.253**

Figure M13 | Différentes étapes du pré-traitement des données mzXML en utilisant le package xcms de R : **p.256**

Tableau M1 | Liste des mutants T-DNA d'*A.thaliana* testés : **p.237**

Tableau M2 | Bilan des caractéristiques spécifiques de chaque essai mené en conditions extérieures : **p.238**

Tableau M3 | Seuil pour les résultats des analyses ELISA : **p.242**

Tableau M4 | Caractéristique des amorces utilisées pour la quantification virale par Q-PCR : **p.244**

Tableau M5 | Description de la composition du mix A & B pour le dosage de l'ascorbate : **p.252**

- Introduction Générale -

L'agriculture du 21^{ème} siècle doit relever un certain nombre de défis afin de subvenir aux besoins de 9 milliards d'individus à l'horizon 2050. Il faudra donc **produire plus** mais également **produire mieux** afin de **préserver les ressources naturelles** de la planète. Il faudra également savoir **adapter notre agriculture aux changements climatiques**. Les plantes évoluent dans des environnements où elles sont soumises à de multiples stress. La combinaison de divers stress a et aura des impacts importants sur les rendements au champ et par conséquent sur les variétés utilisées et les itinéraires culturaux associés. Il est donc indispensable de mener les différentes études de recherche dans des conditions qui vont **se rapprocher le plus des conditions au champ** où stress abiotiques/biotiques se combinent. Les progrès de la biologie moderne – notamment en termes de production, gestion et analyses de données – permettent maintenant de **traiter des situations complexes** correspondant à des **combinaisons de stress**.

Les changements climatiques vont avoir également un impact important sur l'aire de répartition des agents pathogènes et ravageurs des plantes. Cela favorise **l'émergence de nouvelles maladies** dans des régions où elles étaient jusqu'à présent absentes. De plus, ces changements climatiques peuvent augmenter la sensibilité des plantes aux stress biotiques. Parmi les stress biotiques, les **maladies virales** causent d'importants dégâts avec des pertes économiques pouvant atteindre au niveau mondial les 30 milliards de dollars par an. De plus, les maladies virales constituent **50% des nouvelles maladies émergentes** (Anderson *et al.*, 2004). L'absence de méthodes de luttés directes rend ces agents pathogènes encore plus problématiques. Les principaux moyens de lutte reposent sur la prévention des contaminations (utilisation de semences traitées, désherbage des parcelles afin de diminuer les sources d'inoculum) et/ou la lutte contre les insectes vecteurs des virus via l'utilisation de pesticides. Cependant, cette lutte chimique est inadaptée par rapport aux objectifs du plan Ecophyto 2 qui visent à réduire de 50% le recours aux produits phytosanitaires à l'horizon 2025. Dans ce contexte, la **lutte génétique** constitue le moyen de lutte le plus efficace, le plus rentable et le plus respectueux de l'environnement et du consommateur. La mise en œuvre de programmes de sélection permettant d'introduire des gènes de résistance a déjà permis de développer plusieurs variétés résistantes aux virus actuellement commercialisées. Néanmoins, les virus présentent la capacité à s'adapter et contourner les résistances plus ou moins rapidement remettant en question la durabilité des variétés résistantes. La combinaison de plusieurs résistances au sein d'une même variété (pyramidage des gènes) ou le déploiement alterné dans le temps (rotation des gènes) et/ou l'espace (gestion des paysages) peuvent assurer une gestion durable des résistances. Néanmoins, il apparaît comme indispensable d'**identifier de nouvelles sources de résistance** aux maladies virales afin de limiter leur impact sur les espèces cultivées. Les données acquises jusqu'à présent, montrent que certaines bases moléculaires de la sensibilité des plantes aux virus sont relativement bien conservées dans le règne végétal. Ces caractéristiques rendent possible l'utilisation de **plantes modèles** telles que *Arabidopsis thaliana* et par la même occasion l'utilisation de nombreux outils génétiques et moléculaires qui peuvent s'avérer très intéressants pour l'identification et la caractérisation des mécanismes de résistance aux virus. Le **transfert des connaissances** à des espèces cultivées étant facilité par le développement des nouvelles technologies (séquençage haut débit, outils biotechnologiques : mutagenèse dirigée, transgénèse, CRISPR Cas9).

Parmi les différents genres viraux, le genre **Potyvirus** est responsable d'important dégâts et pertes en agriculture. Il constitue de ce fait le support d'étude de l'équipe virologie végétale de l'UMR 1332 Biologie du Fruit et Pathologie à l'INRA de Bordeaux. Différents pathosystèmes impliquant la plante modèle *A.thaliana* et un potyvirus sont analysés. Le pathosystème constitué par ***A.thaliana*/Turnip mosaic virus (TuMV)** est particulièrement intéressant puisqu'il s'agit d'un **pathosystème naturel** : des populations naturelles d'*A.thaliana* ont été retrouvées naturellement infectées par le TuMV.

Le développement des **approches « omiques »** a permis d'appréhender la complexité du vivant à différents niveaux : génomique, transcriptomique, protéomique, **métabolomique**. Caractérisées par des techniques à haut débit, elles permettent d'analyser simultanément un grand nombre de variables et d'échantillons. Il est important de noter que jusqu'à présent, la plupart des études visant à caractériser la réponse des plantes à des infections virales étaient basées sur des traits phénotypiques directement liés à la maladie (symptômes, charge virale). Analyser cette réponse en exploitant d'autres variables telles que les variables métaboliques pourrait permettre d'identifier de nouveaux mécanismes de l'interaction plante/virus et ainsi compléter l'arsenal disponible pour une lutte plus durable. Les virus sont des parasites obligatoires des plantes qui vont de ce fait être à l'origine d'importantes modifications dans leur plante hôte. Ainsi, le métabolisme primaire et secondaire des plantes sont vraisemblablement impactés dès les premières phases de l'infection. Comprendre les **perturbations du métabolisme** lors d'une infection virale constitue un moyen novateur et intéressant de caractériser les mécanismes de réponse des plantes aux virus.

C'est dans ce contexte général que s'inscrit ce travail de thèse qui avait donc pour **objectif** de caractériser la réponse d'*A. thaliana* au TuMV en explorant la diversité naturelle d'*A. thaliana*. Pour cela différents essais ont été mis en place en **conditions contrôlées** mais également et de façon plus originale en conditions **extérieures**. La réponse au virus a été considérée de façon large en analysant des traits de maladie mais également en analysant les variations métaboliques ainsi que leur architecture génétique associée.

La première partie de ce manuscrit fait un état de l'art de l'ensemble des notions concernant les grandes thématiques de ce travail de thèse. Ainsi, il s'agissait de faire un tour d'horizon des travaux réalisés et les connaissances acquises sur **i)** la réponse des plantes aux agents pathogènes et de façon plus spécifique et détaillée sur les interactions plantes/virus, **ii)** sur les études menées en conditions de stress multiples soulignant leur intérêt et importance, **iii)** l'exploration des perturbations du métabolisme dans l'étude des interactions plantes/virus, **iv)** les approches de génétiques quantitatives avec notamment les avantages et limites de deux approches en particulier : la cartographie par analyse de liaison et la génétique d'association, **v)** les acteurs du pathosystème naturel étudié : *A. thaliana* et le TuMV.

Afin de faciliter la lecture, le matériel et méthode de l'ensemble de ce travail de thèse a été placé à la fin de ce manuscrit.

L'ensemble des résultats obtenus au cours de ce travail de thèse ont été organisés en deux chapitres. Le **premier chapitre** s'intéresse à la réponse fine d'un large panel d'accessions d'*A. thaliana* au TuMV en conditions extérieures. Ce chapitre est divisé en deux grandes parties **i)**

la première partie s'intéresse au phénotypage de variables maladie qui ont été utilisées dans le but de caractériser les bases génétiques de la réponse d'*A. thaliana* au TuMV dans ces conditions. Elle regroupe les résultats des analyses de génétique d'association (GWA) menées sur des accessions de la population mondiale mais également sur une population locale, française. Les résultats obtenus sur la population mondiale font l'objet d'une publication, en préparation, **ii) la deuxième partie** concerne la réponse métabolique d'*A. thaliana* au TuMV. Les principaux résultats de cette réponse métabolique sont retrouvés sous la forme d'une deuxième publication en préparation. Ces résultats ont été complétés par des analyses sur l'évolution du contenu en métabolites primaires chez *A. thaliana* en conditions extérieures. Enfin, les variables métaboliques ont été utilisées dans des analyses GWA dans le but de caractériser l'architecture génétique de la réponse métabolique d'*A. thaliana* au TuMV dans ces conditions.

Le **deuxième chapitre** de cette thèse concerne la recherche de facteurs de l'hôte impliqués dans l'interaction *A.thaliana*/TuMV via une stratégie bi-parentale. Un travail de QTL mapping a été mené sur une population de RIL issue du croisement entre deux parents au phénotype contrasté. Les résultats de l'ensemble de ce travail ont été rédigés sous la forme d'une troisième publication également en préparation.

L'ensemble des principaux résultats ayant largement été analysés et discutés au sein des différentes publications et parties de cette thèse, la **discussion** menée dans ce manuscrit avait pour objectif de discuter sur les différents aspects novateurs de ce travail de thèse et leur potentielle contribution à la compréhension des interactions plantes/virus voire même de façon plus large aux interactions plantes/agents pathogènes.

Réponse d'*Arabidopsis thaliana* au Turnip mosaic virus (TuMV) en conditions extérieures et en conditions contrôlées : phénotypage fin de traits de maladie et métaboliques et architecture génétique associée

- OBJECTIFS**
- Caractérisation des bases génétiques de la réponse d'*A. thaliana* au TuMV en conditions extérieures et en conditions contrôlées
 - Identification de nouveaux gènes candidats de l'interaction *A. thaliana*/TuMV
 - Phénotypage fin métabolique afin de mieux caractériser l'interaction *A. thaliana*/TuMV

CHAPITRE 1 = Architecture génétique de la réponse d'*A. thaliana* au TuMV en conditions extérieures et phénotypage fin de la réponse métabolique

PARTIE A = Bases génétiques de la réponse d'*A. thaliana* au TuMV en conditions extérieures

PARTIE B = Phénotypage métabolique fin de la réponse au TuMV en conditions extérieures

Population mondiale

Population française

Population mondiale

Phénotypage fin de la réponse d'*A. thaliana* au TuMV

= Traits liés à la maladie

- Accumulation virale
- Fréquence d'infection
- Symptomatologie
- Hauteur des hampes
- Biomasse sèche

= Traits liés au métabolisme

- Dosage ciblés de métabolites primaires
- Dosages non ciblés de métabolites

- Analyses de **génétique d'association (GWA)** à partir des traits de maladie (population mondiale et française) et à partir des traits métaboliques
- **Analyses statistiques** des données

Publication en préparation N°1

Publication en préparation N°2

CHAPITRE 2 = Recherche de déterminants génétiques de l'interaction *A. thaliana*/TuMV via une approche bi-parentale en conditions contrôlées

Population RIL Bay-0 X Shahdara

Phénotypage

= Traits liés à la maladie

- Accumulation virale
- Fréquence d'infection
- Symptomatologie

- **QTL mapping**
- **Cartographie fine** du locus

Publication en préparation N°3

- Revue Bibliographique -

A. Réponse des plantes aux agents pathogènes : les plantes font de la résistance

1 Mécanismes généraux

Les agents pathogènes sont à l'origine de maladies qui peuvent avoir d'importants impacts notamment en termes de production agricole. Pourtant même si dans la nature les plantes sont en interaction continue avec divers agents pathogènes, peu développent des maladies. La maladie semble donc être une exception qui pourrait s'expliquer par le fait que les plantes sont sensibles vis-à-vis d'un faible nombre de micro-organismes. En effet, une plante peut être résistante si l'agent pathogène n'est pas adapté pour l'infecter. Cette résistance totale de la plante est la forme la plus commune d'immunité, il s'agit de la **résistance non hôte** (Heath, 2000; Mysore & Ryu, 2004). Afin de comprendre ce type de résistance, il est important d'identifier ce dont les agents pathogènes ont besoin et les obstacles qu'ils doivent surmonter afin de pouvoir se développer dans la plante (Thordal-Christensen, 2003). Les principales étapes cruciales à l'installation des agents pathogènes sont présentées sur la figure R.1.

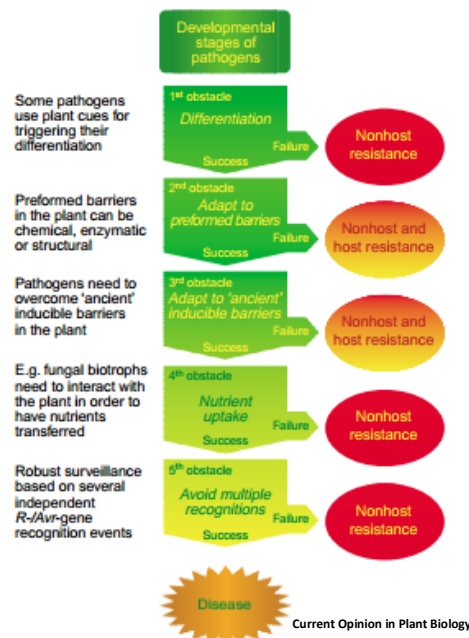


Figure R.1] : Principales étapes que les agents pathogènes doivent franchir afin d'induire la maladie chez la plante. (D'après Thordal-Christensen, 2003)

Les mécanismes de défense non hôte sont notés à chaque étape où ils ont lieu. 1^{er} obstacle, après détection de certains signaux chez les plantes, les agents pathogènes induisent des différenciations cellulaires et expriment des gènes essentiels à leur pathogénicité ; 2^{ème} obstacle, première barrières de l'hôte auxquelles l'agent pathogène doit faire face (paroi cellulaire, enzymes anti-microbiennes) ; 3^{ème} obstacle, les agents pathogènes doivent faire face à des barrières mises en place lors de précédentes attaques (épaississement cuticule, phytoalexines, MAP kinases) ; 4^{ème} obstacle, détournement de la machinerie de l'hôte par les agents pathogènes afin d'assurer leur nutrition ; 5^{ème} obstacle, reconnaissance spécifique des agents pathogènes (interaction gène pour gène)

Le contournement d'un de ces mécanismes conduit à la perte de cette résistance non hôte et donc à la capacité des agents pathogènes à infecter la plante. Les plantes ont de ce fait mis en place d'autres mécanismes de défense qui constituent la **résistance hôte**. Cette résistance se caractérise par deux grands mécanismes (Jones & Dangl, 2006). Ainsi dans un premier temps, la **résistance basale** (ou non spécifique) se met en place. Cette résistance est liée à la reconnaissance de motifs moléculaires associés aux pathogènes (PAMP) également appelée

PTI pour PAMP Triggered Immunity. Ce premier niveau de défense est peu spécifique mais permet de limiter la propagation de nombreux agents pathogènes via les dépôts de callose, la production d'éthylène, des chocs oxydatifs ou encore l'induction de certains gènes liés aux mécanismes de défense (Ingle *et al.*, 2006). Cette résistance va cependant pouvoir être contournée par les agents pathogènes qui produisent des effecteurs qui vont bloquer les mécanismes de défense basale, mais vont également altérer la physiologie de la plante afin de promouvoir leur développement (Chisholm *et al.*, 2006).

En réponse, les plantes ont développé des mécanismes de reconnaissance intracellulaires (directs ou indirects) de ces effecteurs, il s'agit de récepteurs appelés les protéines R. Cette réponse constitue la **résistance spécifique** également appelée immunité déclenchée par des effecteurs : **ETI** pour Effector Triggered Immunity. Cette résistance est souvent associée à une réponse d'hypersensibilité (HR) liée à un phénomène rapide de mort cellulaire, à la production d'espèce réactives de l'oxygène (ROS) ainsi que d'acide salicylique (SA) et à l'expression de gènes de défense (Win *et al.*, 2012).

Le **modèle Zig-Zag** représente bien de façon schématique la co-évolution entre les protéines végétales et les protéines des agents pathogènes impliquées dans l'immunité des plantes (Jones & Dangl, 2006) (Fig. R.2).

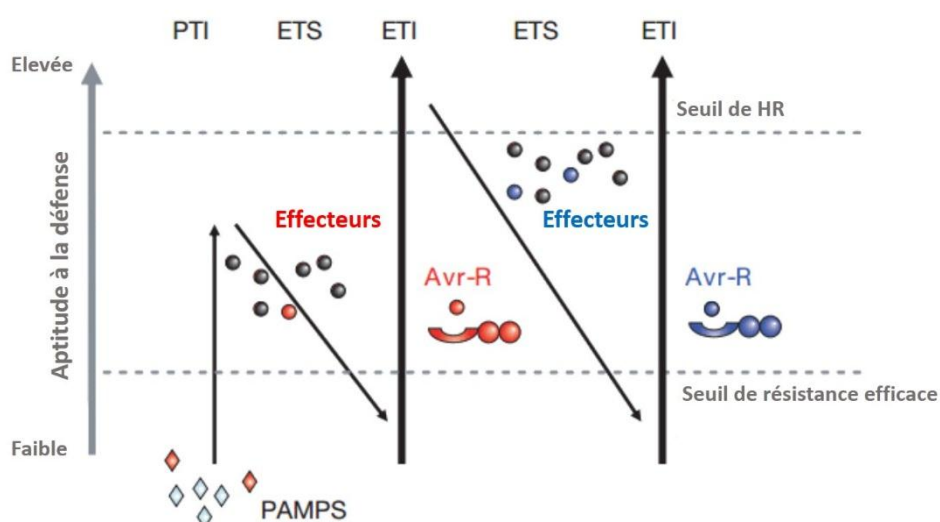


Figure R.2 | Modèle en zig-zag (D'après Jones & Dangl, 2006)

Au départ, reconnaissance par la plante des motifs moléculaires associés aux agents pathogènes, les PAMPS. Cela constitue la première phase de l'immunité à savoir l'immunité basale (**PTI** - PAMP-Triggered Immunity). Au cours de l'évolution, les agents pathogènes ont développé des effecteurs c'est-à-dire des molécules qui interfèrent avec la PTI permettant ainsi leur développement ce qui conduit chez la plante à une sensibilité activée par les effecteurs (**ETS**, Effector-Triggered Susceptibility). En réponse, les plantes ont développé un système de détection des effecteurs qui va activer l'immunité activée par les effecteurs (**ETI**, Effector Triggered Immunity). Les agents pathogènes peuvent acquérir d'autres effecteurs (en bleu), les plantes doivent s'adapter afin de pouvoir reconnaître ces nouveaux effecteurs.

2 Mécanismes de défense contre les maladies virales

Au départ, les virus étaient exclus de la plupart des modèles décrivant les mécanismes d'immunité chez les plantes tels que le modèle Zig Zag de Jones & Dangl puisqu'ils ne produisent pas de PAMPs ou d'effecteurs capables d'activer la PTI et l'ETI tels que décrits pour les autres agents pathogènes. Cependant de récentes études ont émis l'hypothèse que l'immunité basale pouvait également être utilisée par les plantes pour lutter contre les virus avec la mise en place de mécanismes similaires à ceux décrits pour les autres agents pathogènes (Gouveia et al., 2017). Ainsi, les mécanismes liés à l'immunité des plantes face aux infections virales sont complémentaires en terme de timing d'activation (plus ou moins tôt au cours de l'infection), de localisation (au niveau des tissus infectés ou en systémie) de cible d'action (génome ou protéines virales) (Nicaise, 2014).

2.1 RNA interference mediated-resistance

Le **RNA silencing** également appelé **RNA interférence (RNAi)** est un des premiers mécanismes de défense à se mettre en place chez les plantes qui reste très conservé chez les eucaryotes. Chez les plantes, le RNA silencing en plus de sa fonction de défense, assure la protection du génome contre des éléments génétiques mobiles comme les transposons et régule l'expression des gènes. Ce mécanisme est généralement activé par la présence de petits ARN tels que les siRNA (« short interfering RNAs ») qui sont de petites molécules d'ARN double brin générées à partir d'un double brin d'ARN long et les miRNA (microRNA) qui correspondent à des ARN simple brin codés par le génome qui participent à la régulation de l'expression des gènes. Il existe trois mécanismes de RNA signaling mis en évidence par des analyses génétiques et moléculaires (Baulcombe, 2004) (Fig. R.3) :

- Le premier mécanisme est cytoplasmique activé par la présence de siRNA. Il serait notamment mis en place au sein des cellules infectées par un virus.
- Le deuxième concerne le RNAi silencing des ARNm endogènes par les miRNA qui vont assurer la régulation de l'expression de certains gènes par une reconnaissance séquence-spécifique puis une dégradation ou l'arrêt de la traduction des ARNm ciblés

Ces deux premières voies de RNA silencing correspondent à des mécanismes de régulation post-transcriptionnelle également appelés PTGS (« post-transcriptional gene silencing »).

- Le troisième correspond à un mécanisme de régulation transcriptionnelle (TGS pour « transcriptional gene silencing »). Il est associé à la méthylation de l'ADN et à la suppression de la transcription. Un des rôles majeur de ce mécanisme serait de protéger les génomes des dommages pouvant être causés par les transposons.

2.1.1 Post-transcriptional gene silencing : un mécanisme de défense anti-viral

Ce mécanisme de RNA-silencing est déclenché par la présence d'ARN double-brins (dsRNA). En ce qui concerne les virus ces dsRNA peuvent correspondre à des intermédiaires de réplication (cas des virus à ARN dont la réplication est assurée de façon générale dans le cytoplasme grâce à une polymérase ARN dépendante), des structures secondaires intramoléculaires en forme de tige-boucle ou encore d'appariements entre transcrits sens et antisens (cas de certains virus à ADN) (Baulcombe, 2004). Lors de la première étape, les dsRNA sont reconnus par des protéines Dicer-like (DCL) dont l'activité ribonucléase de classe III va générer des siRNA de 21-24 nucléotides de longueur. Les deux brins du siRNA seront ensuite

séparés, seul l'un des deux brins sera incorporé au complexe RISC (« RNA-induced silencing complex ») où il servira de guide pour diriger le clivage séquence-spécifique des ARN viraux. Le brin complémentaire sera quant à lui dégradé. Avant la prise en charge par le complexe RISC, les siRNA seront protégés par l'ajout d'un groupement méthyle à leur extrémité 3' par la méthyltransferase HEN1 (HUA ENHANCER 1) (Baulcombe, 2004; Sharma *et al.*, 2013; Pumplin & Voinnet, 2013) (Fig. R.3c)

2.1.2 Acteurs moléculaires du PTGS

2.1.2.1 Les protéines Dicer-like (DCL)

Chez *A. thaliana*, quatre protéines DCL différentes ont été identifiées. Ainsi, la protéine DCL1 est responsable de la formation des miRNA impliqués dans la voie de régulation des gènes endogènes. La protéine DCL3 serait impliquée dans la voie de régulation transcriptionnelle. Ce sont donc les protéines DCL2 et DCL4 qui ont un rôle dans le PTGS. Elles produisent respectivement des siRNA de 22 et 21 nucléotides de longueur. L'activité anti-virale serait principalement assurée par la protéine DCL4 alors que la protéine DCL2 aurait plus un rôle de substitution en prenant le relais de l'activité de DCL4 en cas de problème (Baulcombe, 2004; Sharma *et al.*, 2013; Pumplin & Voinnet, 2013).

2.1.2.2 Argonautes : les protéines essentielles du complexe RISC

Chez *A. thaliana*, dix protéines différentes de la famille des argonautes (AGO) ont été mises en évidence. Ces protéines sont les effecteurs principaux du complexe RISC. Elles possèdent deux domaines essentiels à leur fonction : le domaine PAZ qui fixe la molécule de siRNA et le domaine PIWI responsable du clivage endonucléolytique de l'ARN complémentaire. Parmi ces 10 protéines AGO, ce sont AGO1, AGO2 et AGO7 qui ont un rôle dans le PTGS avec une activité principale pour AGO1 et un rôle de substitution en cas de problème assuré par les deux autres protéines (Baulcombe, 2004; Sharma *et al.*, 2013; Pumplin & Voinnet, 2013).

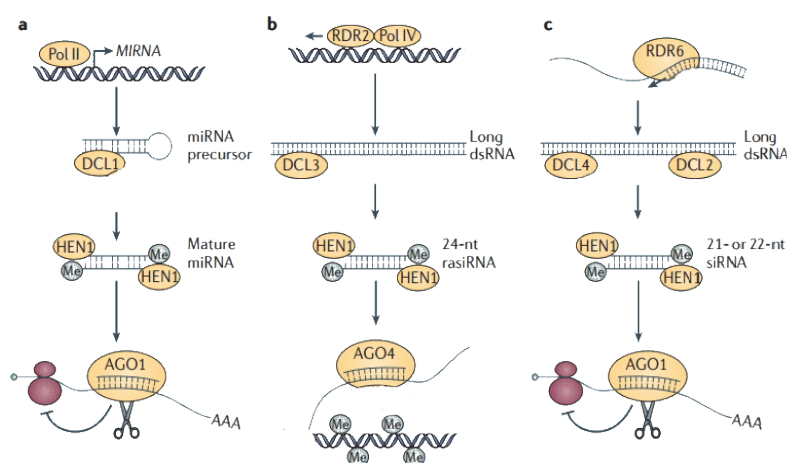


Figure R.3 | Mécanismes de RNA silencing (D'après Pumplin & Voinnet, 2013)

Les trois principaux mécanismes de RNA silencing chez les plantes sont présentés en prenant l'exemple d'*A. thaliana*. **(a)** Mécanisme impliquant les miRNA pris en charge par DCL1 puis stabilisés par HEN1 avant d'être pris en charge par le complexe RISC et en particulier par AGO1 **(b)** Mécanisme de régulation transcriptionnelle qui fait intervenir DCL3 et AGO4 conduisant à la méthylation de l'ADN **(c)** PTGS en réponse notamment à une infection virale (virus ARN) qui met en jeu DCL4/DCL2 ainsi que AGO1 permettant ainsi la dégradation ARN-spécifique de l'ARN viral.

2.1.3 Protéines virales supprimeurs du PTGS

Les virus ont mis en place des systèmes pour contre-attaquer le RNA silencing. Cela passe par la production de protéines VSR (« viral suppressor of silencing ») qui vont donc agir comme des supprimeurs de RNA silencing en interférant par différents modes d'actions à différents niveaux de la voie (Pumplin & Voinnet, 2013). Ainsi, ces protéines VSR peuvent accélérer le turn-over des siRNA en inhibant l'activité de la méthyltransférase HEN1. C'est la stratégie développée par le *Tobacco mosaic virus* (TMV) via une réplicase P126 (Fig. R.4a). Un autre mécanisme d'inhibition du RNA silencing repose sur la séquestration des siRNA empêchant ainsi son recrutement par le complexe RISC. Un exemple est la protéine P19 des tombusvirus qui se fixe spécifiquement au siRNA de 21 nucléotides produits par DLC4 (Fig. R.4b). Les VSR peuvent également cibler les protéines AGO essentielles dans le fonctionnement du complexe RISC. Ainsi, l'ubiquitination de AGO1 par les protéines P0 et P25 du *Potato virus X* (PVX) va entraîner sa dégradation par autophagie ou par l'activité de la sous unité 26 S du protéasome respectivement (Fig. R.4c)

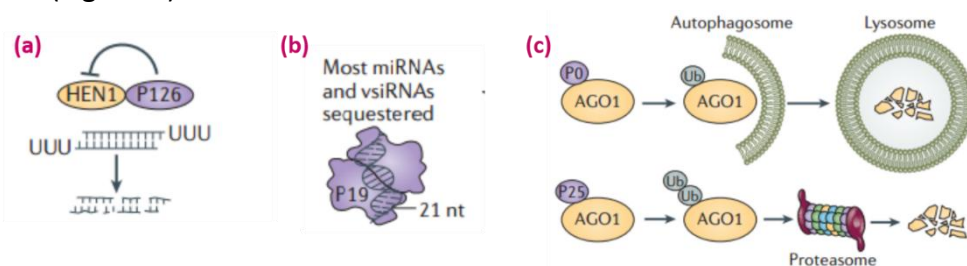


Figure R.4| Exemples de mécanismes de suppression du RNA silencing (D'après Pumplin & Voinnet, 2013)

(a) La réplicase P126 du *Tomaco mosaic virus* (TMV) inhibe l'activité méthylase de la HEN1 ce qui conduit à la dégradation des siRNA après uridylation (U) **(b)** Séquestration des siRNA de 21 nucléotides par la protéine P19 des tombusvirus **(c)** Ubiquitination de AGO1 par les protéines P0 et P25 du *Potato virus X* ce qui conduit à sa dégradation par autophagie ou via la sous-unité 26S du protéasome

2.2 Les résistances dominantes

Une des stratégies dans l'immunité des plantes largement décrite en ce qui concerne les interactions plantes/virus concerne les gènes de résistance dominante appelés **gènes R**. Ces gènes R sont impliqués dans les résistances de type gène pour gène où le produit des gènes R permettra de reconnaître de façon spécifique les protéines virales d'avirulence (*avr*) (Flor, 1971).

2.2.1 Les NB-LRR

La plupart des gènes R codent pour des protéines de la famille des **NB-LRR** (Nucleotide-binding Leucine-rich repeat). Les gènes NB-LRR constituent une famille vaste et diversifiée avec par exemple 149 gènes différents identifiés chez *Arabidopsis* (Meyers, 2003). Les protéines NB-LRR sont capables de reconnaître différents agents pathogènes (Cooley *et al.*, 2000).

2.2.1.1 Structure des protéines R

Les protéines NB-LRR sont des protéines cytoplasmiques constituées principalement de trois domaines (Moffett, 2009; de Ronde *et al.*, 2014) (Fig. R.5) :

- Au centre de la protéine, le « Nucleotide Binding Site (**NBS**) » qui est composé de trois sous-unités (NB, ARC1 et ARC2) comprenant une poche de liaison nucléotidique permettant la liaison et l'hydrolyse de l'adénosine triphosphate (ATP)

- En C-terminal, le domaine **LRR** – Leucine-rich repeat qui est impliqué dans la reconnaissance spécifique de l'agent pathogène
- En N-terminal, deux motifs différents peuvent être décrits permettant ainsi de distinguer **deux classes de protéines NB-LRR** :
 - o Le domaine **TIR** (« Toll and Interleukin-1 Receptor ») retrouvé uniquement chez angiosperm dicotylédones. Ce domaine présente une homologie avec le domaine TLR (« Toll-like receptor ») impliqué dans l'immunité innée chez les animaux
 - o Le domaine **CC** (« Coiled-Coil ») attribué notamment aux protéines NB-LRR dont le domaine en N-terminal n'a pas été encore décrit.

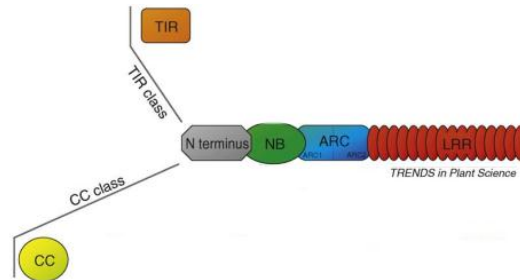


Figure R.5] Représentation schématique de l'architecture linéaire des protéines NB-LRR (D'après Collier & Moffert, 2009)
 Les différents domaines des protéines NB-LRR sont représentés : le domaine LRR (Leucine-rich repeat et le domaine NB (« Nucleotide binding »). Entre ces deux domaines se trouve le domaine ARC (Apaf-1, R proteins, CED-4) qui peut être divisé en deux sous-unités : ARC1 et ARC2. En N-terminal, deux domaines différents constituent deux catégories de protéines NB-LRR : le domaine TIR (« Toll and Interleukine-1 Receptor homology ») et le domaine CC (« Coiled-Coil »).

2.2.1.2 Modèles d'interaction avec les protéines virales

En réponse aux infections virales, 22 gènes de la famille des NB-LRR ont été clonés, pour certains d'entre eux les protéines virales d'avirulence *Avr* ont aussi été identifiées. La majorité de ces gènes codent pour des protéines de type CC-NB-LRR (de Ronde *et al.*, 2014).

Les interactions entre les protéines *R* et les protéines virales d'avirulence *Avr* peuvent s'effectuer de façon directe, il s'agit dans ce cas du modèle d'interaction Ligand-Récepteur. Cependant l'interaction protéines *R/Avr* s'effectue majoritairement de façon indirecte faisant intervenir différents facteurs de l'hôte (Moffett, 2009; Dodds & Rathjen, 2010; de Ronde *et al.*, 2014).

a. Interaction directe : 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* (Fig. R.6a)

b. Interaction indirecte

Il s'agit du modèle le mieux décrit qui fait intervenir une protéine intermédiaire de l'hôte qui sera reconnue par la protéine *Avr*. Cette interaction induit un changement de conformation dans la protéine intermédiaire de l'hôte qui permet la reconnaissance par la protéine NB-LRR. Trois modèles différents ont été décrits sur la base de ce fonctionnement (Moffett, 2009; Collier & Moffett, 2009; Dodds & Rathjen, 2010; de Ronde *et al.*, 2014) :

- **Le modèle de garde** : Dans ce modèle la protéine NB-LRR interagit avec une autre protéine, la protéine gardée, qui lorsqu'elle sera liée à la protéine *Avr* permettra d'activer la protéine NB-LRR activant de ce fait les mécanismes de défense (Fig. R.6b).

- **Le modèle « decoy »** : Dans ce modèle, dérivé du modèle précédent, le facteur de l'hôte qui interagit avec la protéine *Avr* agit comme un leurre (« decoy ») puisque son unique rôle est d'être une molécule perceptrice de la protéine *Avr* (Fig. R.6b).

Le modèle de garde et « decoy » sont difficiles à différencier puisque plusieurs protéines qui interagiraient avec des protéines d'avorulence pourraient correspondre à ces modèles d'interaction. Un exemple est 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 codée le gène *N* du tabac. L'activation du gène *N* nécessite la pré-reconnaissance du complexe NRIP1/protéine d'avorulence p50 (Caplan *et al.*, 2008).

- **Le modèle « bait-and-switch »** : Ce modèle suppose qu'en l'absence de reconnaissance d'un agent pathogène les protéines NB-LRR seraient dans un état « off » avec une conformation repliée de telle façon que le domaine NB-ARC est en relation avec les domaines N-terminal et LRR. La protéine qui sert d'appât (« bait ») est quant à elle liée au domaine N-terminal. Lorsque cette protéine appât interagit avec la protéine *Avr* il y a un changement de conformation du complexe (« on ») permettant l'interaction de la protéine *Avr* avec le domaine LRR ce qui conduit à la transduction du signal de résistance (Fig. R.6c et Fig. R.7). Un exemple est le gène *Rx* mis en évidence chez la pomme de terre en réponse au *Potato virus X*. Des interactions intramoléculaires maintiennent le produit du gène *Rx* dans un état inactif. L'interaction avec la protéine virale effectrice provoque des changements de conformations qui activent les mécanismes de résistance en aval (Slootweg *et al.*, 2013).

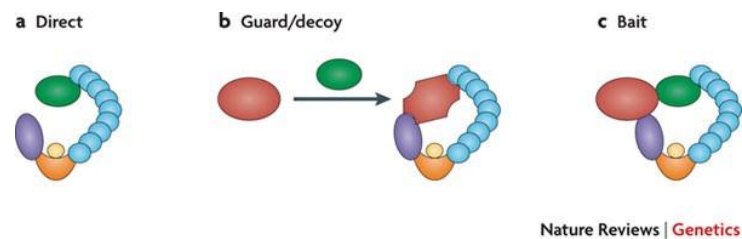


Figure R.6 | Modèles de reconnaissance directe-indirect des NB-LRR (Dodds & Rathjen, 2010)

Les protéines NB-LRR interagissent avec les protéines d'avorulence (*Avr*) de façon directe ou indirecte. **a)** Interaction directe, la protéine *Avr* (en vert) interagit avec le domaine LRR de la protéine R (en bleu) ; **b)** Modèles de garde et « decoy » où la protéine *Avr* (en vert) modifie une protéine accessoire de l'hôte (en rouge) qui peut être sa cible de virulence (modèle de garde) ou une imitation structurelle d'une telle cible (modèle « decoy ») ; **c)** Modèle « bait-and-switch » où l'interaction d'un effecteur avec une protéine accessoire de l'hôte (en rouge) facilite la reconnaissance directe par la protéine NB-LRR. Les protéines NB-LRR représentées ci-dessus sont constituées d'un domaine C-terminal LRR (en bleu), d'un domaine central NB (croissant orange) qui lie un ATP ou ADP (rond jaune) et d'un domaine N-terminal (ovale violet) de type TIR (Toll, interleukin-1 receptor) ou CC (coiled-coil)

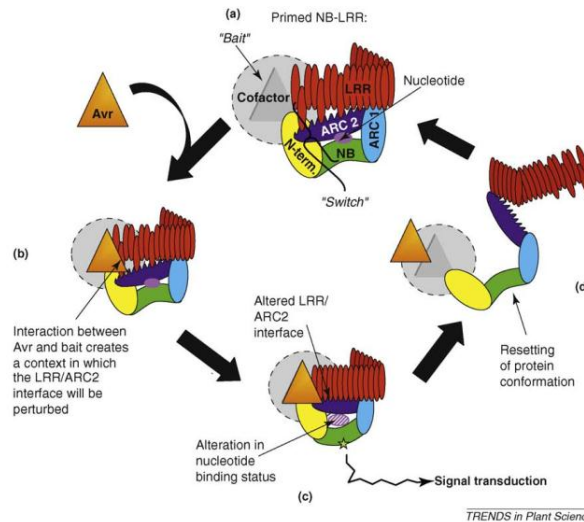


Figure R.7 | Modèle "bait-and-switch" d'interaction des NB-LRR (D'après Collier & Moffett, 2009)

a) La protéine NB-LRR est auto-inhibée (état « off ») dans une configuration telle que le domaine NB-ARC est en relation avec les domaines N-terminal et LRR ; b) La protéine Avr (triangle orange) interagit avec la protéine appât (« bait ») du complexe (en gris) ; c) Cette interaction induit un changement de conformation du complexe (état « on ») qui conduit à une activation du signal de défense ; d) réinitialisation de la conformation de la protéine.

2.2.1.3 Cascades de signalisation et réponses de défense

La première réponse classiquement attribuée à la résistance en lien avec les gènes R est la **réponse hypersensible (HR)**. Cette réponse entraîne une activation de la mort cellulaire programmée afin de limiter la propagation de l'infection. La HR est généralement liée à l'accumulation de l'acide salicylique (SA), jasmonique (JA), d'éthylène, de ROS et de calcium (Ca^{2+}) et est aussi liée à l'expression de gènes codant pour des PR-protéines (« Pathogenesis Related proteins ») (de Ronde *et al.*, 2014). En parallèle à cette réponse de défense localisée au niveau des sites d'infection, un signal de défense dans des tissus adjacents est également induit par les gènes R, il s'agit de la **résistance systémique acquise (SAR)**. La SAR est liée à la voie du SA qui agit comme une molécule signal. Cette réponse se traduit par l'expression systémique de gènes codant pour les PR-protéines. La SAR perdure dans le temps et peut permettre ainsi d'assurer l'immunité des plantes vis-à-vis d'autres infections (Soosaar *et al.*, 2005; de Ronde *et al.*, 2014).

2.2.2 Autres résistances dominantes

Les **gènes RTM (« Restricted TEV Movement »)** ont été les premiers gènes à résistance dominante clonés et impliqués dans la réponse aux virus n'appartenant pas à la famille des NB-LRR. Ces gènes sont impliqués dans la réponse aux potyvirus tels que le *Tobacco etch potyvirus* (TEV), au *Lettuce mosaic virus* (LMV) et au *Plum pox virus* (PPV). La caractérisation génétique de différentes accessions d'*Arabidopsis* a permis de mettre en évidence trois gènes et deux loci à résistance dominante : *RTM1*, *RTM2*, *RTM3*, *RTM4* et *RTM5*. Le gène *RTM1* code pour une protéine appartenant à une grande famille caractérisée par une sous-unité de lectine : les jacalines. Certaines protéines de cette famille ont été identifiées dans la réponse à des agents pathogènes ou des ravageurs (Chisholm *et al.*, 2000). Le gène *RTM2* code pour une protéine avec des domaines transmembranaires présentant des homologues avec les protéines heat shock. Cependant, *RTM2* n'est pas inductible par la chaleur et ne contribue pas à la thermo-tolérance. *RTM3* code pour une protéine présentant en N-terminal un domaine

MATH-TRAF (meprin and TRAF homology) et en C-terminal un domaine coiled-coil. Les loci *RTM4* et *RTM5* ont uniquement été génétiquement caractérisés (Cosson *et al.*, 2012).

Le fait que l'expression des gènes *RTM1* et *RTM2* soit limitée au compartiment phloémien et qu'une interaction entre *RTM1* et *RTM3* ait été mise en évidence en double-hybride et *in planta* ont permis d'émettre l'hypothèse sur la formation par les produits des gènes *RTM* d'un complexe multi-protéique. Ce complexe serait impliqué dans des mécanismes de résistance qui bloqueraient le mouvement en systémie des potyvirus (Cosson *et al.*, 2010).

Un autre gène appartenant à la même famille que *RTM1* a également été identifié en réponse à une infection virale. Il s'agit du gène ***JAX1 (JACALIN TYPE LECTIN REQUIRED FOR POTEXVIRUS RESISTANCE 1)*** qui confère la résistance à un potexvirus *Plantago asiatica mosaic virus* (PIAMV). Cependant, contrairement au mode d'action de *RTM1*, *JAX-1* semble conférer la résistance au PIAMV en agissant lors des premières étapes de l'infection. L'action de ces deux gènes montre le rôle important des lectines dans les résistances aux virus (Yamaji *et al.*, 2012).

Chez une espèce de tomate sauvage *Solanum hirsutum* le gène ***Tm-1*** a été mis en évidence en réponse au *Tomato mosaic virus* (ToMV). Ce gène code pour une protéine qui présente un domaine TIM-barrel. C'est d'ailleurs ce domaine qui assure la liaison aux protéines de réplication du ToMV inhibant de ce fait la réplication de son ARN (Ishibashi *et al.*, 2007; Ishibashi & Ishikawa, 2013).

De façon générale, ces gènes R n'appartenant pas à la famille des NB-LRR semblent intervenir à des étapes clés de l'infection virale ce qui permet de bloquer avec succès le développement de la maladie.

2.2.3 RNA silencing et résistances dominantes : « Le » modèle Zig Zag en réponse aux phytovirus

Le modèle Zig Zag illustre bien « la course à l'armement » entre les plantes et les agents pathogènes excluant au départ les virus de ce modèle. Cependant, les agents pathogènes viraux et non-viraux induisent des réponses immunitaires semblables telles que l'induction de la HR et la SAR dans les tissus non infectés. De plus, les virus produisent des protéines d'avirulence *Avr* (protéines de la capsid, de mouvement, de réplication) reconnues par les gènes R de type NB-LRR (Zvereva & Pooggin, 2012). Un modèle de l'immunité des plantes face au virus a ainsi été décrit par Nakahara & Masuta (2014) en faisant le parallèle avec le modèle Zig Zag de Jones & Dangl (2006). Dans ce modèle, les virus produiraient des dsRNA qui agiraient comme des PAMPs (la notion de VAMP, « virus associated molecular pattern » a d'ailleurs été avancée par Pumplin & Voinnet, 2013). Les plantes activeraient de ce fait le RNA silencing qui correspondrait à une phase PTI-like. Les virus auraient la capacité de contre-attaquer en produisant des RNA silencing suppressor (RSS). Finalement, comme dans une phase ETI-like, les plantes activeraient des protéines R qui reconnaîtraient spécifiquement les RSS produits comme des protéines *Avr*. Cette reconnaissance amenant à l'activation des mécanismes de défense avec notamment la HR et la SAR (Fig. R.8)

Cependant, il est important de noter qu'une particularité propre aux virus crée une réelle différence dans ces mécanismes d'immunité par rapport aux autres agents pathogènes. En effet, contrairement aux champignons ou bactéries qui possèdent différentes protéines d'avirulence, les virus présentent un nombre de protéines limitées qui sont toutes essentielles à la réalisation de leur cycle. Ainsi quand une protéine R reconnaît une protéine virale *Avr*, les virus auront plus de difficultés à pouvoir la remplacer. Cela va avoir un impact sur le nombre de cycles de la phase ETI qui seront moins nombreux par rapport aux autres agents pathogènes (Nakahara & Masuta, 2014).

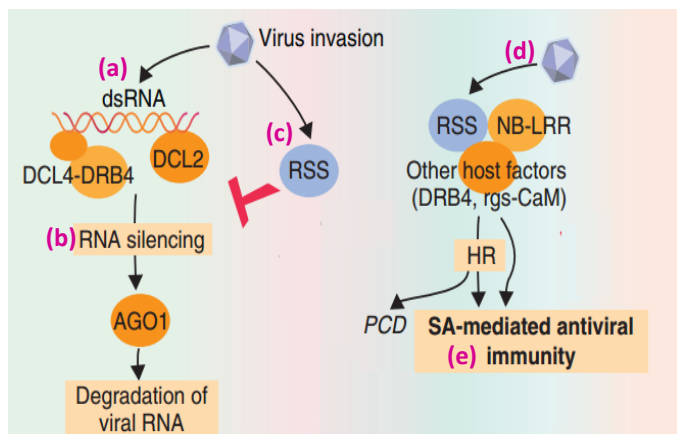


Figure R.8| Modèle des interactions plante/virus impliquant le RNA silencing et les protéines R (D'après Nakahara & Masuta, 2014)

Le RNA silencing (b) est induit par les dsRNA viraux (a) (génomme viral ou transcrits). Ces dsRNA sont pris en charge par le complexe (ici chez *Arabidopsis*) DCL4-DRB4 et DCL2. AGO1 va permettre la dégradation des ARN viraux. Cependant, les virus ont la capacité de contre-attaquer en produisant des RNA silencing suppressors (RSS) (c). Les plantes en réponse vont activer les protéines R qui vont reconnaître ces RSS tels des protéines *Avr* (d). Cela va donc déclencher les mécanismes de défense : HR et SAR (e)

2.3 Résistances récessives

Les résistances récessives se basent sur les interactions moléculaires qui se mettent en place entre le virus et sa plante hôte. En effet, les virus en tant que parasites obligatoires détournent la machinerie cellulaire de leur hôte afin de pouvoir se développer (Wang, 2015). Ainsi l'hypothèse développée par Ronal Fraser largement admise consiste à dire que ces résistances récessives seraient liées à des mutations au niveau des gènes codant pour ces facteurs de l'hôte normalement nécessaires au bon déroulement du cycle viral (Fraser & Van Loon, 1986) (Fig. R.9). En comparaison avec les autres agents pathogènes, les gènes de résistance récessive impliqués dans la réponse des plantes aux virus sont majoritaires. Ainsi, environ 50% des gènes de résistance identifiés seraient des gènes de résistance récessive (Revers & Nicaise, 2014).

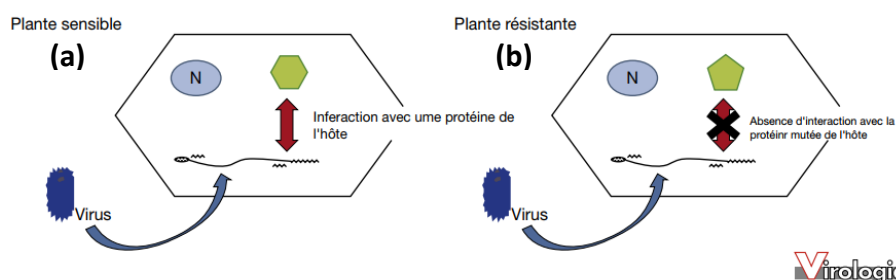
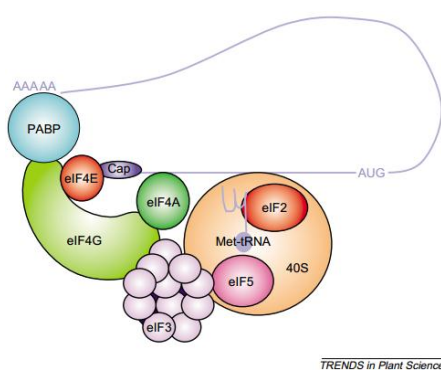


Figure R.9| Mécanisme de résistance récessif aux virus (D'après Lecampion *et al.*, 2009)

(a) Le virus interagit avec une protéine de l'hôte (en vert) afin de réaliser son cycle de réplication, la plante est donc sensible (b) La protéine de l'hôte est mutée, l'interaction n'est donc plus possible, le cycle de réplication viral est arrêté, la plante est donc résistante.

2.3.1.1 Les facteurs d'initiation de la traduction : des déterminants majeurs dans la réponse aux virus

Chez les eucaryotes, la traduction de l'ARNm peut être divisée en trois grandes étapes : initiation, élongation et terminaison. La phase d'initiation constitue une étape clé dans la traduction de l'ARNm faisant intervenir un certain nombre d'acteurs moléculaires constituant le complexe d'initiation de la traduction. L'assemblage de ce complexe va permettre le recrutement du ribosome qui permettra d'enclencher la synthèse protéique. Le **facteur d'initiation de la traduction eIF4E** est le premier acteur moléculaire à intervenir en se fixant en 5' à la coiffe de l'ARNm. Cela permet le recrutement du facteur de la traduction **eIF4G**. L'interaction entre ces deux eIFs constitue le **complexe eIF4F**. L'eIF4G va assurer la liaison avec d'autres acteurs moléculaires tels que la protéine PABP (« Poly A Binding Protein ») ce qui assure la circularisation de l'ARNm qui est essentielle à l'efficacité de la traduction. L'ensemble de ce complexe permet le recrutement de la sous-unité 40 S du ribosome (liée à l'ARN de transfert associé à la méthionine) (Fig. R.10) (Robaglia & Caranta, 2006).



TRENDS in Plant Science

Figure R.10| Complexe d'initiation de la traduction (D'après Robaglia & Caranta, 2006)

Le facteur eIF4E se lie en 5' à la coiffe de l'ARNm (« cap »). En recrutant le facteur eIF4G ils forment le complexe eIF4F qui se lie à la protéine PABP assurant ainsi la circularisation de l'ARN. D'autres acteurs moléculaires interviennent par la suite tel qu'eIF3 (se lie à eIF4G) qui permet le recrutement de la sous-unité 40 S du ribosome. Cette sous-unité forme un complexe ternaire avec eIF2 et l'ARNt associé à la méthionine. Enfin, eIF5 permet le recrutement de la sous-unité 60 S du ribosome ce qui permet d'initier la phase d'élongation de la traduction.

Une étape critique dans le cycle viral concerne la traduction de l'ARN viral. Ainsi, le recrutement de facteurs d'initiation de la traduction semble indispensable pour assurer la réplication virale. Il n'est donc pas surprenant qu'un grand nombre de gènes de résistance récessive caractérisés codent pour des facteurs d'initiation (eIFs) appartenant aux familles des eIF4E et eIF4G ainsi que leurs isoformes (Robaglia & Caranta, 2006; Revers & Nicaise, 2014; Nicaise, 2014; Hashimoto *et al.*, 2016).

La première étude impliquant les protéines eIF4E dans réplication virale concernait un crible double hybride chez la levure dans le cadre de la recherche d'interacteurs de la protéine virale VPg chez le *Turnip mosaic virus* (TuMV) (Wittmann *et al.*, 1997). Ensuite une étude impliquant mutant perte de fonction d'*A. thaliana* du gène *eIF(iso)4E* montrait que cette mutation lui conférait une résistance vis-à-vis du *Tobacco etch virus* (TEV) (Lellis *et al.*, 2002). Le rôle des gènes eIF4E dans la résistance naturelle aux potyvirus a été confirmé par le clonage du gène récessif *pvr2* chez le piment (*Capsicum* spp.) assurant une résistance totale au *Potato virus Y* (PVY) et au TEV (Ruffel *et al.*, 2002). Ainsi, initialement mis en évidence dans le cadre d'infection par un potyvirus le rôle des eIFs a rapidement été relié en réponse à d'autres familles de virus telles que les bymovirus, cucumovirus, ipomovirus, sobemovirus, carmovirus et waikivirus. Ces eIFs pourraient donc contribuer à un vaste mécanisme de sensibilité des plantes aux virus (Revers & Nicaise, 2014).

2.3.1.2 Recherche d'autres gènes de résistance récessive

D'autres gènes de résistance récessive ne codant pas pour des facteurs d'initiation de la traduction ont également été mis en évidence via notamment l'analyse de collections de mutants et/ou en exploitant la diversité des plantes cultivées/sauvages (Revers & Nicaise, 2014). Dans la cette recherche de nouveaux facteurs de l'hôte potentiellement impliqués dans la réponse à des virus, la plante modèle *A. thaliana* a largement été utilisée en tant que source de nouvelles résistances récessives (Ouibrahim & Caranta, 2013).

Ainsi, les loci *rlm1* et *rpv1* qui confèrent respectivement une résistance au LMV et au PPV ont été identifiés dans l'accession Cvi-0 (*Cape Verde Islands*) au sein de régions génétiques ne comprenant pas de gènes de type eIF (Revers *et al.*, 2003; Decroocq *et al.*, 2006). De manière plus récente, le gène *rwm1* qui code pour une phosphoglycerate kinase chloroplastique (cPGK2) a été caractérisé en réponse au *Watermelon mosaic virus* (WMV). La résistance à ce potyvirus est liée à une seule substitution d'un acide aminé au niveau du domaine N-terminal de cPGK2 (Ouibrahim *et al.*, 2014). D'ailleurs dans une approche gène candidat en réponse au PPV, cPGK2 semble être un candidat potentiel de la réponse à ce potyvirus (Poque *et al.*, 2015). Le screening d'une large population EMS d'*A. thaliana* a permis d'identifier les gènes codant pour des protéines membranaires **TOM1**, son homologue **TOM3** et **TOM2A** impliquées dans la formation des complexes de réplication du *Tobacco mosaic virus* (TMV) (Yamanaka *et al.*, 2002).

Si l'ensemble de ces gènes identifiées semblent coder pour des protéines intervenant dans des phases précoces de l'infection virale, le locus *sha3* serait quant à lui impliqué dans la restriction du mouvement en systémie du PPV (Pagny *et al.*, 2012). D'ailleurs, ce locus co-localise avec le gène *RTM3* décrit précédemment. Le transport vasculaire du *Potato virus A* (PVA) serait lui bloqué par le produit du gène *ra*. Sa caractérisation génétique dans *Solanum tuberosum* montre qu'il ferait partie d'un cluster de gènes incluant les gènes *Ry* et *Na* qui sont eux des gènes de résistance dominante (Hämäläinen *et al.*, 2000).

L'utilisation de la variabilité naturelle des espèces cultivées a permis chez l'orge, par des stratégies de clonage positionnel, de caractériser le gène *rym11*. Ce gène code pour la protéine disulfide isomerase PDIL5-1 impliquée dans la résistance récessive contre des bymovirus (Yang *et al.*, 2014). De même, chez le riz Africain (*O. glaberrima*) le gène *RYMV2* a été identifié en réponse au *Rice yellow mottle virus* (RYMV) (Thiémélé *et al.*, 2010; Orjuela *et al.*, 2013). La description de ces différents gènes de résistance récessive identifiés chez des plantes cultivées ne regroupe évidemment pas l'ensemble des facteurs de l'hôte identifiés à ce jour en réponse aux phytovirus. La revue scientifique publiée par Hashimoto et ses collaborateurs (2016) fait état des connaissances actuelles relatives aux résistances récessives en réponse à des virus (pour revue Hashimoto *et al.*, 2016).

2.4 Régulation hormonale et immunité

Les phytohormones ont largement été caractérisées pour leur rôle à la fois dans les mécanismes de développement mais également dans leur implication dans la réponse à différents stress. Certaines de ces phytohormones ont déjà largement été caractérisées dans la réponse immunitaire des plantes (éthylène, acide jasmonique et l'acide salicylique). De nouvelles études montrent que certaines des phytohormones pourraient également avoir un rôle dans les mécanismes de défense alors qu'elles étaient au départ associées aux

mécanismes de développement et de croissance (auxine, acide abscissique, les cytokinines, les gibbérillines et les brassinostéroïdes) (Denancé *et al.*, 2013).

L'**acide salicylique (SA)** est impliqué dans la réponse induite par les mécanismes de défense déclenchés par les protéines R, la SAR. Cette biosynthèse et accumulation de SA entraîne la production au niveau des sites de l'infection d'espèces réactives de l'oxygène, l'expression de protéines PR, le dépôt de callose et l'induction de la HR. Le SA joue un rôle également dans la mise en place du signal de défense en systémie. De façon générale, le SA est impliqué dans les mécanismes de défense contre des agents pathogènes biotiques tels que les virus (Collum & Culver, 2016).

Comme pour le SA, le rôle de l'**éthylène** dans les mécanismes de défense anti-viraux a déjà largement été décrit (Fischer & Dröge-Lasser, 2004 ; Love *et al.*, 2005-2007).

L'**acide jasmonique (JA)** est souvent décrit comme impliqué dans la réponse à des agents pathogènes nécrotrophes ou à des ravageurs (Collum & Culver, 2016). Ce qui fait qu'en réponse aux virus, son action est plus controversée. En effet, lors d'une infection par des geminivirus, les gènes de la voie du JA sont réprimés alors que cette même voie est stimulée chez des tabacs transgéniques exprimant un suppresseur viral du *African cassava mosaic virus* (Calil & Fontes, 2016).

Certaines études ont mis en évidence l'action des **brassinostéroïdes (BR)** dans la réponse aux virus. Ainsi, des tabacs traités avec du brassinolide (premier BR avec une activité biologique isolé) présentent une meilleure résistance au TMV. L'analyse plus fine des mécanismes de réponse sous-jacents laisse penser que la réponse médiée par les BRs est différente des mécanismes mis en place par le SA. En effet, il n'y a pas d'induction des voies du SA, ni expression de PR protéines (Nakashita *et al.*, 2003 ; Calil & Fontes, 2016).

L'**acide abscissique (ABA)** a largement été étudié dans la réponse à des stress abiotiques et biotiques. Son implication dans le cas d'infections virales a été par exemple mise en évidence chez des plantes de tomates résistantes (gène *Tm-1*) dont les concentrations en ABA sont supérieures aux plantes sensibles. D'ailleurs une application exogène d'ABA réduit l'accumulation en systémie du TMV. Il semble également impliqué dans le blocage du mouvement de cellule à cellule du *Bamboo mosaic potexvirus* (BamV) en contribuant à augmenter le dépôt de callose au niveau des plasmodesmes (Calil & Fontes, 2016).

Enfin, les virus pourraient également avoir la capacité d'agir sur les voies des **auxines**, **cytokinines** et **gibbérellines**. Ainsi, par exemple les gibbérellines pourraient contribuer aux mécanismes de défense en modulant la balance entre SA et JA en fonction de l'agent pathogène présent (biotrophe vs. nécrotrophe).

Cette revue rapide concernant l'implication des phytohormones dans la réponse à des virus permet de mettre l'accent sur le rôle crucial de ces molécules qui doivent en parallèle assurer la croissance et le développement de la plante. Il existe des relations complexes entre ces différentes phytohormones. Les mécanismes de défense hormonaux sont donc le résultat d'interactions et régulations complexes entre ces phytohormones (Verma *et al.*, 2016; Berens *et al.*, 2017). D'ailleurs, leur implication semble dépendre fortement du type de pathosystème

analysé. De part cette complexité, il reste donc encore beaucoup de questions à résoudre afin de comprendre les mécanismes hormonaux mis en place par la plante lors d'une infection virale (Calil & Fontes, 2016).

3 Contournement et Durabilité des résistances

3.1 Le contournement des résistances

Les virus sont des parasites obligatoires caractérisés par leur réplication rapide, leur population de grande taille et leur taux de mutation élevé qui ne comprend pas de système de réparation ($\sim 10^{-3}$ à 10^{-5} mutations par nucléotide et par cycle de réplication). Tout cela leur confère une grande variabilité génétique qui leur permet de s'adapter rapidement aux différentes contraintes de leurs hôtes et donc de pouvoir contourner les résistances mises en place dans les plantes cultivées (Montarry *et al.*, 2012).

Pour qu'une résistance soit contournée, il faut l'apparition, dans la population virale, d'un variant virulent. Ce sont des substitutions nucléotidiques au sein du gène d'avirulence qui entraînent un changement d'acide aminé à l'origine du variant virulent. Ces variants virulents vont ensuite devoir s'accumuler dans la plante, en compétition avec les variants avirulents. Cette accumulation des variants virulents au sein de la population virale va dépendre de leur valeur adaptative relative et/ou de leur capacité à s'accumuler dans la plante et à être transmis de plante à plante. En effet, les gènes viraux sont multifonctionnels, par conséquent des substitutions nucléotidiques peuvent avoir un effet pléiotropique important : le gain de virulence peut entraîner un coût important pour le virus, le variant virulent ne sera donc pas conservé dans la population. De plus, la transmission puis le mouvement du virus dans la plante constituent des événements de « bottleneck » forts, à l'origine de la dérive génétique. Ce phénomène aura de ce fait un impact important dans la probabilité de dissémination du variant virulent. Cela sera également à l'origine d'une ségrégation spatiale des différents variants du virus qui s'accumuleront de façon différentielle dans la plante (Fig. R.11) (Gómez *et al.*, 2009)

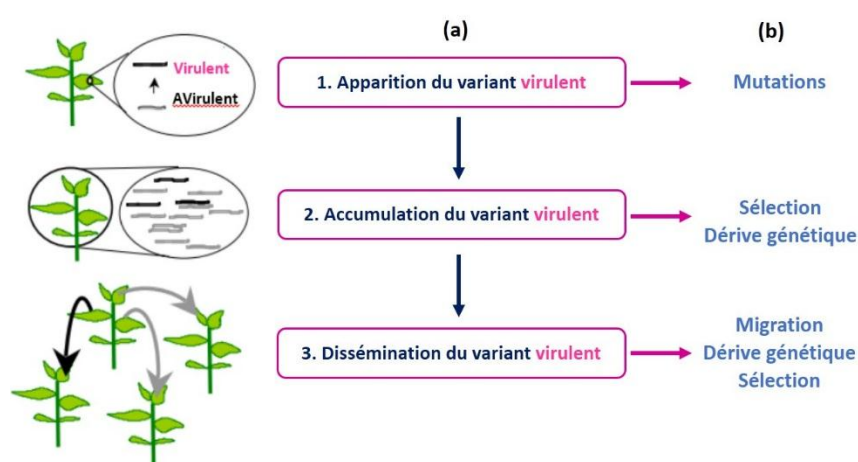


Figure R.11 | Mécanismes du contournement des résistances par les virus (D'après Gomez *et al.*, 2009)
(a) Principales étapes que le virus doit franchir afin de pouvoir contourner une résistance **(b)** Forces évolutives impliquées à chaque étape du contournement

3.2 Durabilité des résistances

D'après Johnson (1981), une résistance est dite durable si elle a conservé son efficacité après avoir été utilisée de manière prolongée sur de grandes surfaces et en présence de l'agent pathogène. De part cette définition, la durabilité d'une résistance ne peut être mesurée qu'*a posteriori*. L'idéal serait de pouvoir la prédire par avance ce qui implique de connaître les facteurs qui sont susceptibles d'avoir un impact sur cette durabilité. Dans une étude menée par McDonald & Linde (2002) sur un large panel d'agents pathogènes, le facteur déterminant de la durabilité des résistances était le potentiel d'évolution de la population de l'agent pathogène. Ce potentiel d'évolution des populations serait également un élément clé dans la durabilité des résistances aux virus (García-Arenal & McDonald, 2003). Ainsi, en tenant compte d'un certain nombre de traits viraux (taille des populations, capacité de dissémination, mode de transmission, taux de recombinaison) un indice EP pour « Evolutionary Potential » qui corrèle avec la durabilité des résistances a été créé. Plus cet indice est élevé, moins la durabilité de la résistance est importante (García-Arenal & McDonald, 2003; Gómez *et al.*, 2009).

Différents facteurs vont avoir un impact sur la durabilité des résistances, sur lesquels il sera possible d'agir afin de l'améliorer. Un de ces facteurs concerne la nature des résistances utilisées. En effet, plus de 80% des résistances aux virus décrites sont monogéniques. Or, en tenant compte des mécanismes que les virus doivent mettre en place pour assurer le contournement d'une résistance, il semblerait logique de dire que plus les virus ont de modifications à réaliser pour devenir virulents, plus la résistance sera durable. En ce sens, les **résistances quantitatives** seraient plus durables en comparaison aux résistances qualitatives largement développées jusqu'à présent car plus faciles à manipuler en sélection (Parlevliet, 2002). Il a également été montré que le fond génétique des plantes est important et va donc avoir un effet sur la durabilité des résistances. Ainsi, dans l'étude de l'interaction piment/*Potato virus Y (PVY)*, la durabilité du gène majeur *pvr2³* est améliorée lorsque ce gène est introduit dans un fond génétique partiellement résistant par rapport à un fond génétique sensible (Palloix *et al.*, 2009).

Le concept de « **gene pyramiding** » est en lien avec cette idée de freiner l'apparition dans la population virale de nouveaux variants virulent. Cette stratégie consiste à associer différents types de résistances qui vont cibler différentes protéines virales à différents stades du cycle viral. Cette stratégie a déjà fait ses preuves dans la réponse à des virus que ce soit avec des gènes majeurs (Werner *et al.*, 2005) ou par combinaison d'un gène majeur et de plusieurs résistances quantitatives (Palloix *et al.*, 2009; Quenouille *et al.*, 2016).

L'environnement ainsi que les pratiques culturales constituent d'autres facteurs pouvant avoir un impact sur la durabilité des résistances. Différentes stratégies de gestion peuvent donc être déployées dans le temps et dans l'espace afin d'assurer par ce biais la durabilité des résistances. Par exemple, la rotation des gènes de résistance se base sur une utilisation rationalisée des gènes de résistance de manière à contrôler la pression de sélection exercée par ces gènes sur les populations virales afin d'éviter la fixation dans ces populations de mutations de contournement (Mundt, 2014). Si ce type de stratégies s'applique plutôt à l'échelle d'une parcelle, de plus en plus d'études s'intéressent au développement de

stratégies qui permettrait d'assurer la durabilité des résistances à l'échelle du paysage (Fabre *et al.*, 2012).

Enfin, le développement de nouvelles approches scientifiques et technologiques va permettre de contribuer à comprendre d'avantage les interactions plantes/virus ce qui permettra par la suite de développer des stratégies optimales permettant de garantir la durabilité d'une résistance (Gómez *et al.*, 2009; Mundt, 2014).

B. Interactions plantes/agents pathogènes en conditions de stress multiples

1 Les plantes face aux stress multiples

La réponse des plantes face à des stress multiples est complexe : des mécanismes de réponse spécifiques à chaque stress mais également des réponses combinées aux différents stress sont mis en jeu. Deux grandes catégories de stress sont décrites : les **stress abiotiques** et les **stress biotiques**. Ainsi, les principaux stress abiotiques déjà largement caractérisés de par leur fort impact sur les cultures sont la sécheresse, la salinité, la chaleur, le froid, le gel, les fortes intensités lumineuses, l'ozone (O₃) et les stress anaérobies. De façon générale, les études menées sur ces stress ont été effectuées en conditions contrôlées. Cependant, ces conditions de laboratoire sont très différentes de celles observées dans la nature ou au champ où les plantes en culture doivent continuellement faire face à de nombreux stress en même temps. La combinaison de ces stress est souvent additive et peut avoir des conséquences bien plus importantes que dans le cas d'un simple stress (Fig. R.12) (Suzuki *et al.*, 2014). D'ailleurs il a été montré que la réponse des plantes à une combinaison de stress ne peut pas être prédite à partir de la réponse caractérisée dans les études menées sur chacun de ces stress de façon indépendante (Rasmussen *et al.*, 2013; Prasch & Sonnewald, 2013).

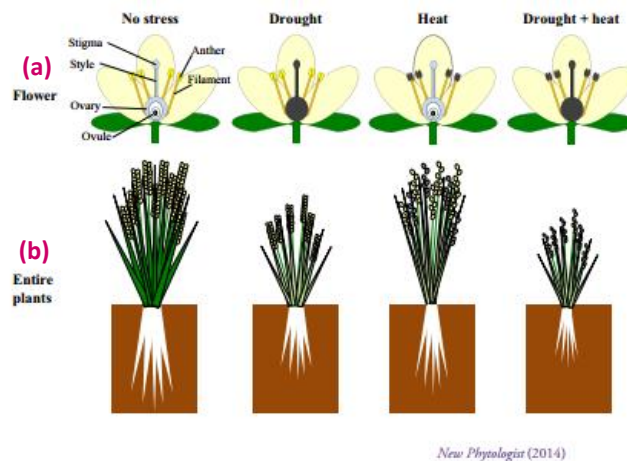
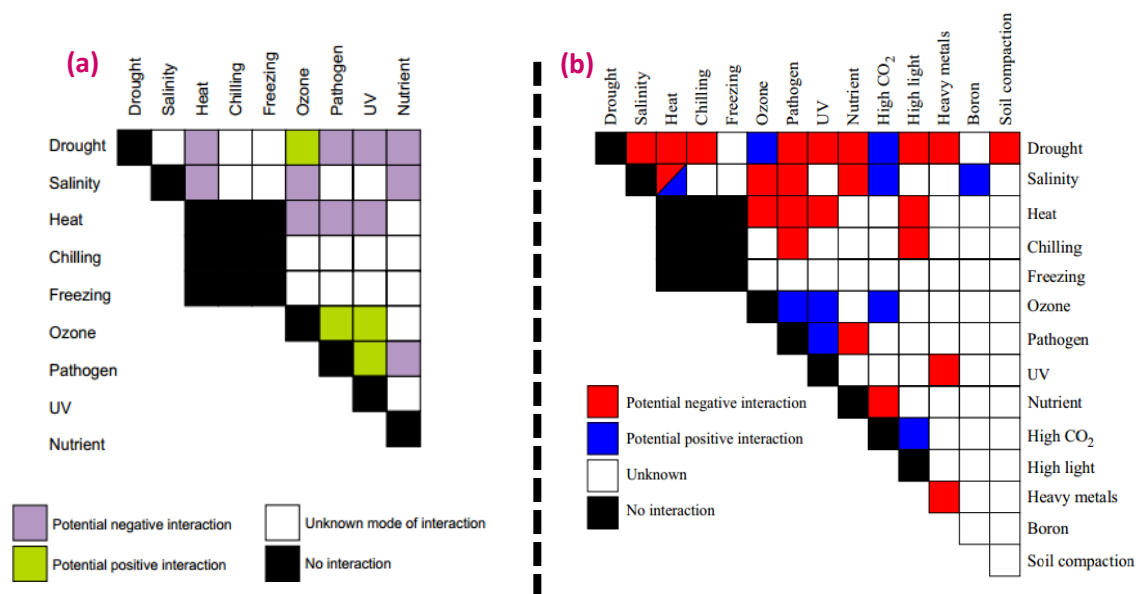


Figure R.12 | Effets de la combinaison de stress en comparaison à des stress uniques (D'après Suzuki *et al.*, 2014)
Exemple de l'effet de la sécheresse et de la chaleur sur le développement du maïs et du blé. **(a)** Sur les organes reproducteurs, la sécheresse a un impact sur le développement des pistils alors que c'est le développement du pollen qui est affecté par la chaleur (organes affectés marqué en noir) **(b)** La sécheresse entraîne une diminution de la taille, du nombre d'épis et du poids des grains/épis. La chaleur va entraîner une augmentation du nombre d'épis avortés (indiqués en noir) et une diminution du nombre de grains/épis. Les effets liés à ces deux stress sont retrouvés de façon simultanée lorsqu'ils sont combinés.

En plus des conséquences importantes qu'ils entraînent ces stress abiotiques vont également avoir un impact sur les aires de répartition des maladies ainsi que sur le niveau d'agressivité et de pathogénicité des agents pathogènes entraînant ainsi l'apparition de maladies émergentes. De plus, ces différents stress abiotiques en altérant la physiologie des plantes vont perturber leurs interactions avec les agents pathogènes (Pandey *et al.*, 2017). Il est donc important de comprendre les mécanismes de réponse des plantes soumises à des stress multiples afin de pouvoir développer des variétés avec un large spectre de résistance.

D'ailleurs des variétés, créées dans le cadre de programmes de sélection, résistantes à un seul stress sans avoir testé leur sensibilité à d'autres contraintes (biotiques/abiotiques) peuvent répondre de manière inattendue lorsqu'elles sont cultivées au champ (Atkinson & Urwin, 2012). Ainsi, une variété de manioc résistante au *Cassava mosaic virus* présente une sensibilité accrue à l'aleurode *Bemisia tabaci* (Otim *et al.*, 2006).

Il est donc important de comprendre les interactions qui peuvent exister entre différents stress et leurs conséquences sur les plantes soumises à ces conditions de multistress. Des « matrices de contrainte » ont ainsi été développées mettant en avant les effets des combinaisons de stress qui pourraient être dommageables au champ (Fig. R.13) (Mittler, 2006; Suzuki *et al.*, 2014).



2 Des éléments communs de réponse aux stress biotiques et abiotiques

Malgré la diversité de différents stress, qu'ils soient abiotiques ou biotiques, des éléments de réponse communs peuvent être décrits. Ces facteurs communs de réponse sont particulièrement intéressants dans la recherche et l'identification de déterminants pouvant par la suite être utilisés dans des programmes de sélection (Atkinson & Urwin, 2012) (Fig. R14).

2.1 Les facteurs de transcription

Les **facteurs de transcription** ou **TFs** présentent un grand intérêt dans la réponse à des stress multiples puisque leur action va avoir un impact sur un grand nombre de mécanismes en aval (régulation des voies hormonales, de l'expression de gènes de défense entre autres). Ainsi des TFs appartenant à la famille des MYB sont associés à la réponse à la fois à des stress biotiques et abiotiques. C'est le cas par exemple du facteur *BOS1* chez *A. thaliana* qui assure la résistance à des stress biotiques en lien avec des champignons nécrotrophes (*B.cinerea* & *A.brassicicola*)

mais assure aussi la tolérance face à des stress abiotiques (stress hydrique, salin) (Mengiste, 2003). La fonction de ces TFs de la famille des MYB est d'ailleurs très conservée entre les espèces. En effet, l'homologue du facteur BOS1 chez la tomate (S1AIM1) présente aussi un rôle dans la résistance à *B.cinerea* et la tolérance à un stress abiotique (AbuQamar *et al.*, 2009). D'autres facteurs de transcription, NAC et AP2/ERF, ont déjà largement été caractérisés dans la réponse à des stress multiples. Ils constituent des cibles privilégiées dans le développement de variétés avec un large spectre de résistance.

En revanche, un grand nombre de ces TFs entraîne un blocage de la croissance des plantes lorsqu'ils sont exprimés. Cela illustre bien l'équilibre qui existe dans la répartition des ressources qui sont allouées à la croissance et à la défense en conditions de stress (Atkinson & Urwin, 2012).

2.2 Mitogen-activated protein kinase (MAPK)

Les cascades de signalisation activées par les MAPK sont cruciales chez les eucaryotes dans la réponse à divers stimuli environnementaux (Rodriguez *et al.*, 2010). Les plantes présentent un grand nombre de MAPK avec un large spectre d'action, impliquées dans la réponse à des stress biotiques et abiotiques. Les MAPK vont induire différentes cascades de signalisation avec l'activation de l'expression de facteurs de transcription ainsi que de voies hormonales. Par exemple, chez *Arabidopsis*, MPK6 est activée en réponse au froid et à des stress salins mais également en réponse à des stimuli induits par des agents pathogènes (Rodriguez *et al.*, 2010).

2.3 Espèces réactives de l'oxygène (ROS)

Les espèces réactives de l'oxygène ont un rôle important dans la réponse aux stress biotiques et abiotiques même si le mode de fonctionnement diffère entre ces deux types de stress. De façon générale, un niveau de production basal de ROS est maintenu par les plantes. Leur concentration va augmenter en réponse à des stress. Si les plantes sont capables de produire des antioxydants afin de limiter les dommages potentiels causés par ces molécules, dans le cas d'infections par des agents pathogènes cette accumulation est favorisée. Il s'agit du « burst » oxydatif qui va contribuer à la réponse d'hypersensibilité qui va permettre de limiter le développement de l'agent pathogène (Atkinson & Urwin, 2012).

Ces ROS constituent également des molécules de signal qui vont activer l'expression de gènes en fonction du type de stress en question. Par exemple, l'expression du gène codant pour le facteur de transcription ZAT12 chez *Arabidopsis* est induite par les fortes concentrations en ROS en réponse à des stress abiotiques et biotiques. De même, ZAT7 est également activé par une augmentation des ROS. Il confère une tolérance à divers stress abiotiques (salinité, froid) mais induit aussi l'expression de gènes impliqués dans les mécanismes de défense (*WRKY70*) (Gadjev, 2006; Atkinson *et al.*, 2013).

Un autre exemple qui souligne le rôle central de ces ROS dans la réponse à des stress se retrouve dans l'étude menée par Coolen et ses collaborateurs (2016). Une analyse transcriptomique de la réponse d'*Arabidopsis* à trois stress appliqués de façon séquentielle a été effectuée. Une analyse d'enrichissement (GO terms) a été menée sur le set de gènes les plus dérégulés en réponses à ces stress. Il apparaît qu'une seule fonction, celle associée à la réponse aux espèces réactives de l'oxygène, est activée quelle que soit la combinaison de stress (Coolen *et al.*, 2016).

2.4 Les petits ARN (miRNA)

Les miRNA assurent la régulation de post-transcriptionnelle de l'expression des gènes chez les plantes en inhibant la transcription ou en dégradant les transcripts. Ainsi, ces miRNA sont impliqués dans les mécanismes de croissance, développement, prolifération cellulaire. Ils auraient également un rôle important dans la réponse à un large panel de stress abiotiques et biotiques (Nejat & Mantri, 2017). Cependant, seulement quelques études se sont intéressées au rôle de ces miRNA dans la réponse combinée à des stress biotiques et abiotiques. Ainsi, l'étude menée par Xin et ses collaborateurs (2010) montre le rôle de ces miRNA chez le blé en réponse à une combinaison de deux stress : la chaleur et le mildiou (*Erysiphe graminis f.sp. tritici*). Ils ont notamment mis en évidence l'implication de différents miRNA en fonction du stress corroborant l'idée d'un rôle spécifique à chaque stress (Xin *et al.*, 2010).

2.5 Contrôle hormonal des interactions stress abiotique-biotique

Les phytohormones présentent un rôle central dans la réponse à des stimuli extérieurs tout en assurant la croissance et le développement des plantes. Ces phytohormones n'agissent pas de façon indépendante, des interactions synergiques ou antagonistes ont ainsi été décrites entre les différentes voies hormonales en réponse à des stress. Ainsi, en réponse aux stress biotiques et en fonction des caractéristiques des agents pathogènes impliqués (biotrophes vs. nécrotrophes) la voie de l'acide salicylique (SA) ou celles de l'acide jasmonique (JA)/éthylène sont activées. L'acide abscissique (ABA) est quant à lui impliqué dans la réponse à des stress abiotiques. L'ABA a également un rôle important en ce qui concerne la régulation des mécanismes de défense face à des stress biotiques. Ainsi, l'ABA serait un régulateur négatif de ces mécanismes de défense en agissant comme antagoniste des voies du SA et du JA/éthylène normalement activées lors des stress biotiques (Kissoudis *et al.*, 2015). Ce rôle de l'ABA dans la suppression des mécanismes de défense est d'ailleurs exploité par les agents pathogènes eux-mêmes. Ainsi, *Pseudomonas syringae* type III sécrète des effecteurs (T3SE) qui ciblent la voie de l'ABA chez *Arabidopsis* pour déclencher la maladie (Asselbergh *et al.*, 2008). Des effets positifs de l'ABA sur les mécanismes de défense à des agents pathogènes ont néanmoins été décrits. Ainsi Melotto et ses collaborateurs (2006) ont montré que l'ABA en induisant la fermeture des stomates empêcherait l'entrée de certains agents pathogènes. L'action de l'ABA sur les mécanismes de défense pourrait donc être dépendante des caractéristiques de l'infection mise en jeu (nature de l'agent pathogène et temps d'infection). La production d'ABA semble donc avoir un rôle crucial dans la réponse des plantes à des stress multiples.

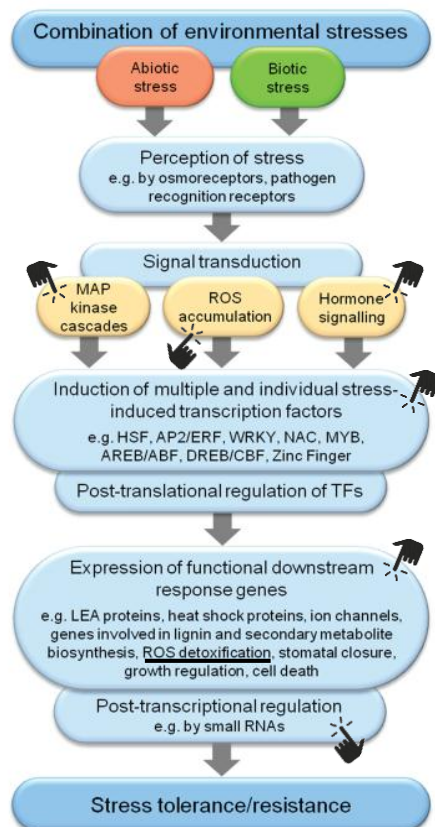



Figure R.14| : Principales étapes des différentes voies de signalisation activées en réponse à la combinaison entre stress abiotique-biotique (D'après Atkinson & Urwin, 2012)

Les interactions entre stress abiotiques et biotiques sont complexes et mettent en jeu divers facteurs impliqués dans différentes voies de signalisation dont l'objectif est de conférer la tolérance/résistance aux plantes face aux stress. Certains de ces facteurs et de ces voies de signalisation sont communs dans la réponse des plantes à des stress abiotiques et biotiques. Les caractéristiques de certains de ces éléments ont été développées tout au long de ce paragraphe et sont signalés par le symbole  sur le schéma

3 Interactions stress abiotiques et infections virales

L'impact des facteurs environnementaux (stress abiotiques) sur les maladies/ravageurs (stress biotiques) correspond au concept formulé par George McNew en 1960, **le triangle de la maladie (« Disease Triangle »)**. Les différents paramètres de ce triangle de la maladie sont en lien avec la sensibilité intrinsèque de la plante hôte, l'agressivité de l'agent pathogène et l'impact de l'environnement sur le développement de la maladie (Fig. R.15) (Scholthof, 2007).

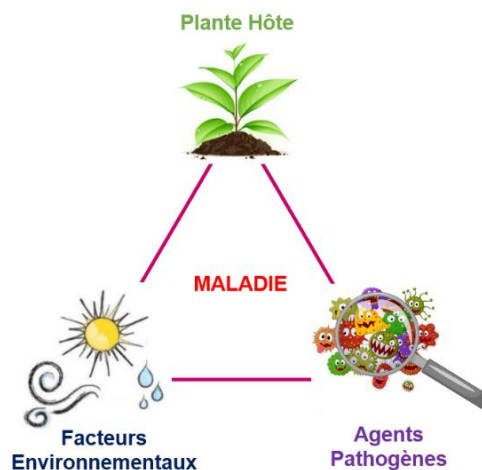


Figure R.15| Représentation schématique du concept de triangle de maladie - "Disease triangle"

Les effets de la combinaison de différents stress abiotiques et biotiques sur les plantes vont dépendre de la nature, la sévérité et la durée de ces stress. L'impact sur la croissance, la physiologie, la productivité des plantes est donc fonction de la combinaison de stress abiotiques et biotiques en question se traduisant par des interactions négatives ou positives (Fig. R.16) (Mittler, 2006; Atkinson *et al.*, 2013; Suzuki *et al.*, 2014; Pandey *et al.*, 2017).

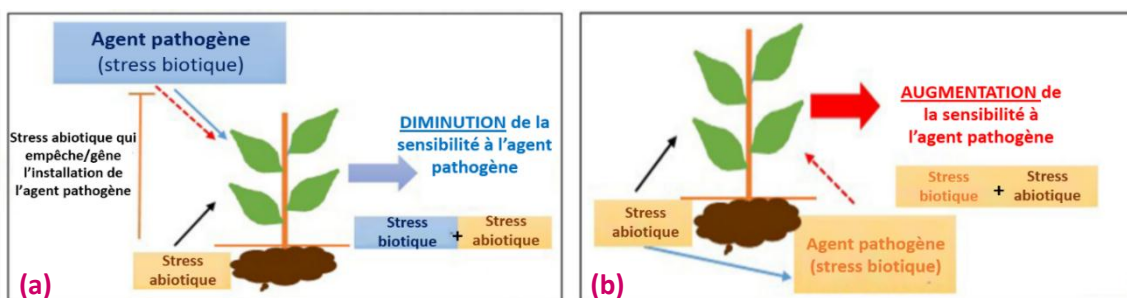


Figure R.16 | Effets des interactions stress abiotique-biotique dans la sensibilité des plantes à un agent pathogène (D'après Pandey *et al.*, 2017)

Selon la combinaison des stress, les stress abiotiques peuvent avoir un impact négatif sur l'installation, développement de l'agent pathogène (a) ou au contraire le favoriser (b) avec des conséquences sur la sensibilité de la plante en réponse à ce stress biotique.

L'étude de ces interactions est donc indispensable dans la compréhension des mécanismes mis en place par les plantes pour faire face à des situations de stress complexes. L'analyse des différentes revues scientifiques s'intéressant aux stress multiples permet de se rendre compte que la majorité des études développées s'intéressent le plus souvent à la combinaison de stress abiotiques. Par exemple, dans la revue de Suzuki et ses collaborateurs (2014), seulement 6 des 33 études recensées portaient sur l'étude de la combinaison entre stress abiotique et biotique (Suzuki *et al.*, 2014). De plus, dans le cadre des combinaisons stress abiotiques/biotiques, les études impliquant des stress abiotiques et des infections virales sont moins nombreuses que celles impliquant des champignons et des bactéries. Des résultats majeurs de certaines de ces études ont été repris permettant de montrer la complexité de ces interactions justifiant d'autant plus leur importance.

Dans leur système multifactoriel, Prasch et Sonnewald (2013) ont testé la réponse d'*Arabidopsis* (accession Col-0) à deux stress abiotiques (la chaleur et la sécheresse) et un stress biotique (une infection virale par le *Turnip mosaic virus*). Ainsi, ils ont montré que ces stress abiotiques augmentaient la sensibilité d'*A. thaliana* au TuMV. Une analyse comparative de transcriptomique a été menée sur des groupes de plantes où les stress étaient appliqués de façon individuelle ou combinée (2 ou 3 stress de façon simultanée). Il ressort que l'expression des gènes en conditions de stress multiples ne peut être prédite à partir de l'expression des gènes caractérisée pour un stress unique. D'ailleurs le nombre de gènes impliqués dans la réponse augmente avec la complexité et la sévérité des stress appliqués. Onze gènes sont retrouvés en communs à l'ensemble des conditions de stress. Deux facteurs de transcription - Rap2.9 et GBF3 - déjà caractérisés en réponse à des stress ont été identifiés parmi ces gènes communs (Sakuma *et al.*, 2002; Fujita, 2005). Cette étude a également permis de mettre en évidence des gènes qui vont être régulés de façon spécifique à une combinaison de stress. Ainsi, 29 gènes sont exclusivement exprimés lors de l'application simultanée des trois stress. (Prasch & Sonnewald, 2013).

D'autres études se sont également intéressées à la combinaison entre stress abiotiques et infection virales. Ainsi, il a été montré sur une étude menée sur le blé, que de fortes concentrations en CO₂ (650 µmol/mol alors que les concentrations normales sont de 400 µmol/mol) augmentent de 36.8% l'accumulation en systémie du *Barley yellow dwarf virus* (BYDV) (Trębicki *et al.*, 2015). Une étude similaire toujours sur ce même pathosystème blé/BYDV, a aussi montré une augmentation de la charge virale du virus chez des blés exposés à des températures élevées (Nancarrow *et al.*, 2014). Chez le pathosystème *Solanum tuberosum*/*Potato leafroll virus* c'est la combinaison des concentrations élevées en CO₂ ainsi que des températures élevées (30°C) qui augmente le pourcentage de plantes infectées (Chung *et al.*, 2017).

Les différents stress abiotiques peuvent avoir un impact à différents niveaux de l'infection virale. Ainsi, l'efficacité de transmission des virus est, comme la multiplication virale, impactée par ces stress. Une augmentation de la transmission par leur insecte vecteur (*Myzus persicae*) du *Cauliflower mosaic virus* (CaMV) et du *Turnip mosaic virus* (TuMV) a été montrée chez des navets soumis à des stress hydriques (van Munster *et al.*, 2017). En revanche, une deuxième étude publiée par la même équipe montre au contraire que l'efficacité de transmission du *Turnip yellow virus* (TuYV) est deux fois moins importante chez les plantes d'*A. thaliana* en condition de stress hydrique alors qu'il n'existe pas de différences significatives dans l'accumulation du virus entre les plantes soumises ou non à ce stress. Une des hypothèses pour expliquer les différences entre les deux études est en relation avec le fait que le CaMV et le TuMV sont des virus non-circulants alors que le TuYV est un virus circulant. Les virus circulants sont acquis et livrés directement au niveau du phloème par les pucerons. Or, en conditions de stress hydrique, les pucerons sont amenés à se nourrir plutôt au niveau du xylème. Cela pourrait expliquer la diminution de particules virales transmises par les pucerons (Yvon *et al.*, 2017).

Ces deux études montrent bien la complexité des réponses des plantes face à des combinaisons de stress. Un autre exemple illustrant ces multiplicités de réponses est l'étude menée par Xu et ses collaborateurs (2008) où l'infection par le *Cucumber mosaic virus* (CMV) de différentes espèces de plantes améliore leur tolérance à la sécheresse. Cette protection conférée pourrait être en lien avec les changements physiologiques importants (notamment en ce qui concerne le métabolisme primaire) induits lors de l'infection virale. Ce résultat pourrait en être en lien avec l'idée que le virus a tout intérêt à prévenir la mort de son hôte dont il est totalement dépendant pour pouvoir accomplir son cycle de vie (Xu *et al.*, 2008).

L'ensemble de ces études combinant stress abiotiques et infection virale montrent au final une telle complexité et multiplicité de réponses qu'il n'est sans doute pas possible d'établir un seul et unique schéma de réponse des plantes en conditions de stress multiples. Il est donc nécessaire de poursuivre le développement de ces études afin d'acquérir un maximum de connaissances sur ces réponses.

Il est intéressant de noter que globalement l'ensemble de ces études ont été menées en conditions contrôlées. D'ailleurs, très peu d'études concernant les interactions plantes/virus ont été menées en conditions naturelles (Hily *et al.*, 2016). Dans ces conditions contrôlées plusieurs questions peuvent se poser concernant l'application des stress abiotiques. En effet, il est pertinent de se demander comment un stress hydrique, salin, thermique peut être

appliqué afin de ne pas être trop intense de façon à se rapprocher au maximum des conditions au champ. Il ne faut effectivement pas oublier qu'en conditions naturelles, les plantes sont soumises à des stress de façon continue avec des variations plus ou moins importantes au cours de la journée (Mittler, 2006). De plus, la question concernant la chronologie d'application de ces stress se pose également : faut-il les appliquer en simultané, en séquentiel ? Par exemple, Coolen et ses collaborateurs (2016) ont analysé la réponse transcriptomique d'*Arabidopsis* à des stress appliqués de façon séquentielle. Il apparaît que dans le cas de deux stress appliqués de cette façon, les profils transcriptomiques obtenus sont similaires à ceux obtenus pour le deuxième stress même si des éléments de réponse caractéristiques du premier stress sont également retrouvés (Coolen *et al.*, 2016). Finalement, en conditions extérieures, au champ, ces questions ne se posent pas puisque dans cette situation les plantes sont naturellement soumises à une combinaison de différents stress. Ce qui devient en revanche plus complexe c'est de caractériser ces stress (intensité, durée, variabilité) et donc de connaître leur impact réel sur les plantes.

C. Elargir le phénotypage de la réponse des plantes aux virus : l'approche métabolomique

Afin de mieux comprendre les mécanismes complexes liés à la réponse des plantes à une infection virale, différentes technologies peuvent être utilisées de façon à analyser la réponse à différents niveaux. Dans le cadre de cette thèse, une approche métabolomique a été utilisée dans le but de caractériser finement l'interaction plante/virus. Les virus sont des parasites obligatoires qui détournent la machinerie de leur hôte pour pouvoir assurer leur cycle de vie. De fortes perturbations au niveau du métabolisme primaire et secondaire sont attendues. S'il est largement admis que le métabolisme secondaire est impliqué dans la réponse des plantes à des stress, le rôle du métabolisme primaire dans ces conditions de stress semble moins évident. Ainsi, dans le cadre de cette thèse il s'agissait de s'intéresser tout particulièrement aux effets de l'infection virale sur le métabolisme primaire, très peu exploré jusqu'à présent. Ainsi, dans le cadre de cette revue bibliographique l'accent a été mis sur **i)** les approches métabolomiques associées à la recherche de marqueurs métaboliques et sur **ii)** les connaissances acquises jusqu'à présent sur les perturbations du métabolisme primaire en réponse à des agents pathogènes et plus particulièrement en réponse à des virus. Des connaissances générales sur le métabolisme secondaire en lien avec la défense des plantes ont été apportées à la fin de cette partie.

1 Métabolomique – Métabolites – Biomarqueurs

1.1 Approches métabolomiques

1.1.1 Notions générales

Afin de mieux comprendre les mécanismes complexes liés à la réponse des plantes à une infection virale, différentes technologies peuvent être utilisées de façon à analyser la réponse à différents niveaux. Ainsi, parmi l'ensemble des approches « omiques » haut-débit (génomique, transcriptomique, protéomique, métabolomique), la **métabolomique** est la plus récemment utilisée. Elle permet l'étude complète qualitative et quantitative des métabolites d'un échantillon biologique grâce au développement de techniques de plus en plus performantes (Gibon *et al.*, 2012). Le métabolome, dernier niveau de régulation après la transcription et la traduction, est plus proche de la fonction biologique et est souvent considéré comme le reflet du phénotype (Fig. R.17) (Feussner & Polle, 2015). Il existe deux approches majeures de métabolomique **i)** le profilage métabolique qui consiste en l'analyse d'un nombre réduit de métabolites appartenant à une même voie ou à une même classe métabolique et **ii)** l'empreinte métabolique qui permet la détection d'un grand nombre de composés de façon non ciblée (Gibon *et al.*, 2012).

En plus des nouvelles technologies qui ont permis d'analyser et d'identifier un grand nombre de métabolites, le développement des outils informatiques et statistiques ont permis le traitement de grands jeux de données dans le but de pouvoir identifier des métabolites caractéristiques d'une réponse biologique. C'est ce qui a donc permis d'identifier des **biomarqueurs** également appelés **marqueurs métaboliques** lorsqu'ils correspondent à des concentrations en métabolites. Les premiers marqueurs métabolites ont été utilisés dans le domaine du médical. Dans le domaine du végétal, les premières utilisations de ces marqueurs

sont à relier avec le développement d'outils de diagnostics tels que Jubil® et N-tester® qui permettent d'assurer une fertilisation raisonnée en azote en fonction respectivement des teneurs en nitrates ou chlorophylles mesurées (Fernandez *et al.*, 2016).

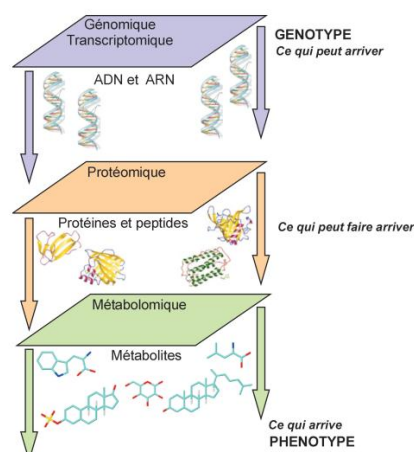


Figure R.17 | Représentation des différents niveaux de caractérisation des systèmes biologiques (D'après iPubli-Inserm – Approches « omiques »)

1.1.2 Les Biomarqueurs

Un **biomarqueur** peut être défini comme « une caractéristique qui est objectivement mesurée et évaluée comme indicateur des procédés biologiques normaux, des procédés pathogènes ou des réactions pharmacologiques liée à une intervention thérapeutique » (NIH Definition Working Group, 2000). Dans le domaine des plantes, cette notion de biomarqueur est souvent associée aux critères de performance (différents traits qui sont suivis au cours du développement des plantes et/ou améliorés dans les programmes de sélection). Trois types de marqueurs moléculaires différents peuvent être décrits (Fernandez *et al.*, 2016) :

- **Marqueurs d'importance agronomique** : ils vont permettre d'effectuer des screening sur un large nombre d'échantillons. Cette prospection peut s'effectuer en utilisant des métabolites désirés (vitamine C, arômes) ou non désirés (toxines).
- **Marqueurs diagnostique** : ils vont permettre de caractériser la réponse biologique à une situation donnée. C'est le cas par exemple de la proline qui s'accumule lorsque les plantes sont soumises à des stress hydriques (Hayat *et al.*, 2012).
- **Marqueurs performance** : il s'agit d'une combinaison de métabolites qui sont associés à une performance. Ainsi, en 2007 pour la première fois des profils métaboliques étaient utilisés pour estimer la production de biomasse chez *Arabidopsis* (Meyer *et al.*, 2007). L'idée finale est de pouvoir utiliser ces associations entre métabolites et performance afin de pouvoir prédire le comportement d'individus sur la base de leurs profils métaboliques.

1.2 Pipeline pour l'identification de biomarqueurs

La caractérisation de la composition métabolique des échantillons biologiques analysés est indispensable. Ainsi, il est nécessaire de mettre en place un protocole allant de la collecte des données métaboliques en passant par leur traitement et analyses permettant d'identifier un ou plusieurs métabolites pouvant être associé(s) à une performance (Fig. R.18) (Zabotina, 2013; Fernandez *et al.*, 2016).

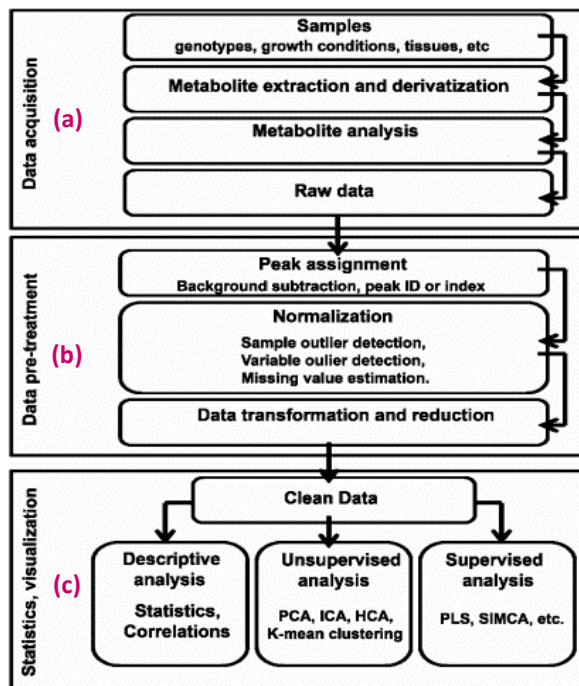


Figure R.18 | Pipeline pour l'identification de biomarqueurs (D'après Zabolina, 2013)

L'ensemble des étapes comprennent (a) l'obtention des données métaboliques qui passe par la production et gestion des échantillons, la détection des métabolites (b) le traitement du jeu de données afin d'obtenir les données (« clean » data) qui serviront aux analyses statistiques qui permettront d'identification de potentiels biomarqueurs (c)

Abréviations: PCA, Principal component analysis; ICA, Independent component analysis; HCA, Hierarchical clustering analysis; PLS, Partial least square

1.2.1 Acquisition des données métaboliques

Les concentrations et les flux métaboliques sont soumis à d'importantes variations en fonction des conditions de culture, des espèces/génotypes/variétés analysés, de leur stade de développement, etc. Ainsi, afin de pouvoir identifier de façon fiable et reproductible des marqueurs métaboliques, une attention particulière doit donc être accordée aux conditions expérimentales (évaluation de l'ensemble des facteurs pouvant avoir un impact sur les métabolites), du design expérimental à la collecte et gestion des échantillons. Le choix de ces conditions va donc être spécifique à chaque étude en fonction de la problématique de travail, des résultats attendus et de l'investissement requis (Fernandez *et al.*, 2016).

Le développement des technologies en métabolomique a rendu possible l'analyse d'un grand nombre de métabolites au sein des échantillons collectés. Le choix des outils utilisés pour le traitement des échantillons est un compromis entre nombre d'échantillons analysés et nombre de métabolites détectés. Ainsi, deux catégories d'approches peuvent être décrites : les **approches à haut-débit verticales et horizontales** (Fig. R.19) (Gibon *et al.*, 2012). L'**approche à haut-débit verticale** également appelée approche à haute densité correspond à l'analyse non ciblée d'un grand nombre de métabolites dans un nombre restreint d'échantillons. Il s'agit d'une approche particulièrement intéressante dans le domaine végétal puisque les plantes sont caractérisées par une grande diversité de métabolites (le nombre de métabolites produits est estimé entre 200 000 et 1 000 000). Identifier cette diversité est donc indispensable pour la compréhension des mécanismes biologiques (Gibon *et al.*, 2012). L'extraction des échantillons va dépendre essentiellement des caractéristiques des métabolites analysés et des techniques de détection utilisées. La détection des métabolites peut être effectuée essentiellement à partir de deux techniques : la spectrométrie de masse (MS) et la résonance nucléaire magnétique (RMN). Ces deux méthodes pouvant être couplées à des techniques de séparation par chromatographie gazeuse (GC) ou liquide (LC) ou par électrophorèse capillaire (CE) (Zabolina, 2013). En fonction des différentes combinaisons de technologies utilisées, le nombre de métabolites détectés varie ainsi que la résolution et la

précision de la détection. Cette approche présente néanmoins deux limites majeures **i)** l'identification des différents métabolites détectés qui peut s'avérer compliquée et coûteuse. Si des bases de données existent (FiehnLib, NIST) seulement un nombre réduit de métabolites spécifiques des plantes y sont recensés ; **ii)** l'obtention de données quantitatives précises qui peuvent être difficiles à obtenir du fait des fluctuations liées à l'ionisation des molécules analysées (Zabotina, 2013).

Dans l'**approche à haut-débit horizontale** un grand nombre d'échantillons sont analysés pour un nombre réduit et ciblé de métabolites. Une des contraintes majeures de cette approche concerne la gestion des nombreux échantillons analysés. Si les premières étapes de la préparation des échantillons (récolte, broyage, pesée) peuvent difficilement être automatisées, la détection des métabolites dans un grand nombre d'échantillon peut actuellement être gérée par une seule personne. En effet, la quantification des métabolites peut être réalisée en microplaques via des réactions chimiques, enzymatiques qui nécessitent l'utilisation de lecteurs spécifiques. Les moins chers, permettant le dosage d'un large panel de métabolites, sont les lecteurs de microplaques avec spectrophotomètre UV/VIS intégré. L'automatisation des dosages peut être augmentée via l'utilisation de robot de pipetage. Les plateformes donnant accès à toutes ces technologies de dosages se développent (Gibon *et al.*, 2012). C'est le cas de la plateforme métabolome de Bordeaux principalement dédiée à l'étude des plantes et de leurs produits dérivés (<https://metabolome.cgfb.u-bordeaux.fr/>).

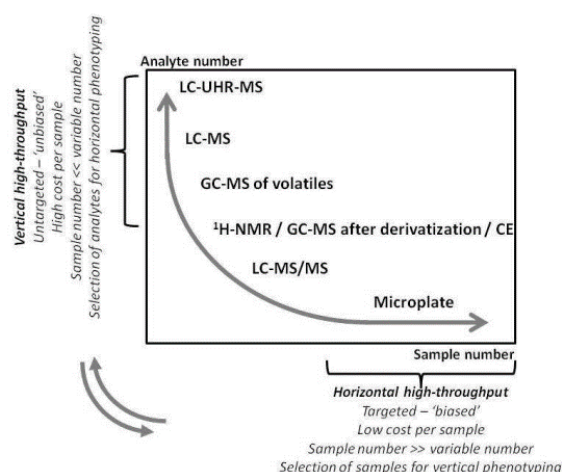


Figure R.19| Approches à haut-débit verticale et horizontale pour l'établissement d'empreintes ou profils métaboliques (D'après Gibon *et al.*, 2012)

Après détection des différents métabolites des échantillons biologiques, par l'une de ces deux approches, l'objectif est de pouvoir extraire l'information qui sera associée au critère de performance dans le but d'identifier de potentiels biomarqueurs de la réponse. Pour cela, un traitement des données est nécessaire afin de dé-complexifier les données, d'obtenir des données dans un format qui soit facile à manipuler par la suite, d'avoir des données normalisées (correction afin d'éliminer au maximum les biais liés à l'expérimentation, cela passe par l'utilisation de différents témoins tels que des blancs, des contrôles qualité, des standards biologiques) (Fig. R.18b) (Zabotina, 2013).

1.2.2 Traitement des données : vers l'identification de marqueurs métaboliques

La détection de marqueurs métaboliques implique l'utilisation de différentes méthodes statistiques afin d'établir des associations entre des profils/empreintes métaboliques et la performance/réponse biologique étudiée (Fig. R.18c). Dans la situation la plus simple où une forte corrélation est mise en évidence entre un métabolite et une performance, l'utilisation de **coefficients de corrélations (Pearson/Spearman)** est suffisante pour la valider. Cependant, la complexité des situations d'étude est souvent plus importante et implique plusieurs métabolites. Ainsi, **l'analyse canonique des corrélations (CCA)** basée sur la matrice des corrélations des variables permet de décrire les corrélations linéaires qui existent entre deux ensembles de variables mesurées sur un même échantillon biologique.

Afin d'avoir une vision globale des données obtenues, il est possible de réaliser une **analyse en composantes principales (PCA)**. Il s'agit d'une analyse multivariée qui a pour but de réduire le nombre de variables en composantes principales en perdant le moins d'information possible c'est-à-dire en gardant le maximum de variabilité totale. Les données réparties au sein de l'espace défini par ces composantes principales permettent une visualisation rapide des données entre elles (similitudes et différences) assurant une bonne discrimination de l'échantillon biologique étudié. La PCA est une méthode non supervisée puisqu'elle ne requiert que les données mesurées. En revanche, des méthodes supervisées peuvent aussi être utilisées. Elles sont basées sur le fait que les individus de l'échantillon constituent des groupes par rapport au critère de performance analysé. Des analyses **PLS-DA «Partial Least Square - Discriminant Analysis** » ainsi que des variantes de cette méthode **l'OPLS-DA « Orthogonal Partial Least Square-Discriminant Analysis** » ont été développées. Ces deux méthodes permettent notamment de caractériser des variables VIP « Variable importance in projection » parmi lesquelles de potentiels biomarqueurs de la réponse peuvent être sélectionnés. Ces deux méthodes statistiques permettent de construire des modèles permettant d'estimer le pouvoir prédictif des données par rapport au critère de performance. Différentes méthodes doivent ensuite être utilisées pour valider ces modèles (Zabotina, 2013; Worley & Powers, 2016; Fernandez *et al.*, 2016).

1.3 Quelle utilité pour les marqueurs métaboliques ?

Si la sélection assistée par marqueurs est largement utilisée dans les programmes de sélection, l'application des marqueurs métaboliques dans de telles procédures n'a pas encore été reportée. Cela pourrait certainement s'expliquer par la forte influence des conditions environnementales et expérimentales sur ces marqueurs qui sont donc moins stables que des marqueurs moléculaires. Cependant, les traits polygéniques, la polyploidie, les traits épigénétiques ou contrôlés par l'environnement sont difficiles à appréhender avec des outils génétiques. De plus, la composition métabolique est plus directement liée au phénotype et est indépendante des ressources génétiques disponibles. De récentes études ont mis en évidence le **potentiel prédictif des profils/empreintes métaboliques qui pourraient dans un futur proche constituer des outils pour la sélection** (Zabotina, 2013). D'ailleurs, l'analyse des métabolomes fait partie intégrante des nouveaux projets de sélection variétale tels que les projets d'investissement d'avenir aMaizING (sur le maïs), SUNRISE (sur le tournesol) ou encore BreedWheat (chez le blé). Ainsi, Riedelsheimer et ses collaborateurs (2012) ont

comparé le pouvoir prédictif de marqueurs moléculaires (SNPs) et métaboliques (130 métabolites dosés) sur différents traits agronomiques (rendement, floraison, contenu en amidon, sucres, lignines) chez des lignées hybrides de maïs. Il apparaît que les 130 métabolites permettent de prédire les traits analysés avec autant de succès qu'avec les 38 000 marqueurs SNPs (Riedelsheimer *et al.*, 2012).

L'utilisation de ces marqueurs métaboliques pourrait également **contribuer à apporter des informations sur le fonctionnement de certaines réponses biologiques**. La corrélation négative entre le contenu en amidon et la biomasse a permis de mettre en évidence l'équilibre fin qui existe entre l'approvisionnement de carbone et la croissance des plantes. Certaines accessions auraient modifié leur rapport carbone/croissance de façon à privilégier la croissance. Ce changement se traduisant par des changements dans le contenu en amidon mais également au niveau transcriptionnel. Ainsi, l'identification de deux gènes candidats dont le polymorphisme semble lié à la biomasse pourraient permettre d'aller rechercher des homologues dans des espèces cultivées afin d'essayer de moduler leur production de biomasse (Sulpice *et al.*, 2009).

Les marqueurs métaboliques peuvent être utilisés comme des **outils pour l'évaluation de critères de performance dans le cadre de process industriels**. Une des premières études à utiliser des marqueurs métaboliques comme outil diagnostic a été effectuée sur la pomme de terre qui est sensible aux noircissements internes (coloration provoquée par les endommagements des tubercules souvent associée à une compression lors du stockage) et aux « brunissements après-cuisson ». Une analyse PLS a permis d'identifier 5 acides aminés (tyrosine, threonine, valine, serine et glutamine) et 2 sucres (glucose et fructose) comme les meilleures variables associées à ces phénomènes impactant la qualité des pommes de terre (variables VIP). Ces données ont d'ailleurs permis de mettre en place des modèles prédictifs dont les forts coefficients de corrélation révèlent leur capacité à donner une indication fiable par rapports aux deux critères commerciaux analysés (Steinfath *et al.*, 2010). L'étude menée par Lopez et ses collaborateurs constitue un autre exemple de l'utilisation des marqueurs métaboliques dans l'évaluation de critères de performance. En effet, la commercialisation des fruits frais de tomate dépend de leur durée de conservation qui est en lien avec la fermeté des fruits. Ainsi, dans cette étude, ces deux traits semblent corrélés au contenu en malate des tomates (López *et al.*, 2015).

Enfin, les métabolites peuvent aussi être utilisés afin d'**évaluer des critères de performance des plantes soumises à des stress**. Un certain nombre d'études ont révélé le rôle de métabolites dans la réponse à des stress abiotiques. C'est le cas par exemple de l'implication du tréhalose dans la réponse à la sécheresse ou encore de l'allantoïne dans le cas de stress salin chez le riz (Quistián-Martínez *et al.*, 2012; Nam *et al.*, 2015). En revanche, l'utilisation de la métabolomique est relativement récente dans les études concernant des plantes soumises à des stress biotiques. Ainsi deux études basées sur des analyses de RMN ont été menées i) chez des plants de vigne infectés par l'esca et ii) sur des plantes de pervenche de Madagascar infectées par des phytoplasmes. Dans d'autres études, l'utilisation de marqueurs métaboliques permet de séparer les plantes infectées des plantes saines (Choi, 2004; Lima *et al.*, 2010). Ainsi, l'analyse des profils/empreintes métaboliques semble constituer un moyen

intéressant afin de caractériser la réponse des plantes face à des stress biotiques et en particulier dans la cadre des interactions plantes/virus.

2 Rôle du métabolisme primaire dans la réponse à des stress biotiques

L'immunité des plantes fait intervenir des mécanismes de réponse innés mais également induits par un agent pathogène suite à sa reconnaissance spécifique. La recherche et caractérisation de la réponse des plantes à des stress biotiques ont révélé la complexité des mécanismes impliqués. De plus, les agents pathogènes vont être à l'origine d'importantes modifications chez leur plante hôte dans le but de pouvoir assurer leur développement. Si de nombreuses connaissances ont déjà été acquises dans ce domaine, très peu de choses ont réellement été décrites sur **l'impact des agents pathogènes sur le métabolisme primaire des plantes**. De même, le **rôle de ce métabolisme primaire dans les mécanismes de défense** à des stress biotiques reste à présent peu caractérisé. Selon une hypothèse admise depuis des années, le rôle du métabolisme primaire serait de produire l'énergie suffisante pour supporter la réponse de défense. En effet, la disponibilité des ressources notamment en carbone est un point crucial dans la mise en place des mécanismes de défense. L'activation de la réponse de défense présente un coût énergétique important lié notamment à l'activation de nombreuses voies de signalisation (Bolton, 2009a; Rojas *et al.*, 2014a). Par exemple, l'infection par *P.syringae* chez *Arabidopsis* conduit à la dérégulation de 650 gènes dans les 10 min suivant l'infection et concerne plus de 2000 gènes après 7h post infection (Scheideler *et al.*, 2002). L'hypothèse du bilan entre croissance et différenciation (GBDH – « Growth Differentiation Balance Hypothesis ») (Herls & Mattson, 1992) décrit bien cet équilibre complexe entre défense et croissance/développement qui est justifié par la surexpression de gènes impliqués dans les mécanismes de défense et la sous-expression de gènes liés à d'autres processus métaboliques. Il y a de ce fait une reconfiguration du métabolisme primaire chez la plante en réponse au stress biotique. Ainsi, les processus liés à la production d'énergie (glycolyse, voie des pentoses phosphate, cycle de Krebs, chaîne respiratoire mitochondriale, synthèse d'ATP et de certains acides aminés tels que la lysine et la méthionine) vont être activés alors que les processus d'assimilation (photosynthèse, métabolisme de l'amidon, des lipides, synthèse de leucine/isoleucine/valine) seront réprimés (Less *et al.*, 2011).

Le métabolisme primaire semble être un élément crucial dans la mise en place des mécanismes de défense face à des stress biotiques. Ainsi une régulation fine de son activité pourrait fortement contribuer à la résistance des plantes aux maladies. La caractérisation de gènes de résistance en lien avec ces processus pourrait contribuer de façon très intéressante à la mise en place de variétés avec un large spectre de résistance aux maladies le plus durables possibles (Rojas *et al.*, 2014a).

A ce jour, seules quelques revues faisant part des connaissances sur le rôle du métabolisme primaire dans les mécanismes de défense à des agents pathogènes ont été publiées (Berger *et al.*, 2007; Bolton, 2009b; Rojas *et al.*, 2014b). Ce qui est encore plus surprenant, c'est le peu de connaissances nouvelles qui sont apportées dans les dernières revues, sept ans après la première revue prise ici en considération. Même s'il ne s'agit certainement pas d'une liste exhaustive des publications du domaine, cela montre qu'il reste encore beaucoup à faire afin

de pouvoir comprendre et caractériser le rôle du métabolisme primaire dans le cadre des interactions plante/agents pathogènes. Si ce manque de connaissances semble évident, il est d'autant plus marqué en ce qui concerne les maladies virales.

3 Métabolisme primaire et infection virales

En réponse aux infections virales, des changements importants concernant différentes voies métaboliques ont été décrits en particulier ceux concernant la photosynthèse et le métabolisme des carbohydrates. En raison du type d'interaction plante/virus analysé ainsi que des conditions d'étude du pathosystème (système d'expérimentation, caractéristiques des prélèvements, etc), une seule et unique description généralisée des réponses métaboliques déclenchées ne peut pas être réalisée (Llave, 2016). De plus, ces réponses métaboliques sont très complexes et dépendent de nombreux facteurs tels que le stade de développement des plantes, le moment de prélèvement, le type de prélèvement (feuilles, plante entière, prélèvement au niveau des sites d'inoculation ou en systémie, etc). Ainsi, les quelques études menées pour essayer de comprendre la reprogrammation métabolique lors d'une infection virale ont analysé la réponse à différents moment de la maladie (Fernandez-Calvino *et al.*, 2014; Sade *et al.*, 2015; Stare *et al.*, 2015; Kogovšek *et al.*, 2016).

Les avancées en métabolomique ont permis d'avoir accès à différentes approches qui ont rendu possible la quantification d'un large spectre de métabolites sur de nombreux échantillons que ce soit chez des espèces modèles ou non (Gibon *et al.*, 2012). Il est donc à présent possible de caractériser et d'évaluer la contribution des composés métaboliques intervenant lors d'une infection virale. Il est intéressant de noter que les résultats obtenus avec ces données relatives au contenu métabolique sont souvent associés à des données de transcriptomiques ce qui permet d'essayer d'identifier les voies métaboliques affectées par les infections virales en analysant la réponse à différents niveaux (Fernandez-Calvino *et al.*, 2014; Sade *et al.*, 2015; Stare *et al.*, 2015; Kogovšek *et al.*, 2016).

3.1 Infection virale et métabolisme des carbohydrates

Un phénomène important est observé au niveau des tissus infectés, c'est la transition source-puits caractérisée par une diminution de la photosynthèse et une augmentation de la demande en composés carbonés. La demande est principalement soutenue par le saccharose qui est synthétisé au niveau des feuilles sources et transporté via le phloème aux tissus puits. Ainsi des concentrations élevées en saccharose ou en sucres solubles ont été quantifiées dans certaines interactions plante/virus. Ces changements dans l'accumulation des carbohydrates sont souvent décrits en relation avec l'activité d'une enzyme membranaire, l'invertase. En effet, les taux élevés en sucres solubles relevés chez des tabacs infectés par le *Potato virus Y* (PVY) sont corrélés avec une augmentation de l'activité de l'invertase qui permet de cliver le saccharose en glucose et fructose. D'ailleurs des plantes *Arabidopsis* surexprimant l'invertase, accumulent une grande quantité de sucres solubles et présentent une résistance contre le *Tobacco mosaic virus* (TMV) (Herbers *et al.*, 1996; Llave, 2016).

Cette augmentation des sucres solubles est retrouvée dans l'étude menée par Fernandez-Calvino et ses collaborateurs (2014) sur la réponse métabolique d'*A. thaliana* au *Tobacco rattle virus* (TRV). En effet, des concentrations plus importantes en glucose, fructose et

saccharose ont été quantifiées au niveau des plantes inoculées avec le TRV par rapport aux plantes non inoculées et ce au moment où l'infection était considérée maximale (Fig. R.20) (Fernandez-Calvino *et al.*, 2014).

Dans leur étude concernant l'infection de *S.tuberosum* par le *Potato virus Y*, Kogovšek et ses collaborateurs ont montré une réponse dynamique de l'accumulation de ces sucres solubles au cours de l'infection. En effet, à un temps très précoce après inoculation (1 dpi) une diminution du contenu en hexoses est notée suivie d'une importante augmentation (3 & 6 dpi). Ce changement d'accumulation reflèterait la transition source-puits qui a lieu au moment où l'infection virale est maximale (Kogovšek *et al.*, 2016).

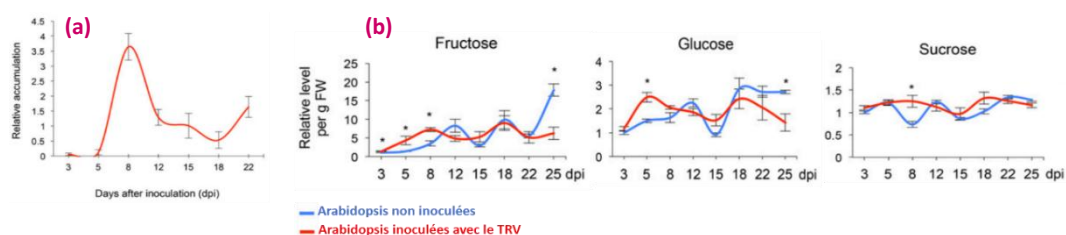


Figure R.20 | Effet de l'infection par le TRV sur les concentrations en carbohydrates chez *Arabidopsis* (D'après Fernandez-Calvino *et al.*, 2014)

(a) Quantification de la charge virale par qRT-PCR à différents temps après inoculation, une accumulation maximale du virus est notée à 8 dpi (b) Concentrations en fructose, glucose et saccharose (sucrose) quantifiées par GC-MS à différents temps au cours de l'infection. Les mesures ont été effectuées en systémie sur des feuilles de plantes inoculées par le TRV (en rouge) et non inoculées par le virus (en bleu). Les astérisques représentent les temps où des différences significatives dans les concentrations en sucres ont été observées entre plantes inoculées et non inoculées par le TRV.

3.2 Infection virale et respiration

Lors d'une infection virale, la respiration est stimulée afin de subvenir aux besoins énergétiques nécessaires à la mise en place des mécanismes de défense et/ou à la multiplication virale. Trois grandes voies métaboliques interconnectées assurant la production d'énergie peuvent être associées à la respiration : la glycolyse, le cycle de Krebs et la chaîne respiratoire mitochondriale (Fernie *et al.*, 2004). Ainsi, une accumulation de composés intermédiaires du cycle de Krebs a été notée dans les feuilles inoculées de pomme de terre infectées par le PVY (Kogovšek *et al.*, 2016). Comme le cycle de Krebs est sensible aux stress oxydatifs, les plantes ont la possibilité de l'éviter via le « shunt » de GABA (4-aminobutyrate). Ainsi, une augmentation de l'expression de la glutamate decarboxylase (GAD) l'une des trois enzymes majeures impliquées dans ce processus est constatée au cours de l'infection de *S.tuberosum* par le PVY (Kogovšek *et al.*, 2016).

D'autres processus métaboliques ont également été décrits comme impliqués dans la production d'énergie lors de la réponse des plantes à des agents pathogènes (la phase oxydative des pentoses phosphates, le cycle du glyoxylate, le processus de β -oxydation). Tout cela montre bien la reprogrammation du métabolisme primaire afin de pouvoir soutenir la demande énergétique nécessaire à la mise en place des mécanismes de défense et/ou probablement à la multiplication virale (Fig. R.21) (Bolton, 2009b)

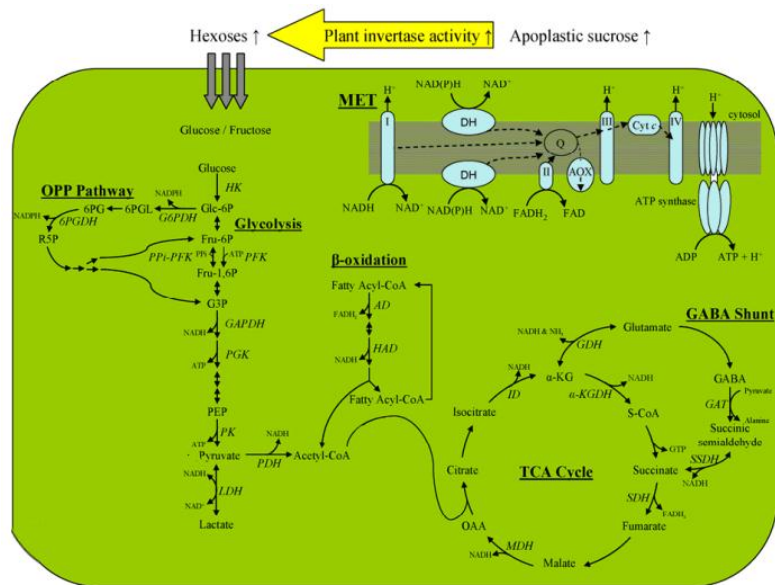


Figure R.21 | Voies du métabolisme primaire impliquées dans la production d'énergie lors de stress biotiques (D'après Bolton, 2009)

Dans cette représentation sont retrouvées les 3 principales voies métaboliques associées à la respiration : la glycolyse, le cycle de Krebs (TCA cycle) et la chaîne respiratoire mitochondriale (MET). L'alternative au cycle de Krebs, le shunt de GABA est également représentée. D'autres voies métaboliques permettant également de produire de l'énergie figurent aussi sur le schéma : la phase oxydative de la voie des pentoses phosphates (OPP Pathway) et la β -oxydation.

Abréviations : HK, hexokinase ; Glc- P, glucose 6-phosphate ; Fru-P, fructose 6-phosphate ; PPI-PFK, pyrophosphate : fructose 6-phosphate 1-phosphotransférase ; PFK, phosphofructokinase ; Fru-1,6P, fructose-1,6-biphosphate ; G3P, glyceraldehyde-3-phosphate ; GAPDH, glyceraldehyde-3-phosphate déhydrogénase ; PGK, phosphoglycerate kinase ; PEP, phosphoenolpyruvate ; PK, pyruvate kinase ; LDH, lactate déhydrogénase ; PDH, pyruvate déhydrogénase ; AD, acyl-CoA déhydrogénase ; HAD, 3-L-hydroxyacyl-CoA déhydrogénase ; ID, isocitrate déhydrogénase ; α -KG, α -ketoglutarate ; α -KGDH, α -ketoglutarate déhydrogénase ; S-CoA, succinyl-CoA ; OAA, oxaloacetate ; GDH, glutamate déhydrogénase ; GAT, 4-aminobutyrate aminotransférase ; SSDH, succinyl semialdehyde déhydrogénase ; DH, internal/external NADH déhydrogénase ; Q, ubiquinone ; AOX, alternative oxidase ; Cyt c, cytochrome c.

3.3 Infection virale et acides aminés

Les infections virales induisent d'importantes modifications dans le contenu en acides aminés qui semblent avoir un rôle important dans la mise en place des mécanismes de défense. En effet, la forte demande en composés carbonés dirige les acides aminés vers les voies métaboliques génératrices d'énergie tel que le cycle de Krebs (Bolton, 2009b). Dans l'étude de Fernandez-Calvino et ses collaborateurs (2014), des concentrations importantes en acides aminés sont relevées chez les plantes d'*Arabidopsis* infectées par le TRV au moment où l'accumulation virale est estimée maximale. Les données de transcriptomique à cette date montrent une sous-expression des gènes codant pour les protéines ribosomiques et une surexpression des gènes codant pour des composés impliqués dans le système ubiquitine/protéasome. Ces données suggèrent que le TRV induirait les mécanismes de protéolyse au niveau des tissus infectés (Fernandez-Calvino *et al.*, 2014). Des changements de concentrations en acides aminés ont été mis en évidence au cours d'une infection par le *Tomato yellow leaf curl virus* (TYLCV) de variétés résistantes et sensibles de tomate. En effet, à un temps précoce après inoculation (1 dpi) les variétés résistantes ont un contenu en métabolites dont le contenu en acides aminés est plus élevé que les variétés sensibles. Au cours de l'infection, le résultat s'inverse avec une accumulation en métabolites qui sera maximale chez les variétés sensibles à 7 dpi (Sade *et al.*, 2015). Malgré l'accumulation en

acides aminés qu'il semble y avoir lors d'une infection virale, la fonction précise de ces composés métabolique dans la réponse à ce stress biotique reste peu connue (Llave, 2016).

3.4 Infection virale et lipides

Les lipides sont des molécules présentant de multiples fonctions biologiques. En effet, ils fournissent des composants structurels pour les parois et membranes cellulaires, ils sont source d'énergie pour le métabolisme. Ils sont impliqués dans divers processus tels que les réarrangements cytosquelettiques, le trafic membranaire, les transductions du signal. Ils sont donc également impliqués dans la réponse à des stress biotiques.

Pendant l'infection, les virus interagissent avec le métabolisme des lipides et des acides gras dans le but de faciliter son entrée dans les cellules, sa réplication/multiplication, son assemblage et sa sécrétion. Par exemple, les virus vont être capables de réorganiser les membranes cellulaires afin d'établir des sites pour sa réplication. D'ailleurs, les virus à ARN positif peuvent former des complexes de réplication à partir des membranes de différents organites (réticulum endoplasmique, appareil de Golgi, endosomes, lysosomes, mitochondries) (Heaton & Randall, 2011). Ainsi, l'inactivation chez *N. benthamiana* des gènes *SMO1* et *SMO2* impliqués dans la biosynthèse des phytostéroïdes conduit à une réduction de l'accumulation virale d'un tombusvirus le *Tomato bushy stunt virus* (TBSV) (Sharma *et al.*, 2010). Lors de l'infection d'*Arabidopsis* par le TRV, des différences majeures du contenu en lipides entre plantes inoculées et non inoculées sont observées uniquement à 8 dpi c'est-à-dire au moment où l'accumulation virale est maximale (Fernandez-Calvino *et al.*, 2014).

3.5 Infection virale et polyamines

Les polyamines sont des composés ubiquistes avec une activité biologique. Il s'agit généralement de molécules contenant deux voire plus de groupements amines dont les plus communes sont la putrescine, la spermidine. Ces molécules semblent avoir un rôle important dans la réponse à des infections virales. En effet, un certain nombre d'études ont mis en évidence leur fonction dans la transduction des signaux de défense. (Llave, 2016). Ainsi, Fernandez-Calvino et ses collaborateurs dans leur étude de la réponse métabolique lors de l'interaction *Arabidopsis*/TRV ont mis en évidence des taux en putrescine deux fois plus élevés dans les plantes infectées, dès les phases précoces de l'infection, qui corrélaient avec la surexpression de gènes de la biosynthèse de ces polyamines (*ARGININE DECARBOXYLASE 1 & 2*). Des études précédentes avaient montré un rôle de la putrescine dans la résistance au froid, les auteurs ont voulu tester l'hypothèse que des plantes infectées par le TRV soient plus tolérantes aux basses températures du fait de l'accumulation de la putrescine. Les résultats ont été concluants puisque les plantes infectées par le TRV présentaient une meilleure tolérance au froid par rapport aux plantes contrôles (Fernandez-Calvino *et al.*, 2014).

4 Généralités sur le métabolisme secondaire

Une reprogrammation massive du métabolisme primaire semble avoir lieu chez les plantes en réponse à une infection virale. Les jonctions entre métabolites primaires et secondaires sont nombreuses. Ainsi, ces fortes perturbations du métabolisme central vont forcément avoir un impact sur le métabolisme secondaire. Le rôle du métabolisme secondaire en réponse à des agents pathogènes ayant déjà largement été démontré, il s'agissait donc dans cette partie d'apporter quelques informations générales sur ces molécules secondaires.

Les métabolites secondaires constituent un groupe très diversifié : trois grandes catégories sont généralement décrites chez les végétaux **i)** les composés phénoliques, **ii)** les alcaloïdes et composés azotés et **iii)** les composés terpéniques (Dixon, 2001). Ils sont produits à partir de différents métabolites primaires ou de leurs intermédiaires selon différentes voies métaboliques mises en place en réponse à différents stress (Fig. R.22). Ainsi, il a par exemple été démontré dans une étude sur le *Tomato yellow leaf curl virus* (TYLCV), que les lignées résistantes au virus présentaient des changements plus coordonnés du métabolisme primaire en comparaison aux lignées sensibles. Ces modifications métaboliques induisaient la production de composés secondaires permettant de contrer efficacement l'infection virale (Sade *et al.*, 2015).

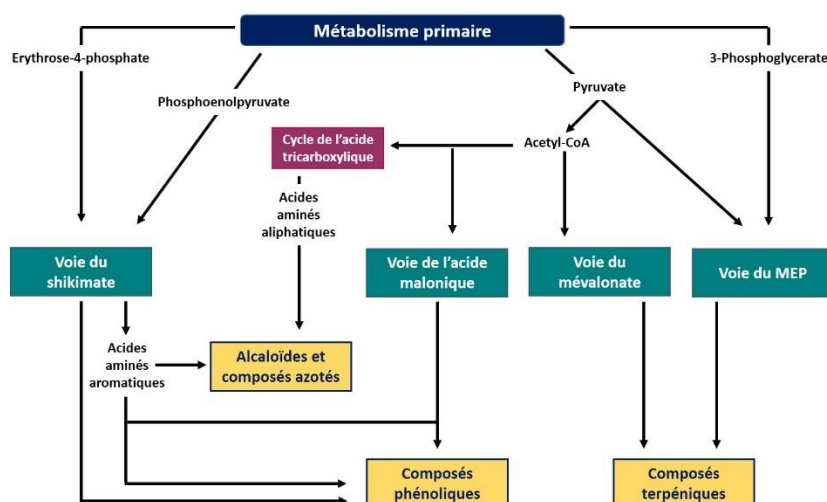


Figure R.22 | Schéma simplifié des principales voies métaboliques impliquées dans la biosynthèse de métabolites secondaires en lien avec les métabolites primaires (D'après Taiz & Zeiger, 2010)
Voie du MEP = Voie du méthylérythitol phosphate

Les composés secondaires impliqués dans les mécanismes de défense ont été caractérisés en deux catégories (Piasecka *et al.*, 2015; Bhadauria, 2016) :

- les **phytoalexines** qui sont les molécules induites au moment d'une infection. Dans cette catégorie, sont retrouvées les camalexines, les coumarines, les terpenoïdes, les flavonoïdes parmi lesquels sont décrits entre autres les anthocyanes, les flavones, les chalcones.
- les **phytoanticipines** qui sont des molécules constitutivement retrouvées au sein des plantes mais dont les caractéristiques biochimiques sont modifiées en réponse à une infection. Les saponines ou encore les glucosinolates font partie de cette catégorie.

Chez *A. thaliana*, les différents composés secondaires ont été caractérisés. L'arsenal de molécules secondaires est donc composé d'anthocyanes, flavones, glucosinolates, terpenoïdes, camalexines et de molécules dérivées du tryptophane. Les glucosinolates constituent le groupe de molécules secondaires le plus diversifié avec plus de 36 structures différentes décrites. Il s'agit de molécules retrouvées de façon constitutive chez *A. thaliana*. Leur activité biologique est due aux produits d'hydrolyse obtenus grâce à l'activité enzymatique de myrosinases. Le rôle des glucosinolates en réponse à des ravageurs et agents pathogènes a largement été décrit (Kliebenstein, 2004). Les camalexines sont d'autres

molécules secondaires dont l'implication dans les interactions plantes/agents pathogènes a été montrée. Il s'agit de phytoalexines dont la synthèse est induite en condition de stress (Kliebenstein, 2004). De nombreux composés secondaires ont un rôle potentiel ou démontré dans les interactions plantes/agents pathogènes. Mais, pour que leur action soit efficace, et pour préserver les tissus de la plante, ces composés doivent être synthétisés au bon moment et au bon endroit (Piasecka *et al.*, 2015). Malgré la caractérisation d'un grand nombre de composés secondaires et la confirmation de leur implication dans la réponse des plantes à des stress, les informations relatives à leurs mécanismes d'action et donc fonction sont moins nombreuses. Ainsi, ces molécules pourraient présenter une activité anti-microbienne mais pourraient également contribuer à la régulation de mécanismes impliqués dans la défense (Piasecka *et al.*, 2015).

Les différentes études citées montrent qu'il existe, en conditions d'infection, d'importantes perturbations du métabolisme primaire. Ces importantes modifications sur le métabolisme central vont avoir des conséquences sur le métabolisme secondaire dont le rôle dans la réponse à des stress a déjà largement été démontré. Il est intéressant de noter que la majorité des études ont été menées en conditions contrôlées. Par ailleurs, très souvent, peu de génotypes ont été étudiés (souvent Col-0 dans le cas d'*A. thaliana*). Les situations les plus souvent décrites sont aussi des situations de sensibilité. L'utilisation d'un phénotype métabolique large et sur un grand nombre de génotypes est une approche nouvelle susceptible d'apporter de nouvelles connaissances sur les interactions plantes/virus. Néanmoins, si la reprogrammation massive du métabolisme semble être fortement liée en réponse aux fortes demandes en énergie il est souvent difficile de savoir si cette réallocation des ressources contribue à la mise en place des mécanismes de défense, à la multiplication du virus ou aux deux. Il reste donc encore beaucoup de questions quant au rôle précis de ces mécanismes métaboliques notamment dans le cadre des interactions plante/virus (Piasecka *et al.*, 2015; Llave, 2016).

D. Etudier l'architecture génétique des interactions plantes/virus

Comme cela a déjà été discuté, le développement de variétés résistantes constitue le moyen de lutte le plus efficace, économique et écologique dans le contrôle des maladies virales. Le potentiel évolutif des virus a un impact important sur la durabilité des résistances allant jusqu'au contournement. Il est donc important de rechercher et d'identifier de nouveaux déterminants génétiques de la réponse des plantes aux virus. La durabilité des résistances développées dépend d'un certain nombre de facteurs. Ainsi, la nature de ces résistances est un de ces facteurs sur lesquels il est possible d'agir. D'ailleurs il a été démontré que les résistances quantitatives seraient plus durables (Parlevliet, 2002). La recherche et l'identification de ces **QTL** (Quantitative Trait Locus) peut s'effectuer à partir de deux approches de génétique quantitative : la **cartographie par analyse de liaison** et la **génétique d'association** (Bazakos *et al.*, 2017). Ces approches visent à associer statistiquement une variation phénotypique aux variations génétiques faisant ainsi ressortir les régions du génome impliquées dans la variation du trait phénotypique et dans lesquelles peuvent se trouver de potentiels gènes candidats de la réponse étudiée. Ces approches ont bénéficié du développement de marqueurs de type marqueurs microsatellites, SSR (Single Sequence Repeat) permettant d'identifier les polymorphismes génétiques mais surtout, avec le développement des méthodes NGS (Next-Generation Sequencing) de marqueurs SNP (Single Nucleotide Polymorphism) (Kumar *et al.*, 2017).

1 Cartographie par analyse de liaison

La première étape dans ce type d'analyse est de construire la population qui sera utilisée. Elle doit être suffisamment grande et comporter un nombre d'individus recombinants suffisant afin de permettre l'étude approfondie du trait phénotypique d'intérêt. Chez les plantes, il existe différents types de populations issues de différents croisements en lien avec le système de reproduction de l'espèce étudiée. L'objectif étant de maximiser le nombre de recombinaisons afin de disposer une variabilité génétique la plus grande possible. De plus, il s'agit également de pouvoir travailler avec des populations les plus homozygotes possibles afin de pouvoir les multiplier et les phénotyper plusieurs fois (Doerge, 2002; Bazakos *et al.*, 2017; Kumar *et al.*, 2017; Xu *et al.*, 2017).

1.1 Populations d'étude

1.1.1 Populations bi-parentales

Ces populations sont issues du croisement entre deux génotypes parentaux présentant des phénotypes contrastés. Ainsi, les **populations F2** sont les plus faciles à générer. Les individus de ces populations sont caractérisés par un grand nombre de régions hétérozygotes ce qui nécessite de générer des populations de grande taille pour la détection de QTL, de plus, cela rend compliqué le phénotypage sur des essais indépendants. Les **populations de RIL (Recombinant Inbred Lines)** permettent de dépasser ces contraintes. Ces populations sont obtenues par auto-fécondation des F2 (6 à 8 générations). Il s'agit donc de populations caractérisées par un fort taux d'homozygotie. Le temps de génération des populations RIL peut être raccourci par la création de **lignées DH (Double-Haploid)**. Enfin des populations **NILS**

(Near-isogenic lines) ont été également générées et déjà largement utilisées dans la recherche de QTLs (Bazakos *et al.*, 2017; Kumar *et al.*, 2017; Xu *et al.*, 2017) (Fig. R.23)

La principale limite de ces populations bi-parentales est liée au faible nombre de recombinaisons obtenu lors de leur création. La détection des QTLs dépend de la diversité phénotypique entre les deux parents qui correspond à une faible partie de la variation génétique qui existe au sein d'une espèce (Xu *et al.*, 2017). L'intervalle de détection dépend quant à lui de la taille de la population et de l'effet du QTL.

1.1.2 Populations multi-parentales

Les populations multi-parentales ont été construites pour dépasser les limites liées à l'utilisation des populations bi-parentales. En effet, l'utilisation de multiples parents permet d'accroître la diversité de réponse relative au trait analysé ce qui améliore la résolution dans la détection des QTLs. Ainsi les populations **NAM (Nested-association mapping)** sont un ensemble de populations RILs impliquant un parent commun. Les populations **MAGIC (Multiparent advanced generation intercrosses)** sont le résultat de croisements entre différents parents menés sur plusieurs générations et rendus homozygote par auto-croisement (Bazakos *et al.*, 2017; Xu *et al.*, 2017) (Fig. R.23)

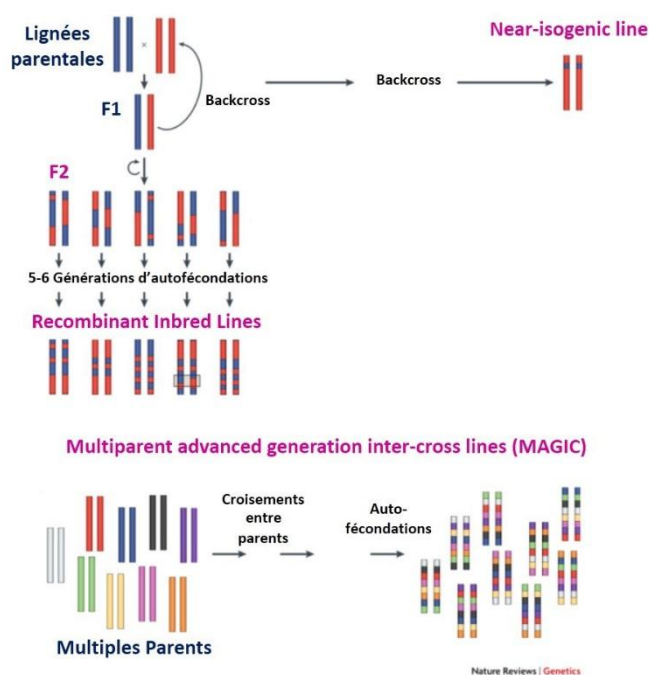


Figure R.23 | Différentes populations utilisées pour l'identification de QTLs par cartographie de liaison (D'après Bergelson & Roux, 2010)

1.2 Méthodes d'analyse de la cartographie par liaison

Le principe de base de l'analyse consiste en une **analyse simple-marqueur** qui teste individuellement l'association de chaque marqueur polymorphe avec le phénotype étudié en utilisant des outils statistiques simples (t-test, ANOVA, régression linéaire simple). Cependant ce type d'analyse s'effectue de façon indépendante marqueur par marqueur et ne tient pas compte des distances génétiques entre ces marqueurs (Doerge, 2002; Xu *et al.*, 2017). Afin d'améliorer la résolution de l'analyse, d'autres méthodes ont été développées. Ainsi, Lander & Botstein (2009) ont proposé l'«**interval mapping**» (**IM**). Basée sur une carte génétique, cette méthode teste la vraisemblance de la présence d'un QTL entre les marqueurs génétiques

à intervalle réguliers. Cette méthode suppose qu'un seul QTL serait impliqué dans la réponse. Cependant, il a déjà largement été démontré que les traits quantitatifs sont susceptibles d'être sous le contrôle de plusieurs loci. D'autres méthodes telles que le « **composite interval mapping** » (CIM) et le « **multiple interval mapping** » (MIM) ont été développées pour permettre de prendre en compte l'effet de plusieurs QTL à la fois.

1.3 Limites de l'approche

L'ensemble de ces méthodes de cartographie par analyse de liaison ont donc permis de caractériser les bases génétiques d'un trait quantitatif d'intérêt (nombre de loci, effets additifs). Cette approche présente néanmoins deux inconvénients majeurs. Le premier concerne la caractérisation des bases génétiques du trait d'intérêt qui ne montre que la différence entre les parents utilisés qui ne sont pas toujours représentatifs de la variation génétique retrouvée dans les populations naturelles. Le second concerne la résolution de l'approche qui amène à identifier de grandes régions génomiques dans lesquelles il faudra déterminer des gènes candidats à tester (Korte & Farlow, 2013).

2 Génétique d'association

La **génétique d'association** ou **GWA** (« **Genome Wide Association** ») permet tout comme la précédente approche de détecter et localiser des QTL impliqués dans la variation d'un trait quantitatif d'intérêt en recherchant les associations significatives entre les génotypes et les phénotypes. La principale différence réside dans le type de population étudiée puisque ces études d'association concernent des populations sans apparentement entre les individus (Myles *et al.*, 2009). Cette approche va donc exploiter les phénomènes de recombinaisons ayant eu lieu pendant toute l'histoire évolutive de la population étudiée. Cela constitue certainement son atout principal puisqu'en fonction du niveau de LD (« **Linkage disequilibrium** ») dans la population étudiée, la résolution des régions génomiques identifiées pourra être meilleure par rapport à l'approche de cartographie par liaison. De plus, les bases génétiques caractérisées ne sont plus uniquement le reflet de la variabilité entre deux parents mais correspondent à la diversité génétique globale du trait d'intérêt disponible sur l'ensemble des génotypes étudiés (Myles *et al.*, 2009). La puissance statistique de résolution de cette approche est sous l'influence de certains paramètres tels que l'étendue du déséquilibre de liaison, de la structuration de la population ainsi que du degré d'apparentement entre les individus de la population (Fig. R.25).

2.1 Paramètres importants du GWA

Le **déséquilibre de liaison** (LD – « **Linkage disequilibrium** ») correspond à l'association non aléatoire entre les allèles de différents loci dans une population donnée. La précision des analyses en GWA dépend de la distance physique sur laquelle décroît le LD dans la population c'est-à-dire la distance à partir de laquelle l'association entre l'allèle d'un locus 1 et celle d'un locus 2 n'est plus significative. Plus le nombre de recombinaisons est grand, plus cette distance est faible et donc plus la précision dans la détection d'un QTL pour le trait d'intérêt est importante. Cette distance physique de décroissance du LD va donc avoir un impact sur la **densité des marqueurs moléculaires** nécessaire et sur le niveau de résolution de détection qui peut être espéré. En effet, si le LD décroît faiblement, une faible densité de marqueurs

moléculaires suffit. Au contraire, pour un LD qui décroît fortement, il faut une densité de marqueurs plus grande. Dans cette deuxième situation la résolution de détection sera meilleure (Fig. R.24) (Xu *et al.*, 2017).

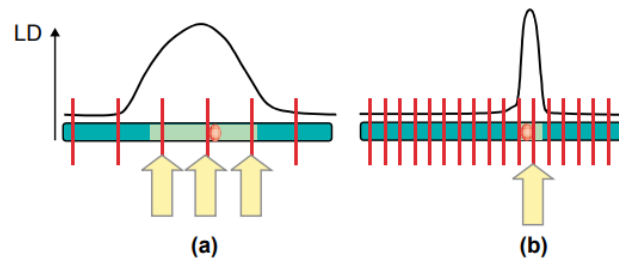


Figure R.24 | Etendue du déséquilibre de liaison et densité de marqueurs moléculaires associée (D'après Rafalski, 2002)

(a) Faible décroissance du LD autour du gène responsable du phénotype (ovale orange) ce qui nécessite une faible densité de marqueurs afin d'identifier les marqueurs associés au phénotypes (flèches jaunes) **(b)** Forte décroissance du LD autour gène causal, une forte densité de marqueurs est requise pour identifier les marqueurs associés au phénotype

Cette étendue du LD peut énormément varier en fonction des espèces étudiées. Cette variation peut également être importante au sein d'une même espèce. Par exemple, chez le maïs le LD peut atteindre 100kb dans les variétés commerciales d'élite alors qu'il n'est que de 2kb dans d'autres lignées pures (Remington *et al.*, 2001).

Le LD est également affecté par d'autres facteurs tels que les mutations, les recombinaisons, la dérive génétique, la sélection, la migration. Ainsi, l'isolement génétique et la subdivision des populations créent du LD alors que la recombinaison permet de faire diminuer le LD en cassant les associations au fur et à mesure des événements de reproduction (Xu *et al.*, 2017).

Un autre facteur important dont il faut tenir compte lors des analyses de GWA est la **structuration de la population** qui peut être à l'origine de faux positifs à cause des patrons de structure génétique complexes liés à l'histoire démographique des populations pouvant provoquer des associations erronées. Des méthodes ont de ce fait été développées pour corriger l'effet lié à cette structuration. Le programme STRUCTURE par exemple permet d'attribuer les individus à des groupes de structure (matrice Q) qui peuvent être ensuite utilisés comme covariables dans les tests d'association génotype/phénotype. Ces groupes de structure (matrice Q) peuvent être déterminés par d'autres outils tels que le « genomic control » ou le « structure association » qui utilisent un set de marqueurs afin de déterminer le nombre de sous-populations existantes et déterminer la place des individus dans ces sous-groupes. Des analyses en composante principale (ACP) peuvent aussi être effectuées.

S'il est très important de tenir compte de la structuration de la population, cela peut poser problème lorsque la variation phénotypique adaptative (en lien avec cette structure de la population) chevauche la variabilité génétique du trait d'intérêt entraînant l'obtention de faux négatifs.

Le **degré d'apparentement entre individus** peut également avoir un impact sur les variations du phénotype dont il est possible de tenir compte en utilisant une matrice de kinship (matrice K). Le coefficient de kinship correspond à la probabilité pour que deux allèles tirés au hasard

au même locus chez deux individus différents soient identiques par descendance (notion de IBD, « Identity By Descent »).

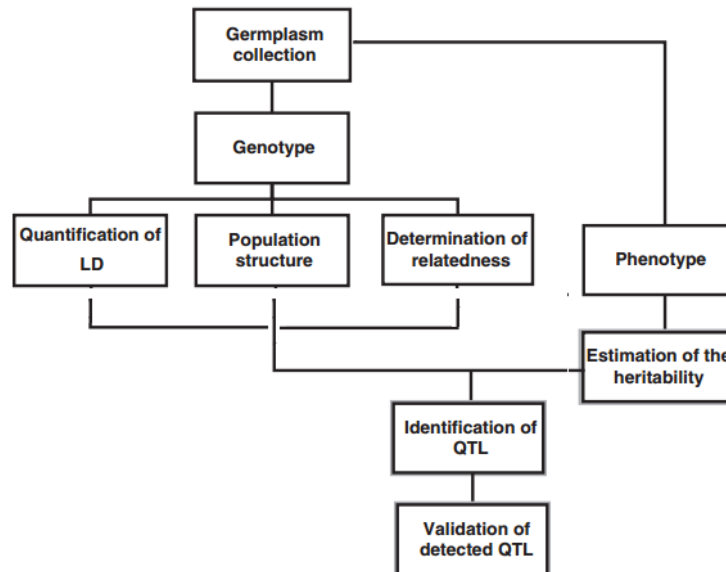


Figure R.25 | Organigramme des analyses GWA (Daprès Xu *et al.*, 2017)

Les grandes étapes de l'analyse GWA sont présentées avec au départ la constitution de la population d'étude qui devra être génotypée et phénotypée. Les différents facteurs susceptibles d'avoir un effet sur le résultat de l'analyse doivent être pris en considération : LD, la structure de la population et le degré d'apparentement entre les individus. L'analyse GWA peut ensuite être réalisée afin de tester l'association génotype/phénotype et permettre d'identifier des QTL qui devront être validés par la suite par d'autres méthodes.

2.2 Méthodes d'analyse de génétique d'association

Un certain nombre de méthodes ont été développées afin de mener ces analyses de GWA tout en tenant compte des différents facteurs qui peuvent influencer les résultats obtenus. Ainsi, le **modèle MLM** (« Mix linear model ») développé par Yu et ses collaborateurs (2006) est utilisé dans de nombreuses études puisqu'il tient compte en particulier de la structure et du degré d'apparentement entre les individus de la population. Cependant ce modèle exige d'avoir des populations de grande taille et une densité de marqueurs moléculaires importante (~ 1 millions de marqueurs). Cela entraîne un temps d'analyse relativement long. C'est d'ailleurs ce paramètre qui est amélioré par les nouvelles méthodes mises en place (Xu *et al.*, 2017). C'est le cas par exemple des modèles linéaires mixtes **EMMA** (« Efficient Mixed Model Association ») et **GEMMA** (« Genome-wide Efficient Mixed Model Association ») qui sont des méthodes exactes qui prennent en compte la variance polygénique (Kang *et al.*, 2008; Zhou & Stephens, 2012). Le développement de ces différentes méthodes d'analyse a conduit à la réalisation d'études comparatives permettant de mettre en avant les limites et avantages de chacune d'entre elles (Eu-ahsunthornwattana *et al.*, 2014).

2.3 Significativité de l'association

Afin de pouvoir considérer un QTL identifié comme ayant un effet significatif sur le trait phénotypique analysé différentes méthodes ont été développées afin de pouvoir disposer d'un seuil de significativité pour analyser les résultats. Plusieurs seuils de significativités peuvent être utilisés. Ainsi, le **critère de Bonferonni** correspond au rapport entre le seuil de significativité global souhaité (fixé généralement à 0.05) et le nombre de tests réalisés (en relation avec le nombre de SNPs). Les **permutations** aléatoires des phénotypes tout en

conservant la structure génétique permet de tester l'hypothèse nulle de non-association. Si cette méthode donne des résultats fiables, les temps de calculs peuvent être importants. Enfin le « **False Discovery Rate** » (**FDR**) évalue la proportion de faux positifs parmi les associations déclarées comme positives (Xu *et al.*, 2017).

2.4 Limites de l'approche

Certaines limites liées à cette approche de génétique d'association ont déjà été évoquées lors notamment de la description des facteurs dont il faut tenir compte lors d'une analyse GWA. Ainsi, la structuration de la population peut amener à considérer comme significative une association génotype/phénotype alors qu'en réalité elle ne l'est pas, ce sont les **faux-positifs**.

Une autre limite concerne cette fois les **allèles mineurs**. En effet, la puissance de détection d'une association est liée à la fréquence des allèles. Ainsi, l'association entre le trait d'intérêt et les polymorphismes dont l'allèle mineur est faible ne sera pas détectée (Myles *et al.*, 2009; Xu *et al.*, 2017).

Enfin, l'**hétérogénéité génétique et allélique** peuvent entraîner une perte de puissance dans la détection des QTL. Ainsi, un même phénotype peut être associé à des locus différents conduisant à une hétérogénéité génétique (Fig. R.26a). Sur le même principe, des polymorphismes différents à un même locus peuvent conduire au même phénotype, il s'agit de l'hétérogénéité allélique (Fig. R26b) (Bergelson & Roux, 2010).

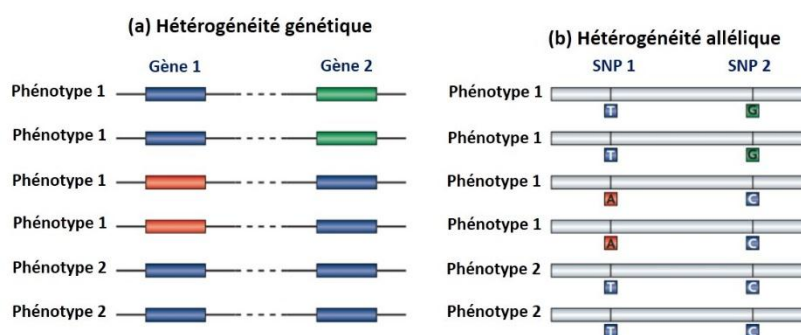


Figure R.26 | Hétérogénéité génétique et allélique (D'après Bergelson & Roux, 2010)
(a) Hétérogénéité génétique quand un même phénotype est en relation avec l'expression de différents gènes. Par exemple, le phénotype 1 peut être le résultat de l'expression du gène 1 (allèle rouge) ou du gène 2 (allèle verte) **(b) Hétérogénéité allélique** quand l'expression de deux allèles sur le même gène (mutations T-> A et C -> G conduisent au même phénotype 1.

2.5 GWA un outil pour l'étude de l'architecture génétique des interactions plantes/agents pathogènes

La première étude publiée concernant l'utilisation du GWA pour mettre en évidence des gènes de résistance a été menée sur le pathosystème *A. thaliana/Pseudomonas syringae* par Aranzana et ses collaborateurs en 2005. Ainsi, parmi les **35 études**, identifiant des QTL en réponse à un agent pathogène par GWA, recensées par Bartoli & Roux (2017) près de la moitié ont été publiées pendant ces deux dernières années. Le boom de ces deux dernières années est certainement en lien avec le développement des NGS qui permettent de disposer d'un set de marqueurs suffisamment importants pour mener ces analyses de génétique d'association.

En ce qui concerne les caractéristiques des pathosystèmes analysés, 2/3 des études impliquent des champignons les reste ont été menées sur des bactéries et des oomycètes. Les plantes d'étude appartiennent quant à elles à 11 espèces différentes réparties dans 3 grandes familles botaniques (*Brassicaceae*, *Fabaceae*, *Poaceae*). D'ailleurs, un peu plus de 60% des pathosystèmes étudiés impliquent une plante à intérêt agronomique. Cela souligne l'impact économique important de ces agents pathogènes sur les plantes cultivées et l'urgence qu'il y a à trouver de nouvelles sources de résistances pour pouvoir lutter efficacement contre ces maladies. Deux plantes modèles sont également utilisées : *A. thaliana* (~30% des études) et *M. truncatula* (~3% des études). Enfin les conditions environnementales dans lesquelles ces études ont été menées concernent pour la majorité d'entre elles des lieux en conditions contrôlées (chambre de culture, serres). Les autres ont été effectuées au champ et pour uniquement trois d'entre elles ces deux conditions environnementales complémentaires ont été utilisées (Chang *et al.*, 2016 ; Desgroux *et al.*, 2016 ; Rinker *et al.*, 2016). Un dernier point intéressant souligné par Bartoli & Roux (2017) concerne le fait qu'uniquement 6 de ces 35 études comprennent des analyses fonctionnelles des gènes candidats dans les régions génomiques identifiées par GWA. Cela pourrait s'expliquer par le fait que l'objectif principal de ces études de GWA est de pouvoir identifier des marqueurs qui pourraient être utilisés dans des stratégies de sélection assistée par marqueurs qui font partie intégrante des programmes de sélection. Cela est d'autant plus vrai dans les études conduites sur des plantes cultivées (Bartoli & Roux, 2017).

E. Acteurs de cette étude : *Arabidopsis thaliana* et le Turnip mosaic virus (TuMV)

L'identification de nouveaux facteurs de l'hôte impliqués dans les interactions plante/potyvirus fait partie des axes de recherche majeurs de l'équipe Virologie de l'UMR BFP. Il a déjà été démontré qu'*A. thaliana* constitue un excellent système biologique pour étudier ces interactions. Différents pathosystèmes sont étudiés au sein de l'équipe. Dans le cadre de cette thèse c'est le pathosystème naturel *A. thaliana* et Turnip mosaic virus (TuMV) qui a été analysé dans l'objectif de caractériser de façon fine cette interaction à travers un phénotypage large de la réponse et d'utiliser ces données de phénotypage pour décrire l'architecture génétique de la réponse d'*A. thaliana* au TuMV. La diversité naturelle d'*A. thaliana* a été utilisée avec des études menées à la fois en conditions extérieures et en conditions contrôlées. Les caractéristiques majeures des deux « acteurs » de ce pathosystème naturel ont été rassemblées et décrites ci-dessous.

1 Modèle d'étude : *Arabidopsis thaliana*

1.1 Descriptions générales

Arabidopsis thaliana (L.) Heyhn (2n = 10) communément appelée l'arabette des dames est une petite plante annuelle appartenant à la famille des *Brassicaceae*. Le genre *Arabidopsis* comprend neuf espèces différentes et huit sous-espèces. Ces neuf espèces sont majoritairement localisées en Europe, seule *A. thaliana* présente une distribution mondiale occupant des habitats ouverts ou perturbés, elle est souvent présente sur des sols sablonneux ou argileux, elle a déjà été retrouvée à plus de 4000m d'altitude (Koornneef *et al.*, 2004; Shindo *et al.*, 2007). Différentes études de phylogénétique moléculaire ont été menées afin d'essayer d'identifier les mécanismes à l'origine de sa distribution géographique actuelle. Ainsi, *A. thaliana* aurait colonisé l'Europe centrale et septentrionale depuis l'Asie avec l'influence également d'un refuge supplémentaire du Pléistocène méditerranéen (principalement la péninsule Ibérique) (Sharbel *et al.*, 2000) (Fig. R.27). La propagation d'*A. thaliana* en Amérique du Nord et en Asie de l'Est serait quant à elle en lien avec les activités humaines (Sharbel *et al.*, 2000; Schmid *et al.*, 2006).

La distribution géographique d'*A. thaliana* dépend fortement des conditions climatiques. En effet, une forte corrélation entre sa distribution et les gradients de températures et précipitations a été mise en évidence. Ainsi, des températures basses au printemps et en automne peuvent limiter son aire de répartition en Europe du Nord alors que des températures élevées et de faibles précipitations tout au long de l'année seraient limitantes pour sa distribution en Afrique du Nord, en Asie du Sud-Est et en Asie centrale (Hoffmann, 2002; Shindo *et al.*, 2007).

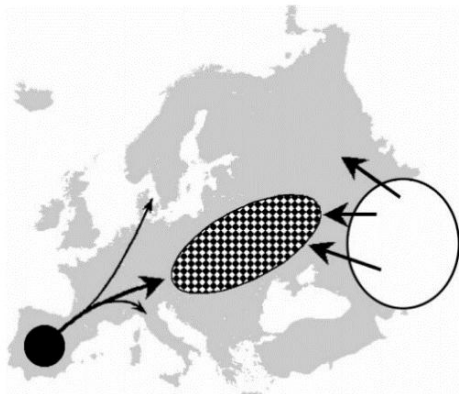


Figure R.27 | Scénario de la colonisation postglaciaire de l'Europe par *A. thaliana* (D'après Sharbel *et al.*, 2000).

Colonisation postglaciaire de l'Europe à partir de la péninsule Ibérique (en noir) et d'Asie centrale (en blanc) = zone centrale (en quadrillé). Les petites flèches à partir de la péninsule Ibérique correspondent à la colonisation de la Scandinavie et de l'Italie.

Cette large distribution géographique d'*A. thaliana* est certainement liée aux caractéristiques de son cycle de vie et en particulier au moment de la germination et à l'initiation de la floraison. Ainsi, deux catégories de plantes peuvent être décrites sur la base de ces deux traits : les accessions annuelles hivernales qui vont germer à la fin de l'été/début de l'automne, passent l'hiver sous la forme végétative (rosettes) puis initient la floraison au printemps. Les accessions annuelles estivales vont pouvoir réaliser l'ensemble de leur cycle de vie au cours de l'été tant que les conditions sont favorables pour la production de graines (Fig. R.28) De façon générale, les accessions du nord de l'Europe sont des plantes annuelles hivernales alors que dans le sud et l'est de l'Europe les deux catégories d'accessions sont retrouvées (Koornneef *et al.*, 2004; Shindo *et al.*, 2007).

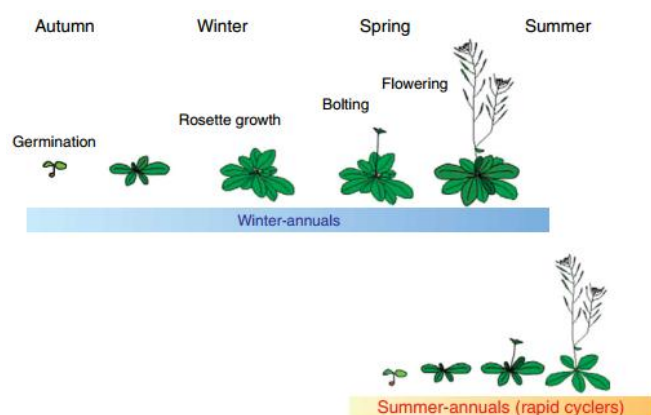


Figure R.28 | Cycle de vie des accessions d'*A. thaliana* annuelles hivernales et estivales (D'après Shindo *et al.*, 2007)

- (a) Accession annuelles hivernales qui vont germer à la fin de l'été/début de l'automne, passage de l'hiver sous la forme de rosette puis initiation de la floraison au début du printemps
 (b) Accessions annuelles estivales capable d'effectuer le cycle de vie au cours de la même saison estivale tant que les conditions restent optimales pour la production de graines

1.2 *A. thaliana*, une plante d'étude modèle

Il y a 75 ans, Friedrich Laibach proposait pour la première fois d'utiliser *A. thaliana* comme un organisme modèle pour les études en génétique, mais ce n'est que dans les années 80, qu'elle fût effectivement considérée comme tel (Meyerowitz, 2001). Ses caractéristiques telles que son petit génome (5 chromosomes, ~120 Mb), son cycle de vie court, la facilité de croisement ont conduit au développement de la recherche sur *A. thaliana*. Ainsi, une étude menée sur la base de données BIOSIS a recensé 54 033 papiers publiés sur ces 50 dernières années en lien avec cette plante modèle (publications dans lesquelles « *A. thaliana* » apparaît dans le titre, le résumé ou dans les mots clés) (Fig. R.29). Il est intéressant de noter que cette « collection » de papiers n'est pas uniquement destinée/utilisée par la communauté travaillant sur *A. thaliana*. Ainsi, 41 682 papiers publiés sur les 50 dernières années ont été au moins cités une fois dont 15 388 ont été cités sur des papiers où *A. thaliana* n'était pas le sujet principal de l'étude. Cela permet de souligner le fait que les recherches menées sur *A. thaliana* ont été et restent essentielles à la découverte de divers mécanismes chez les plantes (Provart *et al.*, 2016). Afin d'avoir une vision plus large des recherches menées sur *A. thaliana* et de leur impact en biologie végétale les publications de Sivasubramanian et ses collaborateurs (2015) ainsi que Provart et ses collaborateurs (2016) sont à consulter (Sivasubramanian *et al.*, 2015; Provart *et al.*, 2016).

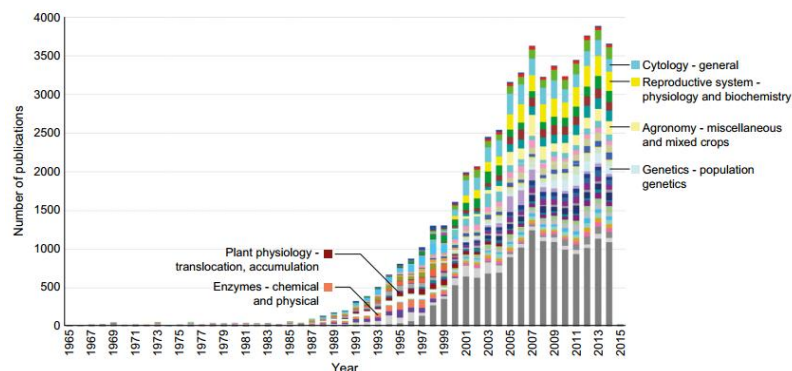


Figure R.29 | Analyse des divers travaux publiés sur *A.thaliana* à l'aide de la base de données BIOSIS (D'après Provart *et al.*, 2016)

Recensement des 54 033 publications sur les 50 dernières années. L'ordonnée correspond au nombre de publications/année. Les couleurs correspondent aux différents domaines d'étude. Les différentes publications/citations peuvent être retrouvées de façon interactive à cette adresse : <http://bar.utoronto.ca/50YearsOfArabidopsis/>.

1.3 Exploration de la variabilité naturelle d'*A. thaliana*

La variabilité naturelle retrouvée parmi les différentes accessions d'*A. thaliana* a constitué un point d'intérêt majeur pour essayer de caractériser les bases génétiques associées à une variabilité phénotypique. Cela passe par l'utilisation des larges ressources génétiques disponibles. *A. thaliana* a été la première plante et le troisième organisme pluricellulaire (après *C. elegans* et *D. melanogaster*) à être complètement séquencé (séquençage achevé en 2000 sur l'accèsion Col-0 - The Arabidopsis Genome Initiative 2000). Le séquençage de 20 autres accessions a contribué à augmenter les ressources génomiques disponibles (Clark *et al.*, 2007). Le programme 1001 Genomes a été lancé en 2008 avec pour objectif de caractériser la variabilité des génomes d'au moins 1001 accessions d'*A. thaliana* (<http://1001genomes.org>). Entre temps, la génétique d'association (GWA) est devenue l'approche la plus populaire afin d'étudier la variabilité génétique. *A. thaliana* a d'ailleurs été un des premiers organismes non-humain pour lequel les études de GWA sont devenues possibles (Atwell *et al.*, 2010). Ainsi, une puce Affymetrix de 250k SNPs a été utilisée pour génotyper une collection de 1307 accessions différentes (collection RegMap – Regional Mapping Lines) (Horton *et al.*, 2012). Plus récemment une étude utilisant la technologie Illumina a rassemblé 1 135 accessions mondiales avec une distribution de ces accessions qui se veut encore plus mondiale que la collection RegMap (avec notamment des génotypes originaires de Suède et de la péninsule Ibérique) contribuant à l'enrichissement des ressources génétiques déjà disponibles (Alonso-Blanco *et al.*, 2016). La séquence de référence a permis l'annotation des gènes. Ainsi, la base de données TAIR10 compte 27 416 gènes codant pour des protéines, 4827 pseudogènes ou éléments transposables et 1359 ncRNA (« non-coding RNA ») (<https://www.arabidopsis.org/>).

1.4 *A. thaliana* vers la découverte de la fonction des gènes

La progression notamment du programme 1001 Genomes a permis d'identifier à travers différentes études, un certain nombre de gènes candidats dont il est intéressant à présent de caractériser la fonction. Afin de pouvoir mener à bien cet objectif, il fallait pouvoir tester des plantes présentant des mutations sur ces gènes sélectionnés. C'est ainsi qu'a été développée la première grande **collection de mutants d'insertion (T-DNA)** (Feldmann & Marks, 1987). Le

développement de méthodes de séquençage à haut-débit a permis d'identifier les sites d'insertion sur l'ensemble de la collection (Sessions, 2002). L'ensemble des données ont été répertoriées dans des bases de données et des stocks de graines créés. Cela a permis d'avoir accès et de commander de façon spécifique le mutant correspondant au gène d'intérêt à tester. Des avancées considérables ont donc été réalisées dans la recherche de la fonction des gènes à la fois chez *A. thaliana* mais également chez d'autres espèces à intérêt agronomique puisque l'utilisation de mutants d'*A. thaliana* est devenu est moyen rapide et courant de tester la fonction de gènes orthologues identifiés (Provart *et al.*, 2016).

Des **collections de mutants TILLING** (« Targeting Induced Local Lesions in Genomes ») ont également été développées au départ chez *A. thaliana*. Comme cette technologie ne requiert pas d'étape de transformation, elle a été rapidement transférée chez d'autres espèces végétales (Till *et al.*, 2003).

Enfin, les **nouvelles technologies** telles que l'utilisation de nucléases zing-finger, de TALENs (« Transcription Activator Like-Effectors ») ou de façon plus récente de CRISPR-Cas9 viennent compléter les outils disponibles en génétique inverse. Tout cela permet de mettre l'accent sur le rôle crucial d'*A. thaliana* dans le développement et l'optimisation de ces différents outils à disposition pour l'ensemble de la communauté scientifique (Provart *et al.*, 2016).

1.5 Des domaines de recherche au transfert de connaissances

L'utilisation de la plante modèle *A. thaliana* a transformé différents domaines de la recherche végétale ces 50 dernières années et son impact est toujours aussi important. L'ensemble des ressources génétiques et génomiques ainsi que d'importantes données de transcriptomique, protéomique et métabolomique recensées dans diverses conditions (environnements, stades de développement) ont fait d'*A. thaliana* une excellente source de référence pour les études concernant les plantes cultivées. Les connaissances acquises ont facilité la compréhension de mécanismes dans divers domaines (Sivasubramanian *et al.*, 2015). Le dernier congrès réunissant la communauté scientifique travaillant sur *A. thaliana*, ICAR2017 (28th International Conference on *Arabidopsis* Research in St Louis, Missouri - <http://icar2017.arabidopsisresearch.org/>) permet de se rendre compte des thématiques de recherche actuelles et des nouvelles orientations de travail prises pour faire avancer les connaissances en biologie végétale dans les prochaines années. Ainsi la compréhension des mécanismes en lien avec le développement, la réponse à des stress reste une priorité avec un rôle important pour les régulations hormonales intervenant à différentes étapes de ces phénomènes. L'épigénétique est devenu une discipline incontournable afin d'étudier les mécanismes moléculaires qui modulent l'expression du patrimoine génétique en fonction des conditions. Le grand nombre de données générées ces dernières années a permis le développement d'approches de modélisation et de bio-informatique qui sont primordiales à leur compréhension. L'objectif final étant de pouvoir transférer toutes ces compétences et connaissances à des plantes à intérêt agronomique. D'ailleurs lors de congrès dans un certain nombre de présentations, *A. thaliana* ne constituait plus le sujet principal d'étude mais était considérée comme un outil pour faire avancer la recherche chez d'autres espèces végétale. Ce rôle d'*A. thaliana* dans le transfert des connaissances se retrouvait d'ailleurs dans le titre du congrès « *From basic dicoverly to translation* ». Il faut bien évidemment garder à l'esprit qu'en tant que plante modèle, *A. thaliana* ne permettra pas de comprendre tous les mécanismes

biologiques mais constituera une bonne référence pour la comparaison. Finalement la bonne question à se poser serait « Est-ce que ce qui est observé ressemble ou pas à ce qui est connu chez *A. thaliana* ? ». Comme le soulignait Provart et ses collaborateurs (2016) au sujet de la recherche sur *A. thaliana* : « Ceci n'est pas la fin, ni même le commencement de la fin, mais c'est peut-être la fin du commencement (W.Churchill) » (Provart *et al.*, 2016).

1.6 *A. thaliana* un hôte compatible pour de nombreux virus

Différentes études basées sur différentes méthodes d'inoculation (mécanique, agro-infiltration, biolistique) ont permis de mettre en évidence le fait qu'*A. thaliana* constitue un hôte favorable pour le développement de 40 espèces virales appartenant à 18 genres différents. La majorité de ces virus (75%) sont des virus à ARN positif appartenant à 12 genres différents. Il existe en revanche peu de connaissances sur la capacité de ces virus à infecter naturellement *A. thaliana* (Ouibrahim & Caranta, 2013). Une étude a néanmoins recensé pendant 4 ans dans des conditions naturelles la présence de 5 virus différents chez *A. thaliana* (Pagan *et al.*, 2010). Si la plupart des infections virales semblent avoir un impact sur la croissance, la reproduction des plantes avec des symptômes caractéristiques de chaque virus, de façon surprenante certaines interactions *A. thaliana*/virus sont asymptomatiques. C'est par exemple le cas du *Lettuce mosaic virus* (LMV) (Revers *et al.*, 2003). Ces observations pourraient être reliées aux connaissances obtenues récemment lors d'une étude de métagénomique où il a été mis en évidence que de nombreux phytovirus ne provoquent pas de symptômes chez leurs plantes hôtes (Roossinck, 2012).

L'analyse de la réponse de différentes accessions à une infection virale a permis de mettre en évidence des gènes de résistance à neuf genres viraux différents. La plupart de ces résistances ont été mises en évidence chez Col-0 (accession référence d'*A. thaliana* largement utilisée dans ce type d'études) en réponse à des virus à ARN. Ces gènes identifiés peuvent assurer une résistance totale à l'infection, limiter la multiplication au niveau des sites d'infection, bloquer sa propagation en systémie ou simplement diminuer l'accumulation virale et/ou limiter la sévérité des symptômes. *A. thaliana* est d'ailleurs considérée comme une source pour l'identification de nouvelles résistances aux virus (Ouibrahim & Caranta, 2013).

2 Un potyvirus : le *Turnip mosaic virus* (TuMV)

2.1 Les Potyvirus

Les virus du genre *Potyvirus* appartiennent à la famille des *Potyviridae*. Cette famille est composée de huit genres et regroupe 158 espèces. Il s'agit de la deuxième plus grande famille de virus après les *Geminiviridae* (Wylie *et al.*, 2017). Les potyvirus infectent plus de 60 familles végétales. Ils peuvent causer d'importants dégâts ce qui en fait à leur actuelle un enjeu économique majeur. D'ailleurs, deux potyvirus le *Potato virus Y* (PVY) et le *Plum pox virus* (PPV) font partie de la liste des dix virus de plante les plus importants d'un point de vue scientifique et économique (Scholthof *et al.*, 2011).

Les potyvirus sont transmis par l'intermédiaire de pucerons via un mode de transmission non-persistant. L'acquisition du virus par le puceron est rapide, le virus ne se multiplie pas dans l'insecte et la durée de rétention des particules virales dans le stylet du puceron est courte.

Cette acquisition rapide par les pucerons fait que les potyvirus peuvent être transmis très efficacement par leur vecteur sur le terrain. De plus, chaque potyvirus peut être transmis par différentes espèces de pucerons et chaque espèce de puceron a la capacité de transmettre plusieurs potyvirus. Certains potyvirus peuvent être transmis par la graine et certains d'entre eux peuvent même être transmis via le pollen (Rajamäki *et al.*, 2004)

Les potyvirus sont des virus à ARN à polarité positive (ARNss(+)) d'environ 10 kb. Lorsque ce simple brin d'ARN est recouvert d'environ 2000 protéines de capsid (CP) il constitue un virus mature, le virion. Cette molécule d'ARNss(+) est polyadénylée à l'extrémité 3' et est liée de façon covalente à une protéine virale la VPg (« Viral Protein genome-linked ») à son extrémité 5'. Le génome des potyvirus contient un ORF majeur, qui après décapsidation, sera traduit en une polyprotéine (~340-370 kDa) qui sera par la suite clivée par une activité auto-catalytique en 10 protéines fonctionnelles (Urcuqui-Inchima *et al.*, 2001). Récemment, un second ORF (PIPO) chevauchant la séquence qui code pour la protéine P3 a été identifié dans le génome d'environ 50 espèces de potyvirus. Cet ORF sera traduit selon un mécanisme de décalage du cadre de lecture et est à l'origine d'une protéine de fusion P3N-PIPO (Chung *et al.*, 2008). Plusieurs rôles à différentes étapes du cycle viral ont déjà été décrits pour chacune des protéines virales des potyvirus (Fig. R.30) (Urcuqui-Inchima *et al.*, 2001; Revers & García, 2015).

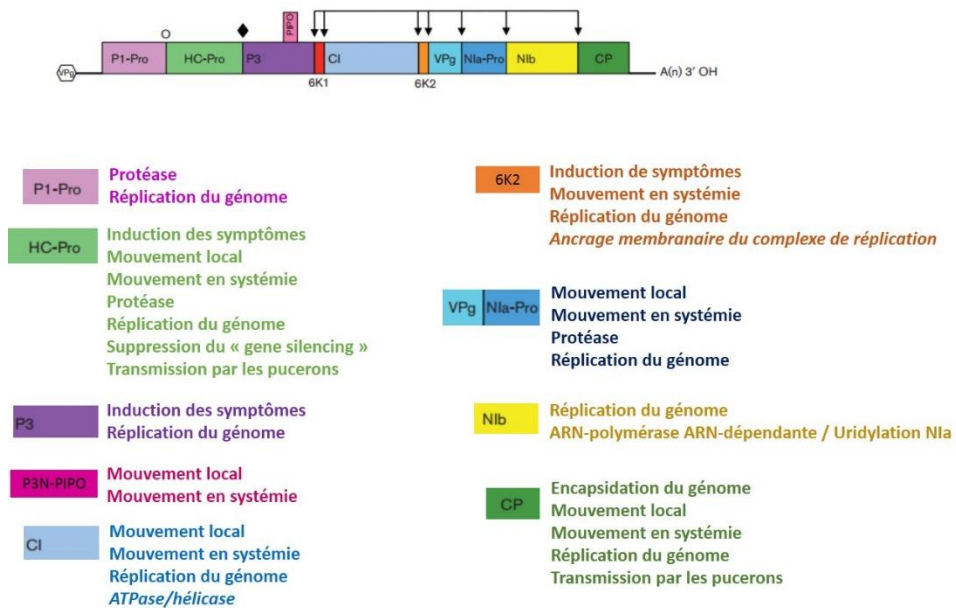


Figure R.30 | Organisation du génome d'un potyvirus et différentes fonctions des protéines virales codées (D'après Wylie *et al.*, 2017)

VPg, viral protein genome-linked ; P1-Pro, protein 1 protease ; HC-Pro, helper component protease ; P3, protein 3 ; PIPO, pretty interesting Potyviridae ORF ; 6K, six kilodalton peptide ; CI, cytoplasmic inclusion ; Nia-Pro, nuclear inclusion A protease ; Nib, nuclear inclusion B RNA-dependent RNA polymerase ; CP, coat protein. Les sites de clivage de P1-Pro (o), HC-PRO (♦) et Nla-Pro (↓) sont représentés.

Les différentes étapes du cycle viral des potyvirus ont largement été documentées dans différentes revues (Rajamäki *et al.*, 2004; Mäkinen & Hafrén, 2014; Ivanov *et al.*, 2014; Revers & García, 2015). Les potyvirus entrent dans les cellules de la plante par des blessures créées par la piqûre lors de la prise de nourriture des pucerons. L'étape suivante est la décapsidation de la particule virale avec le relargage dans le cytoplasme de l'ARN viral traduit en une

polyprotéine. Cette polyprotéine va être clivée sous l'action de trois protéases en 10 protéines différentes. Le décalage du cadre de lecture en lien avec le deuxième ORF, PIPO est à l'origine de la onzième protéine codée (Fig. R.31). Après traduction de tous ces facteurs viraux, il y a la formation de complexes de réplication viraux où aura lieu la synthèse de l'ARN viral. Ces nouvelles copies d'ARN viral pourront **i)** servir de matrice pour la traduction de nouvelles protéines virales, **ii)** être transportées jusqu'aux cellules voisines via l'adressage de vésicule aux plasmodesmes, **iii)** être encapsidées afin de constituer de nouveaux virions qui pourront être transmis à une nouvelle plante après acquisition par un puceron, **iv)** être dégradées (Rajamäki *et al.*, 2004; Ivanov *et al.*, 2014).

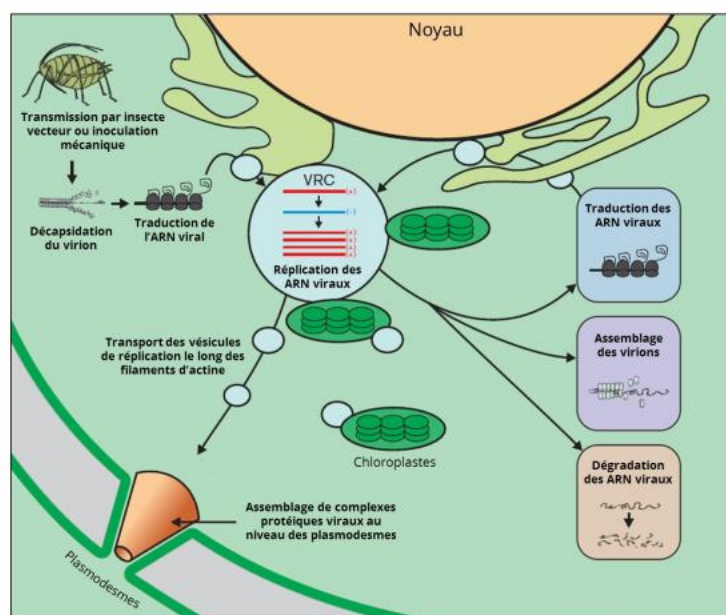


Figure R.31 | Principales étapes du cycle viral des potyvirus (D'après Ivanov *et al.*, 2014)

Entrée des particules virales dans la cellule suite à une piqûre par un puceron suivie par la décapsulation du virion. L'étape suivante est la traduction de l'ARN viral en une polyprotéine qui sera clivée en 11 protéines différentes. L'ARN viral sera répliqué au niveau de complexes de réplication (VRC pour viral replication complex). Ces nouveaux ARN synthétisés pourront **i)** être transportés le long des filaments d'actine jusqu'au plasmodesmes afin de passer aux cellules voisines, **ii)** constituer de nouveaux virions matures, **iii)** être traduits pour donner de nouvelles protéines virales, **iv)** être dégradés.

2.2 Le Turnip mosaic virus (TuMV)

Le TuMV est un membre de la famille des *Potyviridae* avec une distribution mondiale infectant des plantes cultivées et sauvages essentiellement de la famille des *Brassicaceae*. D'ailleurs, il s'agit du seul potyvirus connu à ce jour comme infectant des brassicacées. Il est capable d'infecter une large gamme d'hôtes comprenant 318 espèces, 156 genres et 43 familles différentes. Le TuMV est transmis par 89 espèces de pucerons selon le mode non persistant, en particulier par *Myzus persicae* et *Brevicoryne brassicae* (Walsh & Jenner, 2002). Les principaux symptômes induits par le TuMV sont un retard de croissance, de la mosaïque/des jaunissements/des boursoufflures/des recroquevillements sur les feuilles en systémie (Martín Martín *et al.*, 1999). Des symptômes caractéristiques observés chez le navet (*B.rapa*) et chez *A. thaliana* sont présentés en Figure R.32. Le développement des symptômes dépend de la virulence de l'isolat en question, de la sensibilité de l'hôte et des conditions environnementales. Ainsi, la température semble avoir un effet assez fort sur le

développement de l'infection que ce soit sur sa propagation dans la plante (mouvement en systémie) que sur le développement des symptômes. Une étude menée par Chung et ses collaborateurs (2015) sur le chou chinois (*B.campestris*) montre une augmentation linéaire de la vitesse d'infection ainsi qu'une augmentation de la vitesse d'apparition et de la sévérité des symptômes à partir de 23°C (Chung *et al.*, 2015).

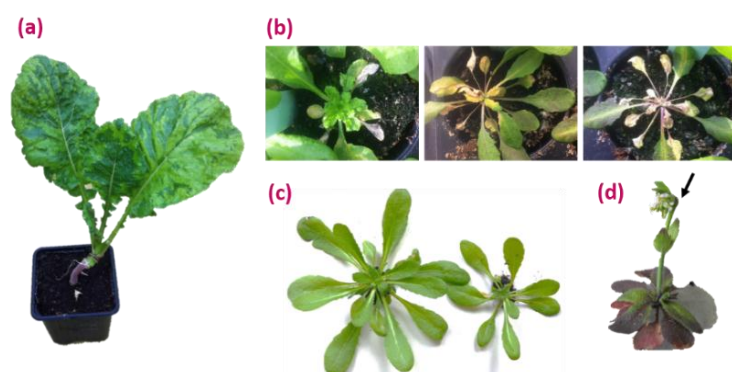


Figure R.32 | Diversité des symptômes causés lors d'une infection par le TuMV chez *B. rapa* et *A. thaliana* (Photo B.Rubio)

(a) Symptômes liés à une infection par le TuMV chez *B. rapa* (15 dpi). Mise en évidence de mosaïques/de boursoufflures au niveau des feuilles en systémie. La croissance des plantes est également affectée (b) Symptômes liés à l'infection par le TuMV chez différentes accessions d'*A. thaliana*. La sévérité des symptômes varie en fonction de la sensibilité des accessions mais varie aussi au cours de l'infection virale. Mise en évidence de recroquevillements/boursoufflures, de jaunissements (c) Mise en évidence des retards de croissance causés par l'infection par le TuMV chez *A. thaliana* : à gauche plante non infectée, à droite plante infectée (d) Mise en évidence d'un symptôme particulièrement caractéristique chez *A. thaliana* infectée par le TuMV : la crosse au niveau de la hampe (↓)

Le TuMV fait partie des 70 potyvirus dont un génome de référence est recensé dans la base de données GenBank. Une étude phylogénétique a décrit 11 lignées différentes chez les potyvirus parmi lesquelles se trouve le groupe « TuMV » qui comprend le TuMV ainsi que le *Japanese yam mosaic virus* (JYMV), le *Narcissus yellow stripe virus* (NYSV) et le *Scallion mosaic virus* (ScMV) tous les trois infectant des monocotylédones. Cette phylogénie montrerait que la capacité à infecter des monocotylédones serait plutôt un caractère ancestral du groupe « TuMV ». La capacité du TuMV à infecter des dicotylédones serait plutôt en lien avec une adaptation récente (Gibbs & Ohshima, 2010). Dans leur étude, Nguyen et ses collaborateurs (2013) ont identifié un cluster composé de virus TuMV-like issus d'orchidées européennes (TuMV-OM) qui semblent avoir les caractéristiques susceptibles d'être trouvées dans les virus à partir desquels les virus TuMV infectant les Brassicaceae (TuMV-BI) sont supposés avoir évolué. L'émergence des lignées TuMV-BI aurait eu lieu selon leur estimation il y a environ 1000 ans (Nguyen *et al.*, 2013).

Le TuMV est un virus très variable, la population de TuMV évolue et se différencie rapidement. Des études sur la variabilité génétique et moléculaire ont permis de comprendre certaines caractéristiques importantes liées à sa biologie, aux changements de virulence, à sa gamme hôtes, à sa distribution géographique. Des analyses des séquences des gènes codant pour la protéine 1 (P1) et les protéines de capsid (CP) menées sur une collection d'isolats mondiaux de TuMV ont permis de caractériser différents clusters sur la base des différences génétiques observées et des localisations géographiques. Ainsi, quatre clusters différents ont été définis : **basal-B** (Brassica), **basal-BR** (Brassica/Raphanus), **Asian-BR** and **world-B**. Le cluster basal-B est le plus variable, les isolats proviennent d'espèces appartenant ou non au genre *Brassica*. Les

isolats dans les clusters basal-BR et Asian-BR semblent plus adaptés pour infecter des plantes du genre *Brassica* et *Raphanus*. La plupart des isolats européens appartiennent au cluster world-B où les isolats infectent des espèces du genre *Brassica* (Ohshima *et al.*, 2002, 2010).

Chez *A. thaliana*, le TuMV est capable d'induire des symptômes sévères (Fig. R.32). L'interaction entre ce potyvirus et 106 accessions différentes d'*A. thaliana* a été testée en 1999 par l'équipe de Martin-Martin. La quantification de la charge virale par ELISA permet de classer ces accessions en trois catégories **i)** les accessions sensibles (33), **ii)** les accessions avec une variabilité de réponse (69) et **iii)** les accessions résistantes (4). Parmi les quatre accessions considérées comme résistantes, seule l'accession Bay-0 présente une réponse stable et répétable (les trois autres accessions accumulent le virus au niveau des feuilles inoculées) (Martín Martín *et al.*, 1999). Cette forte sensibilité d'*A. thaliana* au TuMV pourrait être liée aux caractéristiques évolutives et écologiques d'*A. thaliana*. Quoi qu'il en soit, le TuMV fait partie des virus retrouvés comme infectant naturellement *A. thaliana* ce qui rend ce pathosystème particulièrement intéressant pour la recherche de facteurs de l'hôte impliqués dans la réponse à une infection virale (Pagan *et al.*, 2010).

- Chapitre 1 -

**Architecture génétique de la réponse d'*A. thaliana*
au TuMV en conditions extérieures et phénotype fin
de la réponse métabolique**

Présentation Générale du Chapitre 1

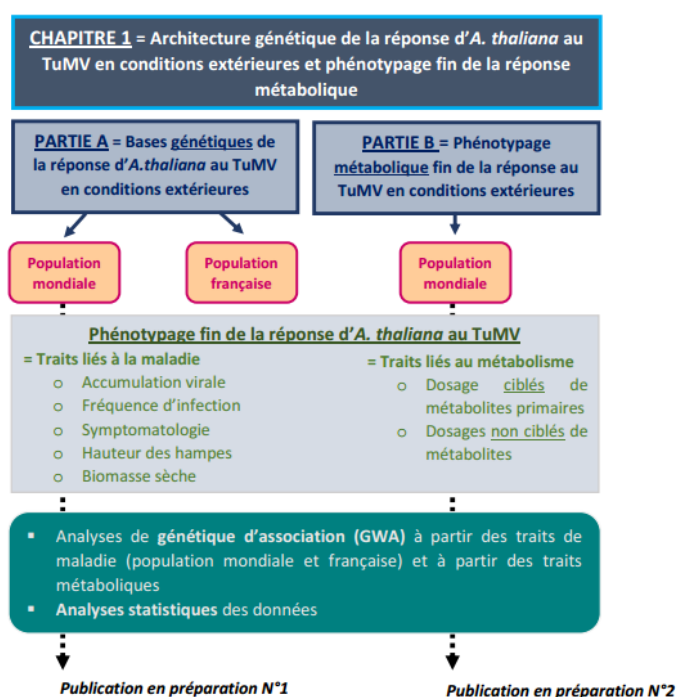
Ce premier chapitre concerne l'ensemble des études réalisées en conditions extérieures au cours de ce travail de thèse. Les principaux objectifs sont **i)** de décrire l'architecture génétique de la réponse d'*A. thaliana* au TuMV dans ces conditions et **ii)** d'effectuer un phénotypage fin de cette interaction en caractérisant notamment différentes variables métaboliques. Ce chapitre est donc divisé en deux parties.

Une première partie, **partie A**, regroupe les analyses en lien avec les bases génétiques de la réponse d'*A. thaliana* au TuMV en conditions extérieures. Dans cette partie A, plusieurs **traits de maladie** ont été utilisés pour phénotyper cette interaction. Deux populations différentes d'*A. thaliana* ont été étudiées : une **population mondiale** et une **population locale française**. L'analyse de l'**architecture génétique** a été menée via des analyses de **génétique d'association** en utilisant les différents traits de maladie. Les résultats obtenus sur l'étude avec la population mondiale sont présentés sous la forme d'une publication en préparation.

La deuxième partie, **partie B**, concerne le **phénotypage métabolique fin** de la réponse d'*A. thaliana* au TuMV dans ces conditions extérieures. Cette partie B, regroupe différentes analyses menées sur différentes accessions de la population mondiale d'*A. thaliana*. Ainsi, dans un premier temps, les résultats relatifs aux analyses métaboliques (métabolisme primaire et secondaire) obtenus sur deux années (2014-2015) et sur un large panel d'accessions (130 et 29 accessions respectivement sur les deux années) sont présentés sous la forme d'une publication en préparation.

Cette première étude a été complétée par des **analyses cinétiques du contenu en métabolites primaires** sur un nombre plus réduit d'accessions.

Enfin, l'architecture génétique de la réponse métabolique d'*A. thaliana* au TuMV en conditions de stress multiples a été recherchée à partir des données du contenu en métabolite primaire du set de 130 accessions étudiées en 2014.



Chapitre 1 - PARTIE A

**Bases génétiques de la réponse d'*A. thaliana*
au TuMV en conditions extérieures**

A-1. Architecture génétique de la réponse d'*A. thaliana* au TuMV en conditions extérieures

Cette partie est présentée sous la forme d'un article scientifique.

Field experiments reveal new loci involved in *A. thaliana* and *Turnip mosaic virus* (TuMV) interaction

Bernadette Rubio¹, Patrick Cosson¹, Mélodie Caballero¹, Frédéric Revers², Joy Bergelson³, Fabrice Roux⁴, Valérie Schurdi-Levraud^{1‡}

¹INRA, Université de Bordeaux, UMR 1332 BFP, BP81, 33883 Villenave d'Ornon cedex, France. INRA, Université de Bordeaux

²UMR 1202 BIOGECO, 69 Route d'Arcachon 33612 CESTAS Cedex, France.

³Ecology & Evolution, University of Chicago, 1101 E 57th St Chicago IL 60637, USA.

⁴CNRS, INRA, Université de Toulouse, LIPM, Castanet-Tolosan, France.

‡ Author for correspondence: Valérie Schurdi-Levraud. [Contact: +33557122377](tel:+33557122377) or valerie.schurdi-levraud@inra.fr

Keywords: *Arabidopsis thaliana*, field experiment, multistress, genetic architecture, genome-wide association mapping, *Turnip mosaic virus* (TuMV), virus

Introduction

In both natural and agricultural settings, plants are exposed to combinations of biotic and abiotic stresses. Thus, due to their sessile lifestyle, plants have evolved mechanisms to face a wide range of stresses such as heat, drought, salinity, pests and pathogens (Osakabe *et al.*, 2014; Liu *et al.*, 2015; Julkowska & Testerink, 2015; Motaung *et al.*, 2017). Recent studies have revealed that the responses of plant to combinations of biotic and abiotic stresses cannot be predicted based on the additive response of plants to individual single stresses (Prasch & Sonnewald, 2015; Coolen *et al.*, 2016). There is thus an urgent need to understand the mechanisms underlying how plants respond to combined abiotic and biotic stresses. To date, research has focused on the joint effects of abiotic stresses such as heat, drought and/or salinity (Rizhsky *et al.*, 2004; Suzuki *et al.*, 2014). This work has revealed both negative and positive interactions, the directionality of which depends on the particular stresses being combined as well as on physiological and molecular crosstalk (Kissoudis *et al.*, 2016). Combined biotic and abiotic stresses have been more sparsely studied, perhaps due to the challenges associated with controlling living organisms to produce desired levels and timings of attack. It has been reported that abiotic stresses may limit the response of plants to pathogens due to the attenuation of resistance and/or enhanced pathogen infection (Sanogo, 2004; Triky-Dotan *et al.*, 2005; Wang *et al.*, 2005; You *et al.*, 2011). On the other hand, some have suggested that abiotic stress may lead to an increase in resistance (Wiese *et al.*, 2004; Achuo *et al.*, 2006; Rouhier & Jacquot, 2008). The mechanisms by which plants regulate and prioritize their adaptive response when exposed to multi-stresses is largely unknown, although it is clear that different stress signaling pathways in plants are interconnected (Coolen *et al.*, 2016).

In the face of global change, it is typically assumed that the natural ecological equilibrium will be altered by the emergence of new infectious agents. This threat is even greater in human created agrosystems where the spread of highly damaging pathogens is facilitated, among other reasons, through global trade, increasing the threat for agricultural sustainability and food security

(McDonald & Stukenbrock, 2016). Among emerging pathogens, plant viruses represent a relatively large threat because global change is likely to modify diverse components of virus epidemics including : resistance to vectors or viruses, vector and virus life cycles, abundance, diversity, reservoirs (Jones, 2009) and the emergence of strains through high rates of mutations (Holmes, 2009; Elena *et al.*, 2011). Potyviruses represent 20% of known plant viruses infecting field-grown plant species and are responsible for a variety of physiological disorders that cause economically and agronomically significant losses. *Turnip mosaic virus* (TuMV) has been ranked among the five most damaging vegetable viruses worldwide (Tomlinson & A, 1987); this virus causes damages in horticultural and arable *Brassica* crops in Asia, North America and Europe (Walsh & Jenner, 2002). Infection of plants by TuMV results in symptoms such as plant stunting and developmental arrest, leaf mosaic, sawed and curly leaves, and aborted flowers (Martín *et al.*, 1999).

Thanks to its small genome (135 Mb) and the abundant molecular tools, *Arabidopsis thaliana* (*A. thaliana*) is a formidable model for plant biotechnological and genetic studies, and more especially when focusing on biotic and abiotic stresses. TuMV, as well as many other viruses have been shown to infect *A. thaliana* in laboratory conditions (Ouibrahim & Caranta, 2013). This opens the door for using *A. thaliana* to identify genes that underlie susceptibility and resistance to viral infection (Yang *et al.*, 2007). These genes can then be introduced into crop species as an ecologically sound means of enhancing resistance to viruses (Revers & Nicaise, 2014).

To date, the screening of *Arabidopsis* accessions for their responses to potyviruses infections has led to the identification of susceptible and resistance candidate genes or QTL (quantitative trait loci). Recessive resistances associated with translation factors are commonly associated with the response to potyviruses, including CIYVV (*Clover yellow vein virus*), LMV (*Lettuce mosaic virus*), TEV (*Tobacco etch virus*), PPV (*Plum pox virus*) and TuMV (Duprat *et al.*, 2002; Lellis *et al.*, 2002; Sato *et al.*, 2005; Decroocq *et al.*, 2006; Nicaise *et al.*, 2007). Recently, recessive resistances driven by a chloroplast phosphoglycerate kinase (cPGK2) have been discovered in

response to WMV (*Watermelon mosaic virus*) and PPV (Ouibrahim *et al.*, 2014; Poque *et al.*, 2015). In addition to conferring partial to complete resistance, phenotypes of long-distance movement restriction have been described. Three dominant resistance genes *RTM1*, *RTM2* and *RTM3* have been cloned (Cosson *et al.*; Chisholm *et al.*, 2000; Whitham *et al.*, 2000) and other loci exist (Cosson *et al.*, 2012). The identification of these genetic determinants have primarily been conducted through mutants screening and traditional linkage mapping, characterizing the offspring of crosses or mutagenized populations under controlled greenhouse or growth chamber conditions. Here we aimed to elucidate the genetic architecture of the response of *Arabidopsis thaliana* to the potyvirus TuMV in a realistic context that takes into account a multi-stress environment. We performed a common garden field experiment, repeated for two years. In these conditions, plants are exposed to temporal abiotic fluctuations that may affect their responses to biotic stress. Through the phenotyping of 327 accessions, we explore the natural variation to TuMV and use GWA mapping to gain insights into the genetic basis of resistance to TuMV.

Materials and Methods

Plant material

The 327 *Arabidopsis thaliana* natural accessions are worldwide accessions from the Hapmap population (http://bergelson.uchicago.edu/?page_id=790#sets). The accessions used are listed in the Supporting information (Table S1).

Virus material

Turnip mosaic virus isolate UK1 (Jenner Walsh, J.A., 1996) was routinely propagated by mechanical inoculations on turnip plants *Brassica rapa L.ssp rapa* NA FR 490001 provided by the BraCySol germplasm center (Ploudaniel, France). Turnip were sown in a professional horticultural soil (106S cope-PletrAcom) in trays maintained in a greenhouse at 20°C and 60% relative humidity. To prepare the inoculum, we introduced virus into three weeks old turnip plants by mechanically scrubbing with a carborandum, and then collected symptomatic leaves two weeks

later. Inoculum was prepared by collecting these 5 week old symptomatic leaves. One g of fresh leaves was grinded in three volumes of sodium dihydrogen phosphate dodecahydrate ($\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$) 30mM and 0.2% of diethyldithiocarbamic acid (DIECA). Inoculum was clarified through 10 min centrifugation at 13000g. Supernatant was recovered and maintained at 4°C until being used to inoculate experimental plants.

Experimental design and *Arabidopsis* growth conditions

We performed common garden experiments in each of two years, 2013 and 2014, using 301 accessions and 326 accessions respectively (Table S1). The experiments were organized in three blocks, each block being an independent randomization of one replicate per natural *A. thaliana* accession. Plants were grown in seedling trays containing 40 cells of 5.5cm of diameter. Two control accessions, Bay-0 and TOU-A1-63 were planted twice in each tray. Bay-0 is the negative control whereas TOU-A1-63 is the positive control. Seeds were sown on 24 March 2013 and 2014 in a professional horticultural soil (106S cope-PletrAcom) in a non-heated and non-lighted glasshouse. At three weeks of age, plants were acclimatized under an opened tunnel, inoculated with TuMV, and then transferred to the common garden. In 2014, we added mock-inoculated plants of each accession treated exactly as inoculated plants except for the presence of the virus (Fig. S1). Soil of the common garden had been tilled so that seedling trays could be slightly buried. Because the bottom of the wells was pierced, roots were able to reach the soil.

TuMV inoculation procedure

Arabidopsis inoculation occurred on 4 weeks-old plants with 8-10 expanded leaves corresponding to 1.09, 1.10 Boyes stages (Boyes *et al.*, 2001). Four young expanded leaves of each plants were mechanically inoculated by depositing 20µL of inoculum on each leaf and slightly scrubbing with a carborandum. Ten minutes after inoculation, plants were rinsed with water. Thirteen days after inoculation (dpi), rosette leaves above the inoculated leaves were collected in vials from Zinsser Analytic ®. In order to evaluate the concentration of the TuMV inoculum, virus was quantified

after inoculation through RNA extraction, cDNA production and Q-PCR as described in (Cosson *et al.*, 2012). After extraction, total nucleic acid was quantified spectrophotometrically. All samples were adjusted to 50 ng/μL with DEPC treated water. Reverse transcription was done using Revertaid/Ribolock reverse transcriptase kit (ThermoFisher Scientific). Quantitative PCR was performed on a Light Cycler 480 II machine (Roche Diagnostic) by using Light Cycler 480 SYBR Green I master mix and one tenth of the synthesized cDNAs. TuMV was amplified and detected by forward (5'TCAAACCATGTGTCAAACCTGC3') and reverse (5'CGAGAAAAGAGTGGCCCTAA3') primers in the capsid gene. To evaluate the relative concentration of inoculum a range dilution from 1 ng/μl to 0.001 ng/μl of a UK1 plasmid (Beauchemin & Laliberte, 2007) was used as a control in Q-PCR. Q-PCR was performed using the following cycling conditions: 95°C for 15 min, and 40 cycles of 95°C for 30 s, 59°C for 30 s and 72°C for 30 s.

Phenotyping

Viral accumulation in each individual plant in the two common garden experiments was estimated using a semi-quantitative double antibody sandwich (DAS) ELISA assay with a commercial anti-potyvirus monoclonal antibody kit (Agdia-Biofords, France, <http://www.agdia-biofords.com/>). The reaction of the substrate (p-nitrophenyl phosphate) was followed at 405 nm. Optical densities (OD) were normalized using an *A. thaliana* positive control deposited on each ELISA plate. Values of OD corresponded to the viral charge of each individual sample. For each genotype, we assessed susceptibility in each of three ways (1) by calculating the average of OD (“Viral charge”), (2) by using the maximum OD (“Viral charge (Max)”), (3) by counting the number of infected plants of each genotype relative to the total number of plants inoculated for that genotype (“Rate of infection”). In 2014, plants were also monitored for the presence or absence of symptoms (“Symptoms”). In addition, the stem length was measured for each sample (“Stem length”). To

consider the effect of infection, we normalize stem length against the average length of mock-inoculated samples of each genotype.

Treatment of phenotypic data sets

For each common garden experiment data set, missing genotypes have been removed. In addition, due to the possibility of obtaining false negative samples as a consequence of failed inoculations (Poque *et al.*, 2015), negative samples were only retained when the same negative profile was obtained in both years of experiment. After filtering our data in this way, the final data sets used for GWA analysis contained 266 genotypes in 2013 and 202 genotypes in 2014. Phenotypic data was subjected to a logarithmic transformation (base -10 logs).

Genome Wide Association mapping and Gene Ontology (GO) Term Enrichment

Genome Wide analysis was based on 216,509 SNPs evenly spaced across the genomes (1 SNP every 500 bp –(Brachi *et al.*, 2010)) of each of our accessions. The design of the Affymetrix genotyping array and genotyping protocols have been detailed elsewhere (Atwell *et al.*, 2010). GWA analysis was performed for viral traits using the linear mixed model GEMMA, which corrects for population structure (Zhou & Stephens, 2012). Only SNPs with minor allele frequency (MAF) higher than 0.05 were retained resulting in a total of 190 286 and 189 707 SNPs in 2013 and 2014 respectively.

For the selection of candidate genes, an arbitrary threshold of $-\log_{10}(P) \geq 4$ was considered as described in (El-Soda *et al.*, 2015). An association peak corresponds to a group of SNPs that co-occur within a physical distance of ~ 20 kb of the top SNP. The 20 kb window is conservative given that linkage disequilibrium in *A. thaliana* decays within an average of 10 kb (Clark *et al.*, 2007). In order to locate candidate genes, LD was calculated using the R packages LDcorSV and plots drawn with LDheatmap.

GO term enrichment was calculated for SNPs within a 20 kb window of an *a priori* candidate gene as described elsewhere (Atwell *et al.*, 2010; Brachi *et al.*, 2010). For each set of top SNPs, a null

distribution of enrichment was generated to determine a 95% confidence interval. The function we utilized calculates the enrichment as the number of SNPs within candidate genes divided by the average of the null distribution for this number.

Statistical analysis

We estimated the broad-sense heritability of each trait by fitting a model including a single random intercept effect for the identity of the accessions (R package lme4; Eq. 1):

$$Y_{i,j} \sim \mu_i + \epsilon_{i,j}, [1]$$

where μ_i is the random effect of the identity of the accessions, ($\epsilon_{i,j} \sim N(0, \sigma^2_1)$). The heritability was estimated using the formula $h^2 = \sigma^2_g / [\sigma^2_g + (\sigma^2_e/n)]$, where σ^2_g is the genetic variance, σ^2_e is the environmental variance and n is the number of replicates. Phenotypic correlations were estimated by Spearman correlations using the R stats package.

Results

Natural variation in the response to TuMV among *A. thaliana* accessions

In both common garden experiments, TuMV symptoms included plant stunting, developmental arrest, leaf mosaic, sawed and curly leaves and curved stem (Fig. S2). No necrotic lesions were seen on any accession.

We observed extensive variation in viral charge with OD ranging from 0.0 to nearly 0.9. This corresponds to accessions that were both highly susceptible and resistant (Fig. 1). Highly susceptible genotypes accumulate virus to an OD that is more than twice the OD of the non-infected *A. thaliana* control. Resistant genotypes do not accumulate virus and present an OD that is equal or under the mean of the OD of the non-infected *A. thaliana* control. Nineteen common accessions were completely resistant, presenting no viral charge as measured through ELISA in 2013 and 2014 (Table S2). The distribution of OD showed a significant difference between the 2 years experiment, OD's were significantly higher in 2013 (0.249) compared to 2014 (0.148). Climatic data revealed (Fig. S3) no significant difference between the dates of the field experiments in 2013 and 2014 for

temperatures, rainfall and PAR (photosynthetically active radiation); the only exception was a higher greenhouse temperature in 2014 than 2013, before plants were moved to the field. Q-PCR failed to detect any significant difference between the two inoculum concentrations, which were estimated at 0.227 ng/ μ L in 2013 and 0.146 ng/ μ L in 2014. The comparison of viral accumulation (as measured by DAS ELISA, Fig. 2) showed that most accessions retained the same behavior between the two years. Accessions could be divided in two main groups corresponding, first, to those that do not accumulate virus (OD<0.0855, in green Fig. 2) and second, to those that accumulate virus at different levels according to the year (in red Fig. 2). The behavior of some accessions (in black Fig. 2) switched from one group to the other group between the 2 years. In that case, lack of viral accumulation was assumed to be a false negative and the data was removed from further analysis. This was the case for 16 genotypes in 2013 and 68 genotypes in 2014. 266 and 202 were kept for further analysis in 2013 and 2014 respectively.

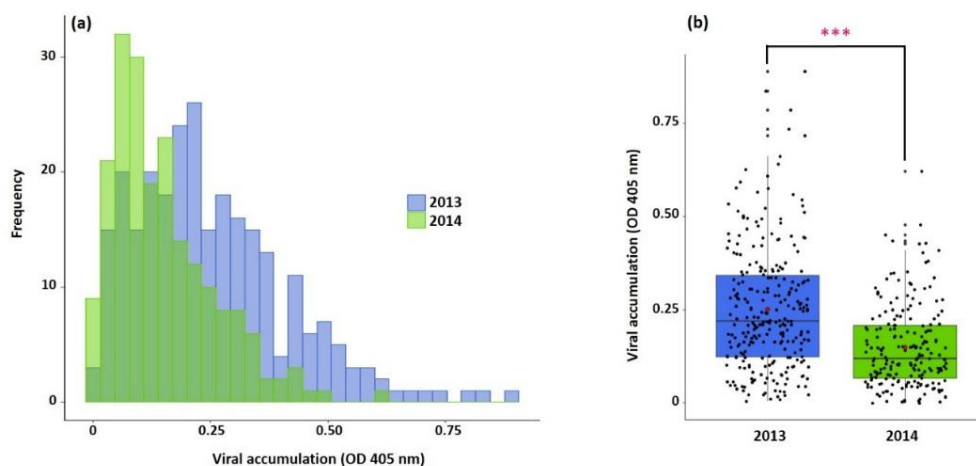


Fig. 1 Distribution of viral accumulations in 2013 and 2014 common garden experiments

(a) Frequency distributions of viral accumulations in 2013 (in blue) and 2014 (in green) in common garden experiments. (b) Boxplot of the distribution of the viral accumulation. Each genotype is represented by a black dot. The mean of the distribution is represented by a red dot. There is a significant difference on viral charge between the two years of experiment (pink asterisk).

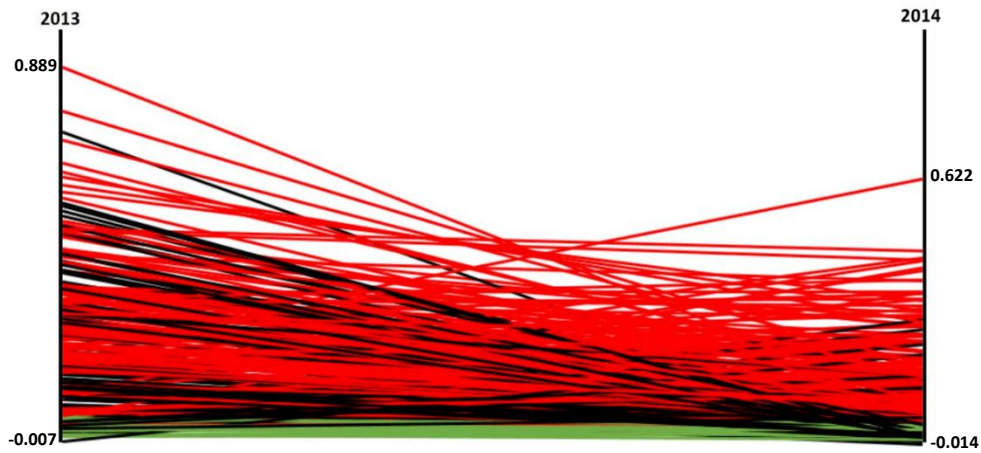


Fig. 2 Slopegraph of the response to virus of *Arabidopsis thaliana* accessions between 2013 and 2014

Each accession is represented by a line. Viral charge measured in 2013 are on the right side and 2014 on the left side. The green lines correspond to the resistant accessions which do not accumulate virus and present an OD that is equal or under the mean of the OD of the non-infected *A. thaliana* control. Red lines correspond to highly susceptible accessions which accumulate virus to an OD that is more than twice the OD of the non-infected *A. thaliana* control. Black lines are accessions that are switching from one group (susceptible versus resistant) to another according to the experiment.

Phenotypic correlations

We considered three different ways to measure the response of *A. thaliana* to the TuMV in 2013: the viral accumulation as measured by the mean OD, the maximum viral OD among replicates and the proportion of inoculated samples that manifest an infection. In 2014, we added two additional phenotypic traits: the normalized stem length, which is the ratio between the stem length of inoculated plants and mean stem length of mock inoculated plants for a given accession, and the presence or absence of symptoms.

All broad sense heritabilities were high with 0.57 and 0.72 for OD traits in 2013 and 2014, respectively, and 0.58 for stem length in 2014.

Spearman correlations were calculated on phenotypic values to explore the relationships among traits (Table 1). In both years, the rate of infection is positively correlated with the viral charge. Surprisingly in 2014, neither stem lengths nor presence/absence of symptoms are correlated with the viral charge: this contrasts to the positive correlations observed between these traits in controlled experiments (Schurdi-Levraud, unpublished data). It seems likely that the lack of correlation between stem length and viral charge under field conditions is the result of relatively

faster development such that accessions bolted before inoculation, which was not the case in controlled conditions. The lack of correlation between symptoms and viral charges may be the consequence of variable morphology under field conditions, making it difficult to assess symptomatology so that symptoms may have been mis- or under-estimated.

Table 1 Phenotypic correlations between the different traits used in the two years of common garden

VIRAL TRAITS	2013 Common Garden Experiment			2014 Common Garden Experiment				
	Viral Charge	Viral Charge (Max)	Rate of infection	Viral Charge	Viral Charge (Max)	Rate of infection	Stem length	Symptoms
Viral Charge	1	0.92 ***	0.75 ***	1	0.89 ***	0.78 ***	-0.11	0.1
Viral Charge (Max)		1	0.56 ***		1	0.52 ***	-0.06	0.07
Rate of infection			1			1	-0.13 -	0.14 *
Stem length							1	0.16 *
Symptoms								1

Spearman correlation coefficient have been used to realize the correlations.

Signif.codes : 0 ‘****’ 0.001 ‘***’ 0.01 ‘**’ 0.05 ‘-’ 0.1 ‘ ’ 1

Genetic architecture of *A. thaliana* response to TuMV infection

We performed a GWA analysis using the mix-model GEMMA (see the Materials and Methods section) employing a threshold of $-\log_{10}(P) \geq 4$ to declare a SNP significantly associated with a trait (Fig. S4). A genome region is considered when a significant set of SNPs occur within a 10kb distance. A comparison of the genome regions detected in 2013 and 2014 is presented as a Venn diagram of each chromosome (Fig. 3). A total of 25 genome regions were found to be consistently detected in 2013 and 2014. In addition, 27 regions unique to 2013 and 26 regions unique to 2014 were identified.

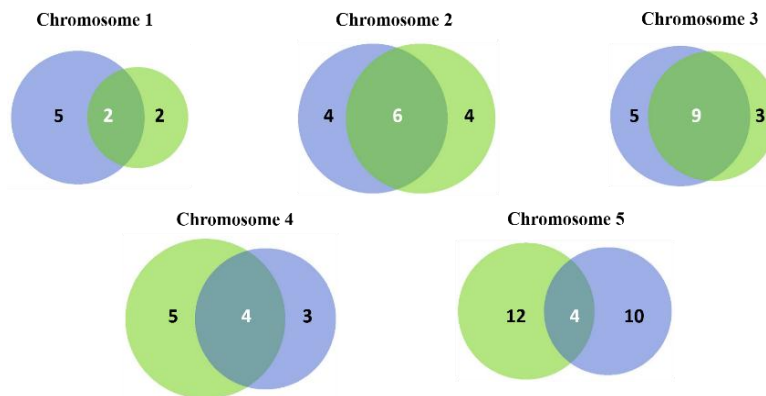


Fig. 3 Venn Diagrams of the comparison of the genome regions detected in 2013 and 2014
Comparisons were realized for each of the five *A. thaliana* chromosomes. The 25 common genome regions are written in white.

GO Term Enrichment of GWA results

To provide global insights into the biological processes that were affected by TuMV infection, we performed an analysis of GO term enrichment for the GWA results (Fig. 4). The algorithm that we used (see the Materials and Methods section) identifies the GO terms that are enriched relatively to a null distribution. Using a 0.95 threshold for the null distribution, we identify 24 significant GO terms in 2013 and 16 significant GO terms in 2014. The two GO terms, “defense response to virus” and “response to virus” appeared particularly important because they were enriched for all the phenotypes tested in both years. Overall, 25% of the GO terms correspond to response to abiotic or biotic stimulus. The other GO terms fall into 3 broad categories: developmental processes (~38 to 56%), cellular processes (~21%) and metabolic processes (~17%). In the metabolic processes

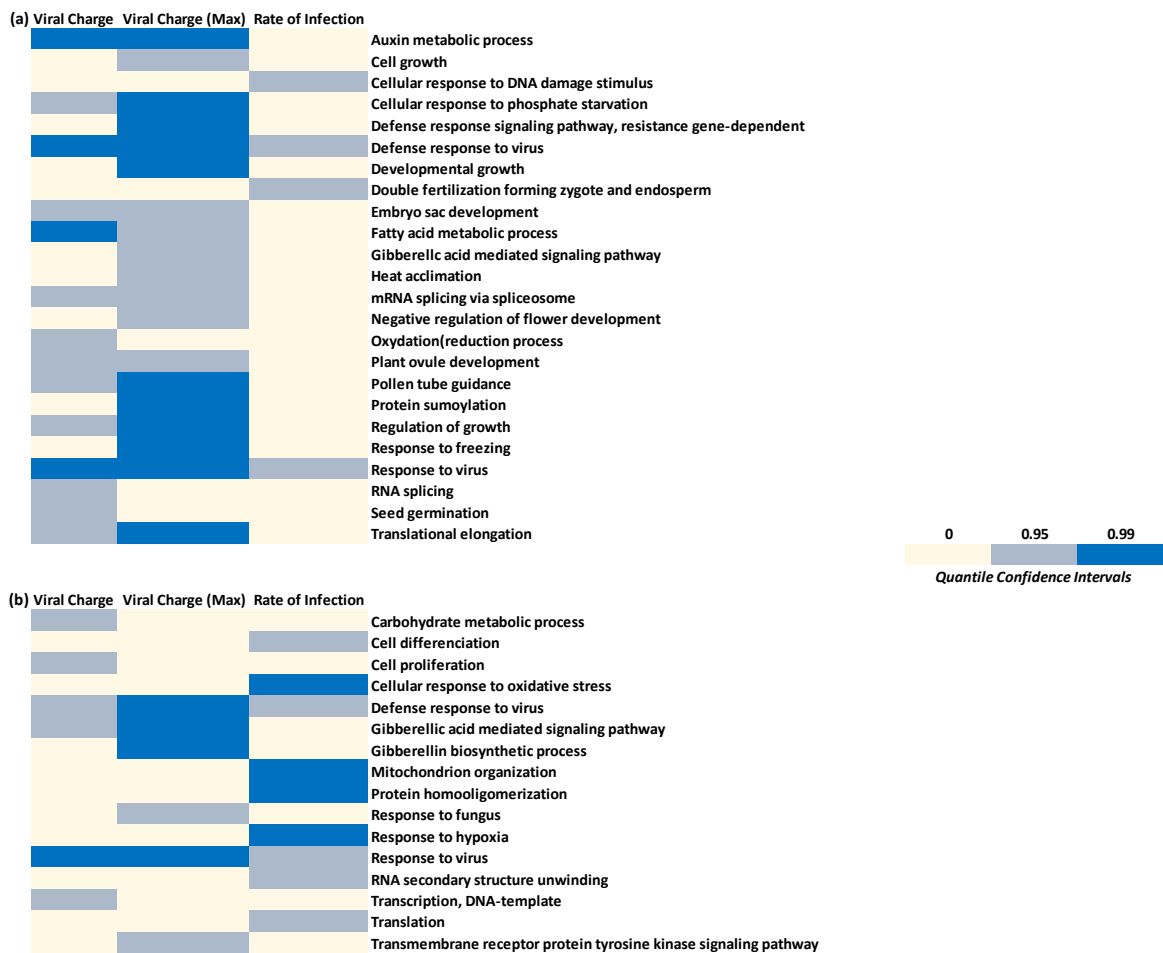


Fig. 4 Heatmap of GO Term enrichment from the GWA results of 2013 and 2014 common garden experiments Heatmaps represent the strengths of the enrichment analysis based on the quantile confidence intervals. The enrichment analysis were realized from GWA results of 2013 (a) and 2014 (b) common garden experiments.

category, enriched GO terms related to hormone signaling through gibberellic acid (2013 & 2014 experiments) and auxin (2013 experiment).

Candidate genes of the response

We sought putative candidate genes in the 25 genomic regions that were mapped in both 2013 and 2014 (Table 2). Year specific regions are listed in Table S3. Among the 25 loci, 3 presented a $\log_{10}p$ -value above 7. The first of these loci maps to a region of chromosome 3 (21,581,213..21,592,778 $-\log_{10}(P)=7.68$) and is significantly associated with the three traits “Viral charge”, “Viral charge (Max)” and “Rate of infection”. This locus co-localize with the RESTRICTED TEV MOVEMENT 3 (*RTM3* - AT3G58350) gene and the locus *sha3* (Cosson *et al.*, 2012; Pagny *et al.*, 2012). *RTM3* encodes a protein with a meprin and TRAF (MATH) homology domain hypothesized to be part of a multiprotein complex that blocks the long-distance movement of several potyviruses (Whitham *et al.*, 1999; Decroocq *et al.*, 2009; Cosson *et al.*, 2010) including *Tobacco etch virus* and *Lettuce mosaic virus*. The LD calculated in this region shows that only the SNPs of *RTM3* are in LD (Fig. S5 a), strongly suggesting the involvement of the product of this gene in TuMV/*Arabidopsis* interaction.

The second and third loci map to chromosome 2 (5,381,663..5,388,087 ; $-\log_{10}(P)= 7.63$) and (5,287,380..5,342,042 ; $-\log_{10}(P)= 7.18$). These two loci are associated with the trait “Rate of infection”. They belong to a gene-poor region. The combined two regions span a 100 kbp which contains only 10 genes. Calculated LD (Fig. S5 b and S5 c) does not reveal blocks of LD but large LD which makes the selection of putative candidate genes difficult.

Six other loci are associated with the three traits “Viral charge”, “Viral charge (Max)” and “Rate of infection”. Among them one belongs to chromosome 3 (205,899..218,658 ; $-\log_{10}(P)= 5.15$). This locus encompasses a DEAD box RNA helicase 1 candidate. The family of DEAD box RNA helicases have already been demonstrated to be involved in plant virus interaction (Kovalev &

Nagy, 2014), especially in plant/TEV (*Tobacco etch virus*) (Huang *et al.*, 2010) and TuMV interactions (Li *et al.*, 2016).

Among the other loci, a region mapped on chromosome 1 (29,894,759..29,896,144 ; $-\log_{10}(P)=4.79$) contains two genes in LD (Fig. S5 d). The gene AT1G79470 encodes a protein with a TIM barrel domain. In *Solanum hirsutum* the gene *Tm-1* encodes a protein with a TIM-barrel domain that binds *Tomato mosaic virus* (ToMV) replication proteins inhibiting its RNA-dependent RNA replication (Ishibashi & Ishikawa, 2013).

Finally, the trait “Rate of infection” mapped to a region of chromosome 4 (8,457,621 ; $-\log_{10}(P)=6.18$) that includes a Bax inhibitor 1 family protein (AT4G14730) (Fig. S5 e). Members of this protein family have been clearly demonstrated to be involved in plant/potyvirus and potexvirus interactions due to their ability to suppress systemic accumulation (Gaguancela *et al.*, 2016). All the other loci revealed candidate genes involved in plant defense responses. It is also interesting to notice that some of them are clearly related to plant metabolism, such as sugar transport, response to phosphate or glutamate pathway.

Table 2 Candidate *A. thaliana* genes resulting from the 25 genomic regions mapped in both 2013 and 2014

Viral Traits	Chr.	Genome Region (start..end)	Best SNP (in bp)	-log10 (P-val)	Genes	Genes description
Rate of Infection	1	8,086,885..8,086,885	8,086,885	5.06	AT1G22860	Vacuolar sorting protein 39
Viral Charge (Max)	1	29,894,759..29,896,144	29,896,144	4.79	AT1G79460 AT1G79470	GA REQUIRING 2 (GA2) - Involved in gibberelin biosynthetic process Aldolase-type TIM barrel family protein - Involved in oxidation reduction
Rate of Infection	2	521,508..538,460	532,612	6.11	AT2G02080 AT2G02090 AT2G02100 AT2G02103 AT2G02120	Indeterminate(ID)-domain 4 - Functions in sequence-specific DNA binding transcription factor activity ETL-1 - Functions in helicase activity, DNA binding, ATP binding, nucleic acid binding Low-molecular-weight cysteine rich 69 - Involved in defense response Unknown protein PDF2.1 - Peptidase inhibitor activity - Involved in defense response
Rate of infection	2	4,071,319..4,071,319	4,071,319	5.01	-	-
Rate of Infection	2	5,287,380..5,342,042	5,289,312	7.18	AT2G12875 AT2G12880 AT2G12900 AT2G12905 AT2G12935 AT2G12940 AT2G12945	Unknown protein Zinc knuckle (CCHC-type) family protein - Functions in zinc ion binding, nucleic acid binding Basic-leucine zipper (bZIP) transcription factor family protein - Involved in regulation of transcription Unknown protein Unknown protein Unfertilized embryo sac 4 (UNE4) - Involved in double fertilization forming a zygote and endosperm Unknown protein
Rate of infection	2	5,381,663..5,388,087	5,387,832	7.63	AT2G13100	Encodes a member of the phosphate starvation-induced glycerol-3-phosphate permease gene family
Rate of infection	2	13,145,936..13,151,930	13,145,936	6.45	AT2G30880 AT2G30890 AT2G30900	Pleckstrin homology domain containing protein - Signal transduction Cytochrome b561/ferric reductase transmembrane protein family Trichome birefringence like 43
Viral Charge (Max)						
Rate of infection	3	205,899..218,658	218,658	5.15	AT3G01516 AT3G01520 AT3G01530 AT3G01540 AT3G01550	Unknown protein Adenine nucleotide alpha hydrolase like superfamily protein - N-terminal protein myristoylation - defense to stress Myb domain protein 57 DEAD box RNA helicase 1 Phosphoenolpyruvate (pep)/phosphate translocator 2 (PPT2) - transport
Viral Charge (Max)						
Rate of infection	3	1,253,054..1,265,320	1,253,054	4.87	AT3G04610 AT3G04620 AT3G04630 AT3G04640 AT3G04650 AT3G04660	Flowering locus KH domain - Positive regulation of flower development Alba DNA/RNA binding protein WVD2 like 1 - Circumnutation, root morphogenesis Glycine rich protein FAD/NAS(P) binding oxidoreductase family protein F-box associated interaction domains containing protein
Viral Charge						
Rate of infection	3	2,974,305..2,978,314	2,977,794	5.12	AT3G09690 AT3G09700 AT3G09710	Alpha/beta hydrolases superfamily protein Chaperone DnaJ-domain superfamily protein - Protein folding - Heat shock protein binding IQ-domain 1 - Calmodulin binding - Glucosinolates biosynthetic process - Defense respons
Rate of infection	3	7,683,665	7,683,665	4.93	AT3G21800	UDP-glucosyl transferase 71B8 - Metabolic process
Rate of infection	3	8,747,302	8,747,302	6.47	AT3G24190	Protein kinase superfamily protein - Protein amino acid phosphorylation
Viral Charge (Max)						
Viral Charge Symptoms	3	9,624,341..9,627,838	3,627,838	4.59	AT3G26270	Cytochrome O450 family 71 subfamily B polypeptide 25 - Oxydation reduction
Rate of infection	3	21,670,298	21,670,298	4.24	AT3G58600	Adpatin ear-binding coat associated protein 1 NECAP-1 - Endocytosis
Viral Charge						
Rate of infection	3	23,291,586	23,291,586	5.11	AT3G63010	GA INTENSITIVE DWARF 1B - Floral organ morphogenesis, raffinose family oligosaccharide
Viral charge	3	21,581,213..21,592,778	21,592,353	7.68	AT3G58280 AT3G58290 AT3G58300 AT3G58310 AT3G58320 AT3G58330 AT3G58340 AT3G58350 AT3G58360	Arabidopsis phospholipase-like protein (PEARU 4) with TRAF-like domain TRAF-like superfamily protein Arabidopsis phospholipase-like protein (PEARU 4) family Unknown protein Ubiquitin-specific protease family C 19-related protein Arabidopsis phospholipase-like protein (PEARU 4) family TRAF-like superfamily protein RESTRICTED TEV MOVEMENT 3 (RTM3) TRAF-like superfamily protein
Viral charge (Max)						
Rate of infection						
Viral charge (Max)						
Rate of infection	4	383,752..383,825	383,752	4.18	AT4G00100	Ribosomal protein S13A - Translation, trichome morphogenesis
Viral Charge (Max)						
Viral Charge Symptoms	4	408,695..426,783	421,592	4.93	AT4G00955 AT4G00960 AT4G00970 AT4G00980	Unknown protein Protein kinase superfamily protein - Protein amino acid phosphorylation Cysteine rich RLK 41 - Protein amino acid phosphorylation Zinc knuckle family protein
Viral Charge (Max)						
Rate of infection	4	900,423..907,590	907,59	4.25	AT1G02050	Sugar transporter protein 7
Viral Charge (Max)						
Rate of infection	4	8,457,621	8,457,621	6.18	AT4G14730 AT4G14740	Bax inhibitor 1 family protein Phosphoinositide binding - Signal transduction
Viral Charge (Max)	5	5,838,190..5,849,031	5,840,307	4.25	AT5G17710	Embryo defective 1241 - Protein folding, embryo development ending in seed dormancy
Symptoms						
Viral Charge Symptoms	5	6,002,985..6,004,834	6,004,834	4.38	AT5G18160 AT5G18170	F-box ans associated interaction domains containing protein Glutamate dehydrogenase 1 (GDH1) - Response to cadmium ion, to salt stress
Rate of Infection	5	16,668,555..16,673,932	16,668,555	5.19	AT5G41670 AT5G41680	6-phosphogluconate dehydrogenase family protein - Involved in response to glucose, fructose sucrose stimulus Protein kinase superfamily protein - Involved in Protein amino acid phosphorylation
Viral charge	5	18,717,743..18,724,002	18,717,743	4.49	AT5G46180 AT5G46190 AT5G46200	Ornithine-delta-aminotransferase (DELTA-OAT) - Involved in ornithine catabolic process RNA-binding KH domain-containing protein Unknown protein
Viral charge (Max)						
Rate of Infection						

Discussion

The genetic architecture of plant virus responses have been mainly considered under controlled conditions, in non-natural and model pathosystems and using bi-parental offspring. This study, aimed to consider plant virus responses in an experimental system that mimics natural field conditions as closely as possible. We chose to decipher the genetic architecture of the *Arabidopsis thaliana*/TuMV interaction. This enable us to make use of *A. thaliana* world-wide genetic diversity and exploit the fact that TuMV is a potyvirus that naturally infects this species (Pagan *et al.*, 2010). We established common garden experiments in each of two years, timing the experiments to match the phenology of the system in nature. To our knowledge this is the first time that GWAS is performed in a field experiment dealing with plant/virus response in *Arabidopsis* (Bartoli & Roux, 2017).

Arabidopsis thaliana infected by TuMV shows a quantitative response from full susceptibility, with high viral charge and symptoms, to complete resistance with no viral charge detected. Part of this variability had been raised by (Martín *et al.*, 1999) who proposed a three class phenotyping for 106 *Arabidopsis* genotypes tested with TuMV UK1 in controlled conditions. The three classes have been defined based on the rate of infection.

We observed that plant phenotypes differ substantially between the field and controlled conditions. Unlike controlled conditions, plants in the field are dark green and red with thick leaves. Stem length was not correlated to viral charge as observed in controlled conditions because they were already developed or at a bolting stage prior to inoculation. This makes symptomatology more challenging to observe and could explain the dearth of loci associated with the presence or absence of symptoms.

Detection of 25 genomic regions mapped in both years reveals the very high reliability of our study. Among the detected loci, some of them are associated with several traits. This is the case of the *RTM3* gene, which belongs to the *sha3* locus previously detected for the PPV/*A. thaliana*

interaction (Cosson *et al.*; Pagny *et al.*, 2012). Although not identified before as a factor in resistance to TuMV, this gene has been detected for response to several potyviruses, in various experimental settings, affirming its major role in potyvirus/*Arabidopsis* interaction. RTM3 has been shown to be involved in long distance virus movement but its precise role is still unknown (Cosson *et al.*, 2012).

We could also map a new candidate gene that was associated with the response to TuMV in both years: the Bax inhibitor 1 family. Bax Inhibitor 1 (BI-1 - AT5G47120) has been demonstrated in *Arabidopsis thaliana*/TuMV interaction to inhibit Potyvirus systemic accumulation (Gaguancela *et al.*, 2016). BI-1 is an endoplasmic reticulum (ER) stress sensor that regulates cell death in response to environmental stresses. Dead box helicase (AT4G00660) has also been shown to be involved in Potyvirus VPg interaction in a yeast two-hybrid screening, and the role of this gene product has been confirmed in KO lines (Huang *et al.*, 2010).

In addition to environmental effects, differences between the two experiments could be attributed to different physiological stages of the plants. In particular, the greenhouse conditions in 2014 were warmer during the germination period before plants moved in to the field.

Surprisingly, eukaryotic translation initiation factors (eIFs), mainly eIF4E, eIF4G and their isoforms (Wang & Krishnaswamy, 2012), were not mapped in our experiment even though they have been identified as key factors driving plant/potyvirus interactions. This could be due to the capability of TuMV to use different forms of these factors as shown in *Brassica rapa* (Jenner *et al.*, 2010), or due to a lack of allelic variability as shown in melon (Nieto *et al.*, 2007).

The influence of abiotic stresses on plant virus interactions was described years ago by (Moury *et al.*, 1998) in their study of hyper-susceptibility to TSWV (*Tomato Spotted Wilt Virus*) in pepper under high temperature. More recently, several studies on Chinese cabbage (Xu *et al.*, 2008; Chung *et al.*, 2015; Hily *et al.*, 2016; van Munster *et al.*, 2017) and *Arabidopsis thaliana* have described the impacts of temperature, water deficit and variable light on responses to viruses. At a more

mechanistic level, (Prasch & Sonnewald, 2013) found that combined heat and drought stress increased the susceptibility of one genotype (Col-0) of the plant *A. thaliana* to TuMV infections through broad transcriptomical and metabolomics changes. These studies, all conducted in controlled conditions, give some clues as to why abiotic and biotic stresses interact. In this regard, agronomically important genes are most likely to be identified when plants are subjected to realistic field conditions.

Acknowledgements

This work was supported in parts by the Research Federation on Integrative Biology and Ecology (FR BIE) and by ANR-13-KBBE-0006 COBRA.

We are grateful to Thierry Mauduit and Pascal Briard for their help in field experiment managing.

Author contributions

VSL conceive the project. VSL and BR planned and designed experiments. BR, PC, MC, FR and VSL phenotyped the plants. BR and VSL made the analyses. BR and VSL wrote the manuscript with input from FR and JB. All the authors proof read the final version of the manuscript.

Supporting Information

The following Supporting Information is available for this article:

Fig. S1 Experiment design of the 2013 and 2014 common garden experiments

Fig. S2 Symptomatology of *A. thaliana* infected by TuMV in common garden experiments

Fig. S3 Climate raw data for 2013 and 2014 common garden experiments and comparison between the 2 years

Fig. S4 Manhattan plots of the GWA analysis with the different phenotypic traits measured in 2013 and 2014 common garden experiments

Fig. S5 Linkage disequilibrium heat maps corresponding to five genome regions mapped in GWA analysis in 2013 and 2014

Table S1 List of the different *A. thaliana* accessions used in the two common garden experiments

Table S2 List of the 19 resistant *A. thaliana* accessions to TuMV and their respective OD values in 2013 and 2014

Table S3 Candidate *A. thaliana* genes resulting from the year specific genome regions detected from both 2013 (a) and 2014 (b) GWA analyses

Fig S1 Experiment design of the 2013 and 2014 common garden experiments. The seedling of all the accessions was realized in an unheated and unlighted greenhouse. After 21 days of growing, all the experiment was transferred under tunnel for acclimation before the transfer to the common garden. The plants were mechanically inoculated just before the transfer. The plants stayed in the common garden device until the first symptoms appear. It corresponds to a period of 13 days post infection.

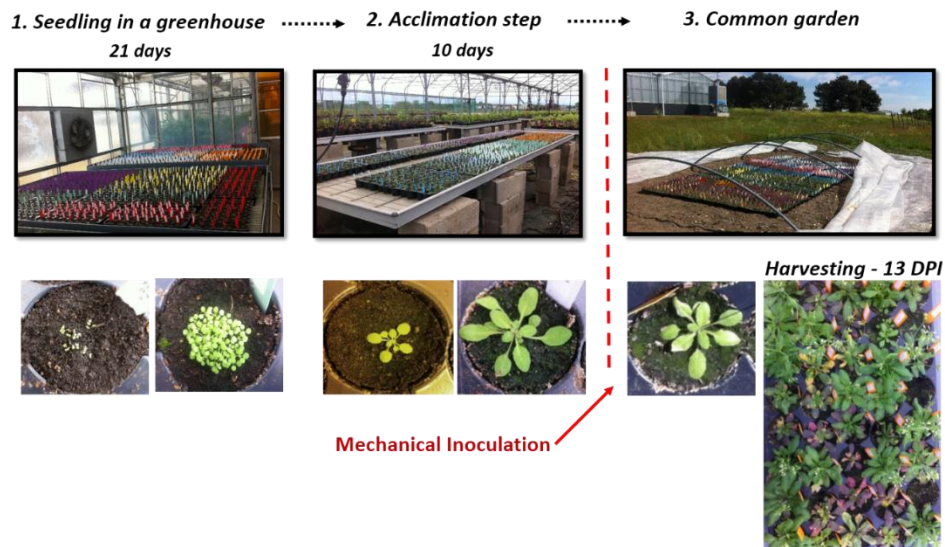


Fig S2 Symptomatology of *A. thaliana* infected by TuMV in common garden experiments

(a) Different accessions of *A. thaliana* mechanically inoculated with TuMV. The red arrows show the plants with typical symptoms: 1. Stunting, 2. Yellowing, 3. Stem curvature. The purplish color of the rosettes is a characteristic phenotype of common garden multistress conditions. (b) Pictures of a characteristic symptom caused by TuMV infection: the stem curvature. All these pictures correspond to plants 13 days after mechanical inoculation by TuMV

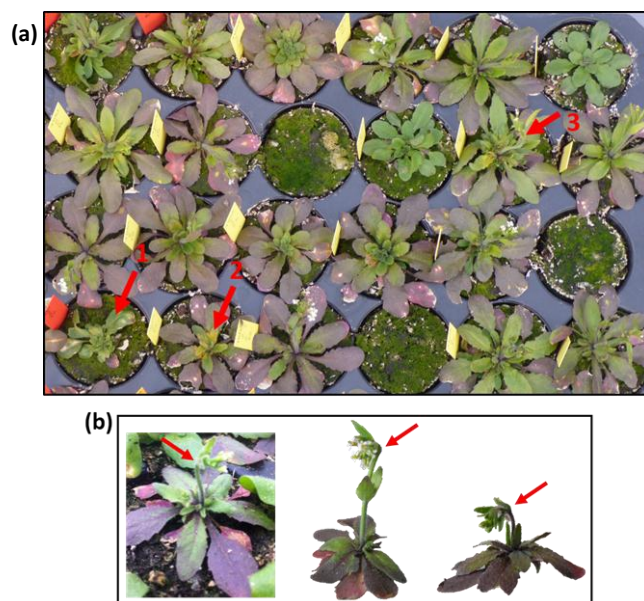


Fig S3 Climate raw data for 2013 and 2014 common garden experiments and comparisons between the two years

Data are sampled from https://intranet.inra.fr/climatik_v2

(a) Rainfall (mm) and average temperatures (°C) in 2013 and 2014 between March and May

(b) PAR photosynthetically active radiation (J/cm²) in 2013 and 2014 between March and May

(c) Boxplot of climatic data of 2013 and 2014 common garden experiments

(d) Average of temperatures (°C) for the three main periods of 2013 (in blue) and 2014 (in green)

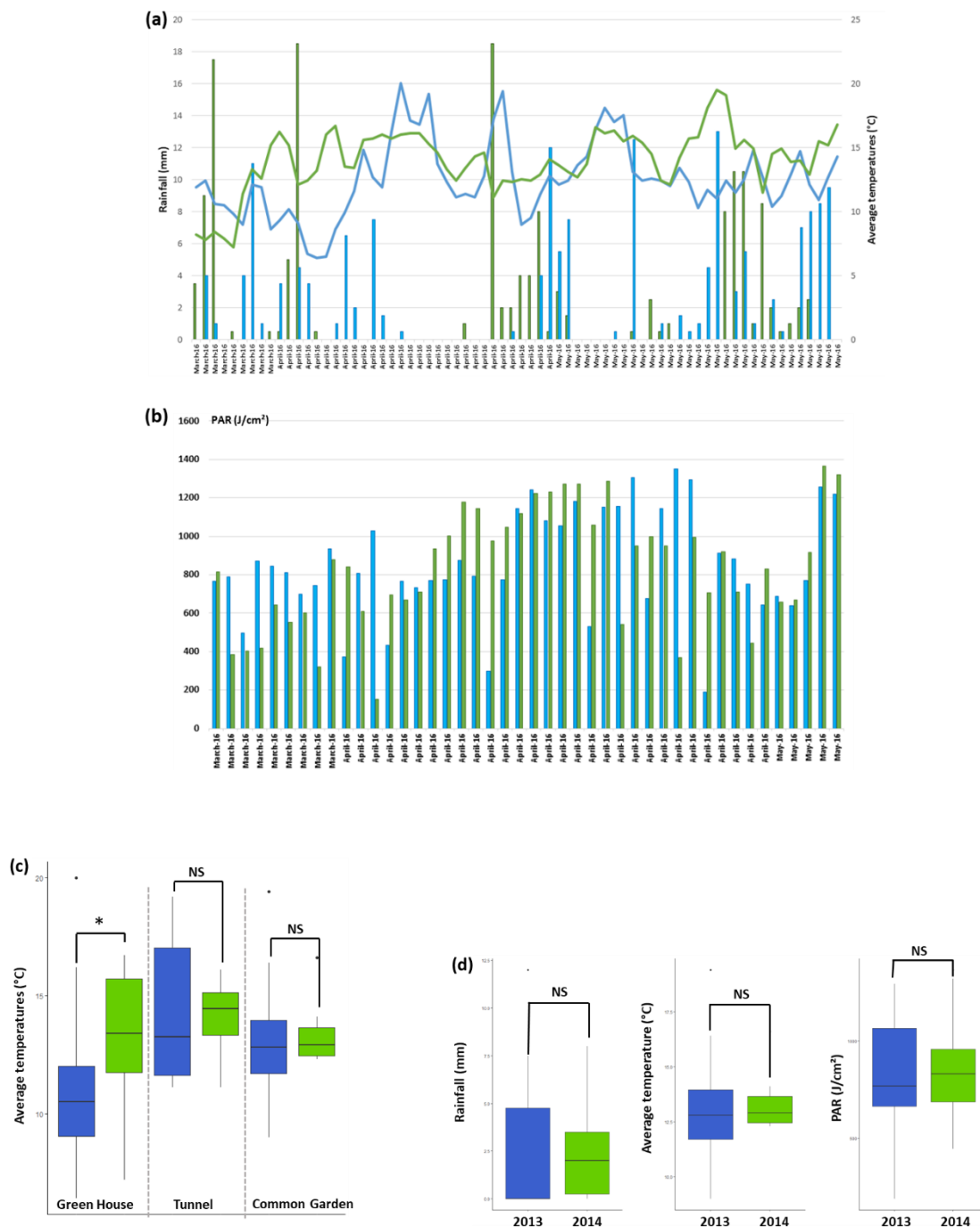
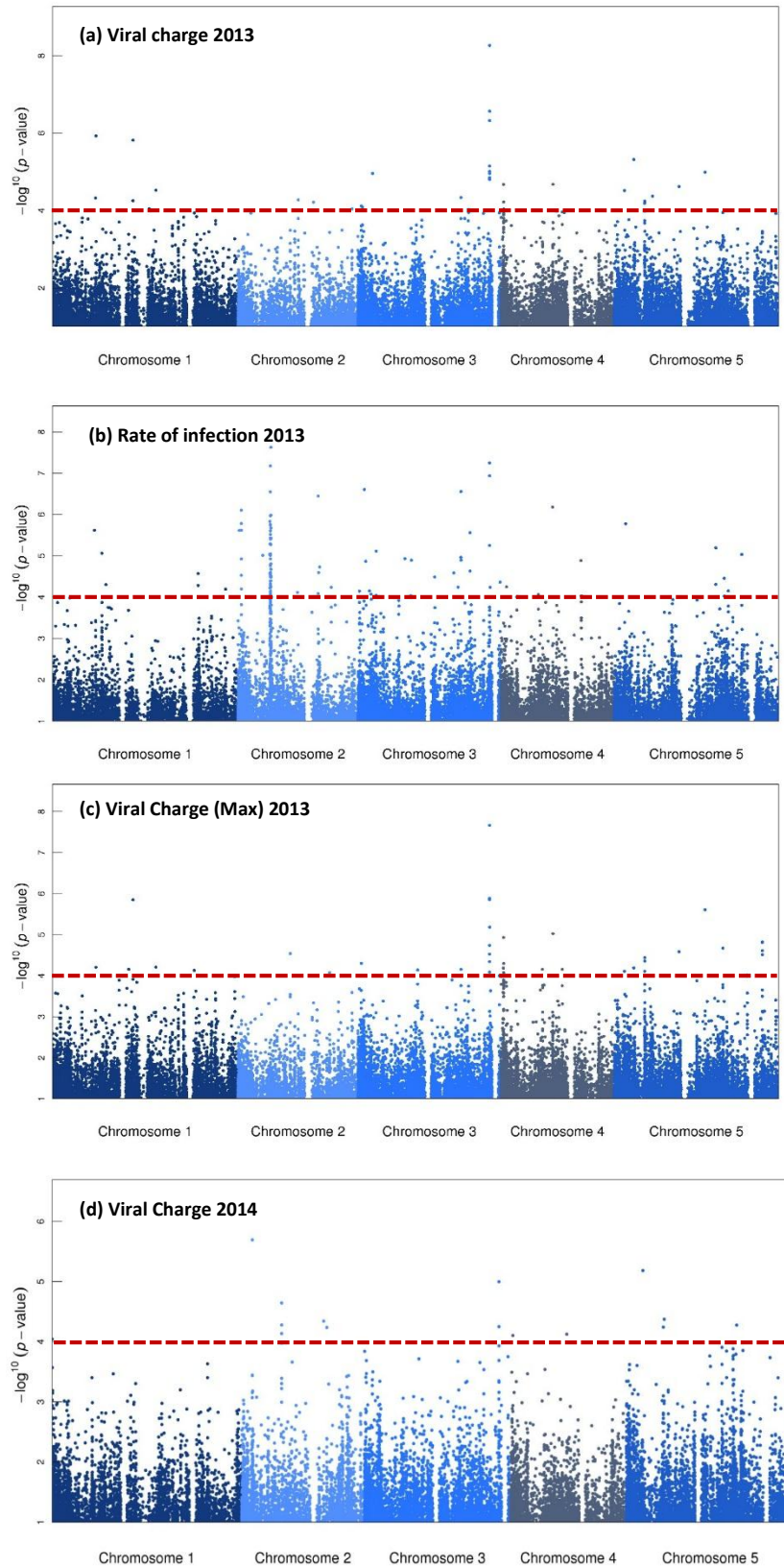


Fig S4 Manhattan plots of the GWA analysis with the different phenotypic traits measured in 2013 and 2014 common garden experiments

Manhattan plots corresponding to GWA analysis of (a) viral charge, (b) rate of infection and (c) viral charge (max) of 2013 common garden experiment and (d) viral charge, (e) rate of infection, (f) viral charge (max) and (g) symptoms of 2014 common garden experiment. The red dotted line corresponds to the chosen significant threshold.



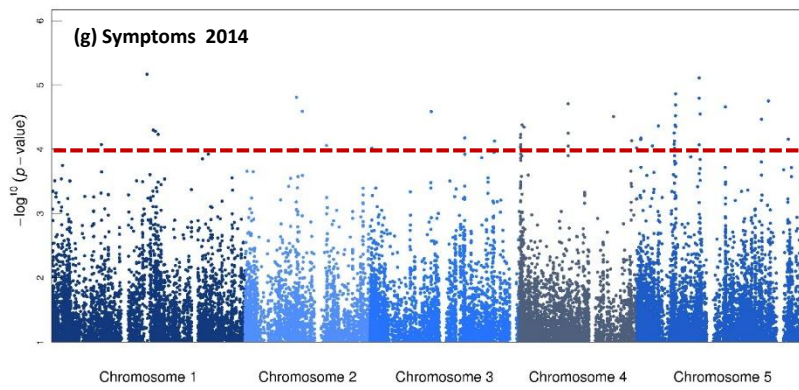
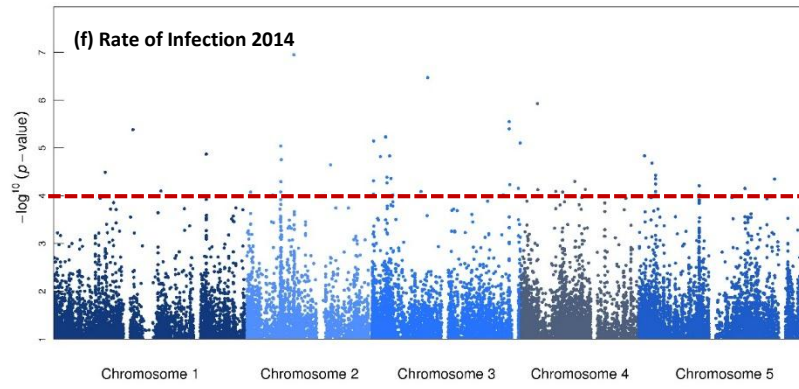
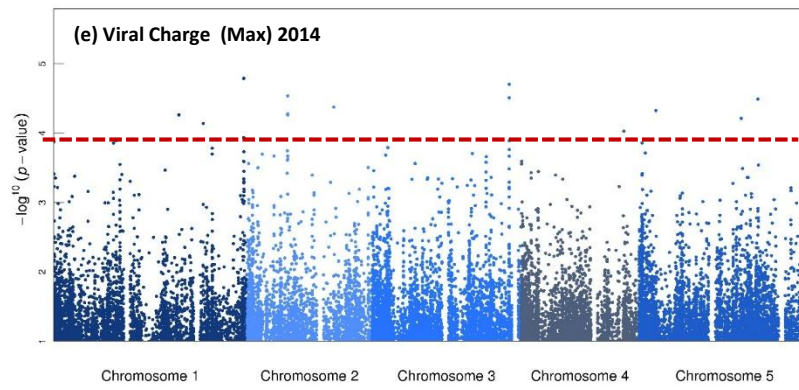


Fig. S5 Linkage disequilibrium (LD) heat maps corresponding to five loci mapped in both 2013 and 2014 GWA analyses

LD heat maps corresponding to loci mapped on (a) chromosome 3, 21.580.246 – 21.603.630 (b) chromosome 2, 5.377.306 – 5.398.494 (c) chromosome 2, 5.279.043 – 5.315.999 (d) chromosome 1, 29.886.883 – 29.906.578, (e) chromosome 4, 8.447.067 – 8.468.170

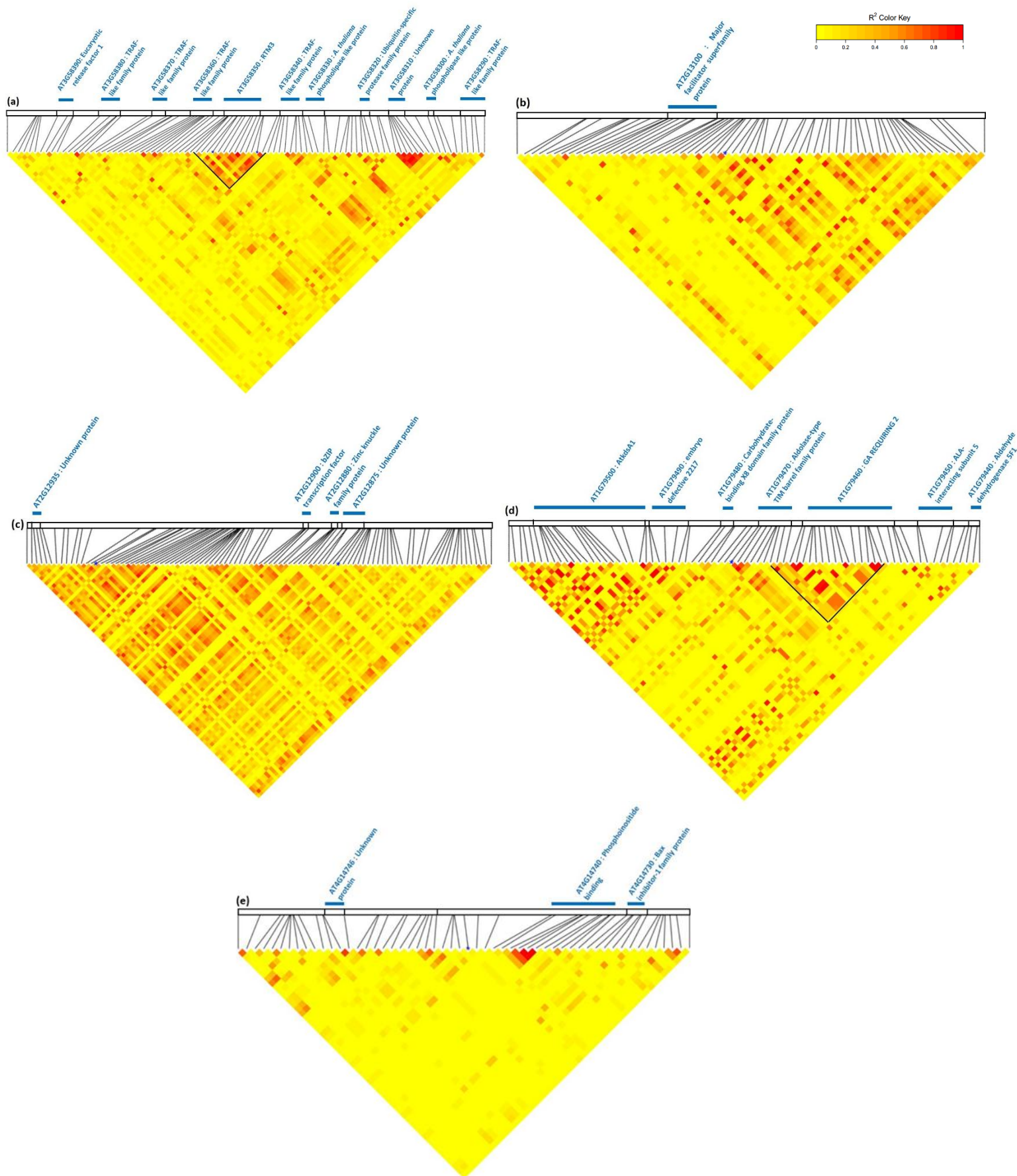


Table S1 List of the different accessions of *A. thaliana* used in the two common garden experiments. Each accession is characterized by a native name, an ecotype ID, the country of origin with geographical coordinates (latitude & longitude). In the last two columns, the accessions used in each experiment are noted by a cross. The hyphens correspond to the missing accessions.

<i>Native Name</i>	<i>Ecotype_ID</i>	<i>Country</i>	<i>Latitude</i>	<i>Longitude</i>	<i>2013 Assay</i>	<i>2014 Assay</i>
Ag-0	6897	FRA	45	1.3	X	X
Alc-0	6988	ESP	40.31	-3.22	X	X
ALL1-2	1	FRA	45.2667	1.48333	-	X
ALL1-5	4	FRA	45.2667	1.48333	X	X
ALL1-6	5	FRA	45.2667	1.48333	X	X
ALL1-8	7	FRA	45.2667	1.48333	X	X
ALL1-9	8	FRA	45.2667	1.48333	X	X
ALL2-1	9	FRA	45.2667	1.48333	X	X
ALL2-4	12	FRA	45.2667	1.48333	X	X
An-1	6898	BEL	51.2167	4.4	X	X
Ang-0	6992	BEL	50.3	5.3	X	X
Ann-1	6994	FRA	45.9	6.13028	-	X
BÅ1-2	8256	SWE	56.4	12.9	X	X
BÅ4-1	8258	SWE	56.4	12.9	X	X
BÅ5-1	8259	SWE	56.4	12.9	X	X
Bay-0	6899	GER	49	11	X	X
Bg-2	6709	USA	47.6479	-122.305	X	X
Bla-1	7015	ESP	41.6833	2.8	X	X
Bor-1	5837	CZE	49.4013	16.2326	X	X
Bor-4	6903	CZE	49.4013	16.2326	X	X
Br-0	6904	CZE	49.2	16.6166	-	X
Bs-1	8270	SUI	47.5	7.5	X	X
Bu-0	8271	GER	50.5	9.5	X	X
Bur-0	5719	IRL	54.1	-6.2	X	X
C24	6906	POR	40.2077	-8.42639	X	X
CAM-16	23	FRA	48.2667	-4.58333	X	X
CAM-25	32	FRA	48.2667	-4.58333	X	X
CAM-42	45	FRA	48.2667	-4.58333	X	X
CAM-45	48	FRA	48.2667	-4.58333	X	X
CAM-48	51	FRA	48.2667	-4.58333	X	X
CAM-56	60	FRA	48.2667	-4.58333	X	X
CAM-58	62	FRA	48.2667	-4.58333	X	X
CAM-61	66	FRA	48.2667	-4.58333	X	X
CAM-64	69	FRA	48.2667	-4.58333	X	X

CAM-7	74	FRA	48.2667	-4.58333	X	X
Can-0	8274	ESP	29.2144	-13.4811	X	X
Cen-0	8275	FRA	49	0.5	X	X
Chat-1	7071	FRA	48.0717	1.33867	X	X
CIBC-17	6907	UK	51.4083	-0.6383	X	X
CIBC-5	6730	UK	51.4083	-0.6383	X	X
Co	7081	POR	40.2077	-8.42639	X	X
Col-0	6909	USA	38.3	-92.3	X	X
Com-1	7092	FRA	49.416	2.823	X	X
Ct-1	6910	ITA	37.3	15	X	X
CUR-10	79	FRA	45	1.75	X	X
CUR-2	80	FRA	45	1.75	X	X
CUR-3	81	FRA	45	1.75	-	X
CUR-4	82	FRA	45	1.75	X	X
CUR-5	83	FRA	45	1.75	X	X
CUR-6	84	FRA	45	1.75	X	X
CUR-7	85	FRA	45	1.75	X	X
CUR-8	86	FRA	45	1.75	X	X
Cvi-0	6911	CPV	15.1111	-23.6167	X	X
CYR	88	FRA	47.4	0.683333	X	X
DAM1	89	FRA	48.7	1.98333	X	X
Di-1	7098	FRA	47	5	X	X
Drall-1	8284	CZE	49.4112	16.2815	-	X
Drall-1	8285	CZE	49.4112	16.2815	X	X
Duk	6008	CZE	49.1	16.2	X	X
Eden-2	6913	SWE	62.877	18.177	-	X
Edi-0	6914	UK	56	-3	X	X
Ei-2	6915	GER	50.3	6.3	X	X
En-1	8290	GER	50	8.5	X	X
Est-1	6916	RUS	58.3	25.3	X	X
Fei-0	8215	POR	40.5	-8.32	X	X
Ga-0	6919	GER	50.3	8	X	X
Gd-1	8296	GER	53.5	10.5	X	X
Ge-0	8297	SUI	46.5	6.08	-	X
Got-7	6921	GER	51.5338	9.9355	-	X
Gr-1	8300	AUT	47	15.5	X	X
Gu-0	6922	GER	50.3	8	X	X
Gy-0	8214	FRA	49	2	X	X
H55	7461	CZE	49	15	X	X
Hi-0	8304	NED	52	5	X	X

Hod	8235	CZE	48.8	17.1	-	X
Hov2-1	8423	SWE	56.1	13.74	X	X
Hov4-1	8306	SWE	56.1	13.74	-	X
Hovdala-2	6039	SWE	56.1	13.74	X	X
HR-10	6923	UK	51.4083	-0.6383	X	X
HR-5	6924	UK	51.4083	-0.6383	X	X
Hs-0	8310	GER	52.24	9.44	X	X
HSm	8236	CZE	49.33	15.76	X	X
In-0	8311	AUT	47.5	11.5	-	X
Is-0	8312	GER	50.5	7.5	X	X
JEA	91	FRA	43.6833	7.33333	X	X
Jm-0	8313	CZE	49	15	X	X
Ka-0	8314	AUT	47	14	X	X
Kelsterbach-4	8420	GER	50.0667	8.5333	-	-
Kent	8238	UK	51.15	0.4	X	X
Kin-0	6926	USA	44.46	-85.37	X	X
Kno-18	6928	USA	41.2816	-86.621	X	X
Koln	8239	GER	51	7	X	X
Kondara	6929	TJK	38.48	68.49	X	X
Kulturen-1	8240	SWE	55.705	13.196	-	X
Kz-1	6930	KAZ	49.5	73.1	X	X
Kz-9	6931	KAZ	49.5	73.1	X	X
LAC-3	94	FRA	47.7	6.81667	X	X
LAC-5	96	FRA	47.7	6.81667	X	X
Lc-0	8323	UK	57	-4	X	X
LDV-14	104	FRA	48.5167	-4.06667	-	X
LDV-16	106	FRA	48.5167	-4.06667	X	X
LDV-25	116	FRA	48.5167	-4.06667	X	X
LDV-3	121	FRA	48.5167	-4.06667	X	X
LDV-30	122	FRA	48.5167	-4.06667	X	X
LDV-31	123	FRA	48.5167	-4.06667	X	X
LDV-34	126	FRA	48.5167	-4.06667	X	X
LDV-44	137	FRA	48.5167	-4.06667	X	X
LDV-53	146	FRA	48.5167	-4.06667	X	X
LDV-57	148	FRA	48.5167	-4.06667	X	X
LDV-58	149	FRA	48.5167	-4.06667	X	X
LDV-6	151	FRA	48.5167	-4.06667	X	X
LDV-68	153	FRA	48.5167	-4.06667	X	X
LDV-70	156	FRA	48.5167	-4.06667	X	X
LDV-8	157	FRA	48.5167	-4.06667	X	X

Ler-1	6932	GER	47.984	10.8719	X	X
Liarum	8241	SWE	55.95	13.85	X	X
Lip-0	8325	POL	50	19.3	X	X
Lis-1	8326	SWE	56	14.7	X	X
Lis-2	8222	SWE	56	14.7	-	X
Lisse	8430	NED	52.25	4.5667	-	X
LL-0	6933	ESP	41.59	2.49	X	X
Lm-2	8329	FRA	48	0.5	X	X
Lom1-1	6042	SWE	56.09	13.9	X	X
Lp2-2	7520	CZE	49.38	16.81	X	X
Lp2-6	7521	CZE	49.38	16.81	X	X
Lu-1	8334	SWE	55.71	13.2	X	X
Lund	8335	SWE	55.71	13.2	-	X
Lz-0	6936	FRA	46	3.3	X	X
MAR2-3	159	FRA	47.35	3.93333	X	X
MIB-1	160	FRA	47.3833	5.31667	X	X
MIB-100	162	FRA	47.3833	5.31667	X	X
MIB-11	163	FRA	47.3833	5.31667	X	X
MIB-14	165	FRA	47.3833	5.31667	X	X
MIB-15	166	FRA	47.3833	5.31667	X	X
MIB-16	167	FRA	47.3833	5.31667	X	X
MIB-17	168	FRA	47.3833	5.31667	X	X
MIB-19	169	FRA	47.3833	5.31667	-	X
MIB-2	170	FRA	47.3833	5.31667	X	X
MIB-20	171	FRA	47.3833	5.31667	X	X
MIB-22	173	FRA	47.3833	5.31667	X	X
MIB-23	174	FRA	47.3833	5.31667	X	X
MIB-24	175	FRA	47.3833	5.31667	X	X
MIB-28	178	FRA	47.3833	5.31667	X	X
MIB-29	179	FRA	47.3833	5.31667	X	X
MIB-32	183	FRA	47.3833	5.31667	X	X
MIB-33	184	FRA	47.3833	5.31667	X	X
MIB-35	186	FRA	47.3833	5.31667	X	X
MIB-36	187	FRA	47.3833	5.31667	X	X
MIB-37	188	FRA	47.3833	5.31667	X	X
MIB-39	190	FRA	47.3833	5.31667	X	X
MIB-40	191	FRA	47.3833	5.31667	X	X
MIB-43	194	FRA	47.3833	5.31667	X	X
MIB-47	196	FRA	47.3833	5.31667	X	X
MIB-50	198	FRA	47.3833	5.31667	X	X

MIB-54	200	FRA	47.3833	5.31667	X	X
MIB-55	201	FRA	47.3833	5.31667	X	X
MIB-57	202	FRA	47.3833	5.31667	X	X
MIB-58	203	FRA	47.3833	5.31667	X	X
MIB-60	204	FRA	47.3833	5.31667	X	X
MIB-61	205	FRA	47.3833	5.31667	X	X
MIB-62	206	FRA	47.3833	5.31667	X	X
MIB-63	207	FRA	47.3833	5.31667	X	X
MIB-64	208	FRA	47.3833	5.31667	X	X
MIB-67	210	FRA	47.3833	5.31667	X	X
MIB-69	212	FRA	47.3833	5.31667	X	X
MIB-70	213	FRA	47.3833	5.31667	X	X
MIB-73	214	FRA	47.3833	5.31667	X	X
MIB-75	215	FRA	47.3833	5.31667	X	X
MIB-76	216	FRA	47.3833	5.31667	X	X
MIB-77	217	FRA	47.3833	5.31667	X	X
MIB-80	219	FRA	47.3833	5.31667	X	X
MIB-83	222	FRA	47.3833	5.31667	X	X
MIB-84	223	FRA	47.3833	5.31667	X	X
MIB-86	224	FRA	47.3833	5.31667	X	X
MIB-87	225	FRA	47.3833	5.31667	X	X
MIB-89	227	FRA	47.3833	5.31667	X	X
MIB-9	228	FRA	47.3833	5.31667	X	X
MIB-90	229	FRA	47.3833	5.31667	X	X
MIB-92	230	FRA	47.3833	5.31667	X	X
MIB-93	231	FRA	47.3833	5.31667	X	X
MOG-11	236	FRA	48.6667	-4.06667	X	X
MOG-12	237	FRA	48.6667	-4.06667	X	X
MOG-37	242	FRA	48.6667	-4.06667	X	X
MOG-40	244	FRA	48.6667	-4.06667	X	X
MOG-55	252	FRA	48.6667	-4.06667	X	X
Mr-0	7522	ITA	44.15	9.65	-	X
Mrk-0	6937	GER	49	9.3	X	X
Mt-0	6939	LIB	32.34	22.46	X	X
Mz-0	6940	GER	50.3	8.3	X	X
N13	7438	RUS	61.36	34.15	-	-
Na-1	8343	FRA	47.5	1.5	X	X
Nc-1	7430	FRA	48.6167	6.25	X	X
Nd-1	6942	SUI	50	10	X	X
NFA-10	6943	UK	51.4083	-0.6383	X	X

NFA-8	6944	UK	51.4083	-0.6383	X	X
Nok-3	6945	NED	52.24	4.45	X	X
Nw-0	8348	GER	50.5	8.5	X	X
Ost-0	8351	SWE	60.25	18.37	-	X
Oy-0	6946	NOR	60.23	6.13	X	X
Pa-1	8353	ITA	38.07	13.22	X	X
PAR-10	257	FRA	46.65	-0.25	X	X
PAR-3	258	FRA	46.65	-0.25	X	X
PAR-4	259	FRA	46.65	-0.25	X	X
PAR-6	261	FRA	46.65	-0.25	X	X
PAR-8	262	FRA	46.65	-0.25	X	X
PAR-9	263	FRA	46.65	-0.25	X	X
Per-1	8354	RUS	58	56.3167	X	X
Petergof	7296	RUS	59	29	X	X
PHW-35	7506	FRA	48.6103	2.3086	-	X
PHW-36	7507	FRA	48.6103	2.3086	X	X
Pla-0	8357	ESP	41.5	2.25	X	X
Pn-0	7307	FRA	48.0653	-2.96591	X	X
Pna-17	7523	USA	42.0945	-86.3253	-	X
Pro-0	8213	ESP	43.25	-6	X	X
Pu2-23	6951	CZE	49.42	16.36	X	X
Pu2-7	6956	CZE	49.42	16.36	X	X
Pu2-8	6957	CZE	49.42	16.36	X	X
Ra-0	6958	FRA	46	3.3	X	X
Rak-2	8365	CZE	49	16	X	X
RAN	266	FRA	48.65	-2	X	X
Rd-0_CS1482	8366	GER	50.5	8.5	X	X
Rd-0_CS1483	8411	GER	50.5	8.5	X	X
Ren-1	6959	FRA	48.5	-1.41	X	X
Ren-11	6960	FRA	48.5	-1.41	X	X
Rmx-A180	7525	USA	42.036	-86.511	X	X
Rou-0	7320	FRA	49.4424	1.09849	X	X
RRS-10	7515	USA	41.5609	-86.4251	-	X
RRS-7	7514	USA	41.5609	-86.4251	X	X
Rsch-4	8374	RUS	56.3	34	X	X
Rubezhnoe-1	7323	UKR	49	38.28	X	X
Sanna-2	8376	SWE	62.69	18	-	X
Sap-0	8378	CZE	49.49	14.24	X	X
Sav-0	8412	CZE	49.1833	15.8833	X	X
Seattle-0	8245	USA	47	-122.2	-	-

Shahdara	6962	TJK	38.35	68.48	X	X
Sorbo	6963	TJK	38.35	68.48	X	X
Sq-1	6966	UK	51.4083	-0.6383	X	X
Sq-8	6967	UK	51.4083	-0.6383	X	X
Sr:5	8386	SWE	58.9	11.2	X	X
St-0	8387	SWE	59	18	X	X
Stw-0	8388	RUS	52	36	X	X
Ta-0	8389	CZE	49.5	14.5	X	X
Tottarp-2	6243	SWE	55.95	13.85	X	X
TOU-A1-105	273	FRA	46.6667	4.11667	X	X
TOU-A1-109	275	FRA	46.6667	4.11667	X	X
TOU-A1-111	277	FRA	46.6667	4.11667	X	X
TOU-A1-112	278	FRA	46.6667	4.11667	X	X
TOU-A1-114	280	FRA	46.6667	4.11667	X	X
TOU-A1-115	281	FRA	46.6667	4.11667	X	X
TOU-A1-116	282	FRA	46.6667	4.11667	X	X
TOU-A1-117	283	FRA	46.6667	4.11667	X	X
TOU-A1-120	287	FRA	46.6667	4.11667	X	X
TOU-A1-124	290	FRA	46.6667	4.11667	-	X
TOU-A1-125	291	FRA	46.6667	4.11667	X	X
TOU-A1-128	292	FRA	46.6667	4.11667	X	X
TOU-A1-129	293	FRA	46.6667	4.11667	X	X
TOU-A1-131	295	FRA	46.6667	4.11667	X	X
TOU-A1-133	296	FRA	46.6667	4.11667	X	X
TOU-A1-134	297	FRA	46.6667	4.11667	X	X
TOU-A1-137	298	FRA	46.6667	4.11667	X	X
TOU-A1-143	302	FRA	46.6667	4.11667	X	X
TOU-A1-18	306	FRA	46.6667	4.11667	X	X
TOU-A1-23	309	FRA	46.6667	4.11667	X	X
TOU-A1-26	310	FRA	46.6667	4.11667	X	X
TOU-A1-27	311	FRA	46.6667	4.11667	X	X
TOU-A1-33	314	FRA	46.6667	4.11667	X	X
TOU-A1-37	316	FRA	46.6667	4.11667	X	X
TOU-A1-39	318	FRA	46.6667	4.11667	X	X
TOU-A1-41	320	FRA	46.6667	4.11667	X	X
TOU-A1-60	326	FRA	46.6667	4.11667	X	X
TOU-A1-61	327	FRA	46.6667	4.11667	X	X
TOU-A1-62	328	FRA	46.6667	4.11667	X	X
TOU-A1-63	329	FRA	46.6667	4.11667	X	X
TOU-A1-65	331	FRA	46.6667	4.11667	X	X

TOU-A1-66	332	FRA	46.6667	4.11667	X	X
TOU-A1-68	334	FRA	46.6667	4.11667	X	X
TOU-A1-69	335	FRA	46.6667	4.11667	X	X
TOU-A1-70	337	FRA	46.6667	4.11667	X	-
TOU-A1-73	338	FRA	46.6667	4.11667	X	X
TOU-A1-74	339	FRA	46.6667	4.11667	X	X
TOU-A1-75	340	FRA	46.6667	4.11667	X	X
TOU-A1-77	341	FRA	46.6667	4.11667	X	X
TOU-A1-79	343	FRA	46.6667	4.11667	X	X
TOU-A1-81	346	FRA	46.6667	4.11667	X	X
TOU-A1-84	348	FRA	46.6667	4.11667	X	X
TOU-A1-93	355	FRA	46.6667	4.11667	X	X
TOU-A1-96	357	FRA	46.6667	4.11667	X	X
TOU-A1-98	359	FRA	46.6667	4.11667	X	X
TOU-C-1	360	FRA	46.6667	4.11667	X	X
TOU-C-2	361	FRA	46.6667	4.11667	X	X
TOU-D-1	363	FRA	46.6667	4.11667	X	X
TOU-D-5	364	FRA	46.6667	4.11667	X	X
TOU-E-7	368	FRA	46.6667	4.11667	X	X
TOU-F-1	369	FRA	46.6667	4.11667	X	X
TOU-H-10	372	FRA	46.6667	4.11667	-	X
TOU-H-13	374	FRA	46.6667	4.11667	X	X
TOU-H-3	375	FRA	46.6667	4.11667	X	X
TOU-I-17	378	FRA	46.6667	4.11667	X	X
TOU-I-6	380	FRA	46.6667	4.11667	X	X
TOU-K-2	385	FRA	46.6667	4.11667	X	X
TOU-L-10	387	FRA	46.6667	4.11667	X	X
TOU-L-17	388	FRA	46.6667	4.11667	X	X
TOU-L-5	389	FRA	46.6667	4.11667	X	X
Ts-1	6970	ESP	41.7194	2.93056	X	X
Ts-5	6971	ESP	41.7194	2.93056	-	X
Tsu-1	6972	JPN	34.43	136.31	X	X
Tu-0	8395	ITA	45	7.5	X	X
UII1-1	8426	SWE	56.06	13.97	X	X
UII-2-3	6973	SWE	56.0648	13.9707	X	X
Uod-1	6975	AUT	48.3	14.45	X	X
Van-0	6977	CAN	49.3	-123	X	X
Vinslov	9057	SWE	56.1	13.9167	X	X
VOU-10	391	FRA	46.65	0.166667	X	X
VOU-2	392	FRA	46.65	0.166667	X	X

VOU-4	393	FRA	46.65	0.166667	X	X
VOU-5	394	FRA	46.65	0.166667	X	X
VOU-6	395	FRA	46.65	0.166667	X	X
VOU-7	396	FRA	46.65	0.166667	X	X
Wa-1	7394	POL	52.3	21	X	X
Wei-0	6979	SUI	47.25	8.26	X	X
Wil-1_Dean	100000	LTU	NULL	NULL	X	X
Ws-0	6980	RUS	52.3	30	X	X
Ws-2	6981	RUS	52.3	30	X	X
Wt-5	6982	GER	52.3	9.3	X	X
Yo-0	6983	USA	37.45	-119.35	X	X
Zdr-1	6984	CZE	49.3853	16.2544	X	X
Zdr-6	6985	CZE	49.3853	16.2544	X	X

Table S2 List of the 19 resistant *A. thaliana* accessions to TuMV and their respective OD values in both years.

Gen	OD_2013	OD_2014
Bay-0	0.0754	0.017
CAM-42	0.0228	0.032
CAM-48	-0.0009	0.044
CAM-61	0.0038	0.021
Is-0	0.0162	0.01
MAR2-3	0.002	0.044
MIB-15	0.0021	0.053
MIB-39	0.0107	0.004
MOG-40	0.0055	0.03
Per-1	0.0372	0.038
Petergof	0.0065	0.044
Ra-0	0.0237	0.02
Rak-2	0.0038	0.025
Stw-0	0.0287	0.021
TOU-A1-18	0.0304	0.033
TOU-A1-65	0.0317	0.047
TOU-A1-69	-0.0016	0.009
TOU-A1-84	0.0276	0.031
TOU-F-1	0.0227	0.022

Table S3 Candidate *A. thaliana* genes resulting from the year specific genome regions detected from both 2013 (a) and 2014 (b) GWA analyses

Viral Traits	Chr	Genome Region (start..end)	Best SNP (in bp)	Log10 (P-val)	Genes	Genes description
(a) 2013						
Viral Charge	1	13.194.587..13.199.493	13.197.526	7.6	AT1G35660	Unknown protein
Viral Charge (Max)						
Viral Charge	2	8.560.604..8.567.534	8.560.604	4.54	AT2G19830	SNF7.2 - Vesicle mediated transport
Viral Charge (Max)						
Viral Charge (Max)	3	16.888.294..16.889.872	16.888.294	6.56	AT3G45940	Glycosyl hydrolases family 31 protein
Rate of infection					AT3G45950	Pre-mRNA splicing Prp18-interacting factor
Rate of infection	3	18.356.535..18.360.215	18.356.535	5.56	AT3G49520	F-box and associated interaction domains-containing protein
Viral Charge	4	8.529.779	8.529.779	5.02	AT4G14910	Histidine biosynthesis 5B
Viral Charge (Max)						
Viral Charge	5	24.292.732..24.295.626	24.294.066	4.82	AT5G60400	Unknown protein
Viral Charge (Max)					AT5G60410	SIZ1 - Plant small ubiquitin like modifier (SUMO) E3 ligase
Viral Charge	5	26.533.930..26.536.996	26.536.571	4.43	AT5G66450	Phosphatidic acid phosphatase (PAP2) family protein
(b) 2014						
Rate of infection	1	23.996.043..23.998.024	23.998.024	4.87	AT1G64600	Methyltransferase copper ion binding
					AT1G64610	Transducin/WD40 repeat-like superfamily protein
Viral Charge	2	1.702.032..1.703.100	1.703.100	5.69	AT2G04845	Acyl-CoA N-acyltransferase (NAT) superfamily protein
Viral Charge (Max)						
Viral Charge	2	6.408.710..6.415.074	6.413.732	4.64	AT2G14900	Gibberellin-regulated family protein
Viral Charge (Max)					AT2G14910	Unknown protein
Symptoms	4	7.731.838..7.733.154	7.733.154	4.71	AT4G13270	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein
					AT4G13280	Terpenoid synthase 12
Symptoms	5	6.053.613..6.055.019	6.054.184	4.86	AT5G18280	Apyrase 2
					AT5G18290	SIP1;2 - Aquaporins family protein
Symptoms	5	9.778.165..9.789.175	9.789.175	5.11	AT5G27630	Acyl-CoA binding protein 5
					AT5G27640	Translation initiation factor 3B1
					AT5G27650	Tudor/PWWP/MBT superfamily protein

References

- Achuo EA, Prinsen E, Hofte M. 2006.** Influence of drought, salt stress and abscisic acid on the resistance of tomato to *Botrytis cinerea* and *Oidium neolyopersici*. *PLANT PATHOLOGY* **55**: 178–186.
- Atwell S, Huang YS, Vilhjalmsdottir BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, et al. 2010.** Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**: 627–631.
- Bartoli C, Roux F. 2017.** Genome-Wide Association Studies In Plant Pathosystems: Toward an Ecological Genomics Approach. *FRONTIERS IN PLANT SCIENCE* **8**.
- Beauchemin C, Laliberte JF. 2007.** The poly(A) binding protein is internalized in virus-induced vesicles or redistributed to the nucleolus during turnip mosaic virus infection. *J Virol* **81**: 10905–10913.
- Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Gortchakov J. 2001.** Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* **13**: 1499–1510.
- Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F. 2010.** Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet* **6**: e1000940.
- Chisholm ST, Mahajan SK, Whitham SA, Yamamoto ML, Carrington JC. 2000.** Cloning of the *Arabidopsis* RTM1 gene, which controls restriction of long-distance movement of tobacco etch virus. *Proc Natl Acad Sci U S A* **97**: 489–494.
- Chung BN, Choi KS, Ahn JJ, Joa JH, DO KS, Park K. 2015.** Effects of Temperature on Systemic Infection and Symptom Expression of Turnip mosaic virus in Chinese cabbage (*Brassica campestris*). *Plant Pathol. J.* **31**: 363–370.
- Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, et al. 2007.** Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**: 338–342.
- Coolen S, Proietti S, Hickman R, Davila Olivas NH, Huang P-P, Van Verk MC, Van Pelt JA, Wittenberg AHJ, De Vos M, Prins M, et al. 2016.** Transcriptome dynamics of *Arabidopsis* during sequential biotic and abiotic stresses. *PLANT JOURNAL* **86**: 249–267.
- Cosson P, Schurdi-Levraud V, Le QH, Sicard O, Caballero M, Roux F, Le Gall O, Candresse T, 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, Le QH, Leger V, Schurdi-Levraud V, Whitham SA, Yamamoto ML, Gopalan S, Le Gall O, Candresse T, et al. 2013.** 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 Physiol* **154**: 222–232.
- Cosson P, Sofer L, Schurdi-Levraud V, Revers F. 2010.** 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**: 1321–1323.
- Decroocq V, Salvador B, Sicard O, Glasa M, Cosson P, Svanella-Dumas L, Revers F, García JA, Candresse T. 2009.** The determinant of potyvirus ability to overcome the RTM resistance of

Arabidopsis thaliana maps to the N-terminal region of the coat protein. *Molecular plant-microbe interactions* **22**: 1302–1311.

Decroocq V, Sicard O, Alamillo JM, Lansac M, Eyquard JP, Garcia JA, Candresse T, Le Gall O, Revers F. 2006. Multiple resistance traits control Plum pox virus infection in Arabidopsis thaliana. *Mol Plant Microbe Interact* **19**: 541–549.

Duprat A, Caranta C, Revers F, Menand B, Browning KS, Robaglia C. 2002. The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant J* **32**: 927–934.

Elena SF, Bedhomme S, Carrasco P, Cuevas JM, de la Iglesia F, Lafforgue G, Lalic J, Prosper A, Tromas N, Zwart MP. 2011. The evolutionary genetics of emerging plant RNA viruses. *Mol Plant Microbe Interact* **24**: 287–293.

El-Soda M, Kruijjer W, Malosetti M, Koornneef M, Aarts MGM. 2015. Quantitative trait loci and candidate genes underlying genotype by environment interaction in the response of Arabidopsis thaliana to drought. *PLANT CELL AND ENVIRONMENT* **38**: 585–599.

Gaguancela OA, Zuniga LP, Arias AV, Halterman D, Flores FJ, Johansen IE, Wang A, Yamaji Y, Verchot J. 2016. The IRE1/bZIP60 Pathway and Bax Inhibitor 1 Suppress Systemic Accumulation of Potyviruses and Potexviruses in Arabidopsis and Nicotiana benthamiana Plants. *MOLECULAR PLANT-MICROBE INTERACTIONS* **29**: 750–766.

Hily J-M, Poulicard N, Mora M-A, Pagan I, Garcia-Arenal F. 2016. Environment and host genotype determine the outcome of a plant-virus interaction: from antagonism to mutualism. *NEW PHYTOLOGIST* **209**: 812–822.

Holmes EC. 2009. The Evolutionary Genetics of Emerging Viruses. *ANNUAL REVIEW OF ECOLOGY EVOLUTION AND SYSTEMATICS* **40**: 353–372.

Huang T-S, Wei T, Laliberte J-F, Wang A. 2010. A Host RNA Helicase-Like Protein, AtRH8, Interacts with the Potyviral Genome-Linked Protein, VPg, Associates with the Virus Accumulation Complex, and Is Essential for Infection. *PLANT PHYSIOLOGY* **152**: 255–266.

Ishibashi K, Ishikawa M. 2013. The resistance protein Tm-1 inhibits formation of a Tomato mosaic virus replication protein-host membrane protein complex. *J Virol* **87**: 7933–7939.

Jenner Walsh, J.A. CE. 1996. Pathotypic variation in turnip mosaic virus with special reference to European isolates. *Plant Pathology* **45**: 848–856.

Jenner CE, Nellist CF, Barker GC, Walsh JA. 2010. Turnip mosaic virus (TuMV) is able to use alleles of both eIF4E and eIF(iso)4E from multiple loci of the diploid Brassica rapa. *Mol Plant Microbe Interact* **23**: 1498–1505.

Jones RA. 2009. Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Res* **141**: 113–130.

Julkowska MM, Testerink C. 2015. Tuning plant signaling and growth to survive salt. *TRENDS IN PLANT SCIENCE* **20**: 586–594.

Kissoudis C, Sunarti S, van de Wiel C, Visser RGF, van der Linden CG, Bai Y. 2016. Responses to combined abiotic and biotic stress in tomato are governed by stress intensity and resistance mechanism. *JOURNAL OF EXPERIMENTAL BOTANY* **67**: 5119–5132.

- Kovalev N, Nagy PD. 2014.** The Expanding Functions of Cellular Helicases: The Tombusvirus RNA Replication Enhancer Co-opts the Plant eIF4AIII-Like AtRH2 and the DDX5-Like AtRH5 DEAD-Box RNA Helicases to Promote Viral Asymmetric RNA Replication. *PLOS PATHOGENS* **10**.
- Lellis AD, Kasschau KD, Whitham SA, Carrington JC. 2002.** Loss-of-susceptibility mutants of *Arabidopsis thaliana* reveal an essential role for eIF(iso)4E during potyvirus infection. *Curr Biol* **12**: 1046–1051.
- Li Y, Xiong R, Bernardis M, Wang A. 2016.** Recruitment of *Arabidopsis* RNA Helicase AtRH9 to the Viral Replication Complex by Viral Replicase to Promote Turnip Mosaic Virus Replication. *SCIENTIFIC REPORTS* **6**.
- Liu J, Feng L, Li J, He Z. 2015.** Genetic and epigenetic control of plant heat responses. *FRONTIERS IN PLANT SCIENCE* **6**.
- Martín ÁM, y Poch HLC, Herrera DM, Ponz F. 1999.** Resistances to turnip mosaic potyvirus in *Arabidopsis thaliana*. *Molecular plant-microbe interactions* **12**: 1016–1021.
- McDonald BA, Stukenbrock EH. 2016.** Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES* **371**.
- Motaung TE, Saitoh H, Tsilo TJ. 2017.** Large-scale molecular genetic analysis in plant-pathogenic fungi: a decade of genome-wide functional analysis. *MOLECULAR PLANT PATHOLOGY* **18**: 754–764.
- Moury B, Selassie KG, Marchoux G, Daubeze AM, Palloix A. 1998.** High temperature effects on hypersensitive resistance to Tomato Spotted wilt Tospovirus (TSWV) in pepper (*Capsicum chinense* Jacq.). *EUROPEAN JOURNAL OF PLANT PATHOLOGY* **104**: 489–498.
- van Munster M, Yvon M, Vile D, Dader B, Fereres A, Blanc S. 2017.** Water deficit enhances the transmission of plant viruses by insect vectors. *PLOS ONE* **12**.
- Nicaise V, Gallois JL, Chafiai F, Allen LM, Schurdi-Levraud V, Browning KS, Candresse T, Caranta C, Le Gall O, German-Retana S. 2007.** Coordinated and selective recruitment of eIF4E and eIF4G factors for potyvirus infection in *Arabidopsis thaliana*. *FEBS Lett* **581**: 1041–1046.
- Nieto C, Piron F, Dalmais M, Marco CF, Moriones E, Gomez-Guillamon ML, Truniger V, Gomez P, Garcia-Mas J, Aranda MA, et al. 2007.** EcoTILLING for the identification of allelic variants of melon eIF4E, a factor that controls virus susceptibility. *BMC Plant Biol* **7**: 34.
- Osakabe Y, Osakabe K, Shinozaki K, Tran L-SP. 2014.** Response of plants to water stress. *FRONTIERS IN PLANT SCIENCE* **5**.
- Ouibrahim L, Caranta C. 2013.** Exploitation of natural genetic diversity to study plant-virus interactions: what can we learn from *Arabidopsis thaliana*? *Mol Plant Pathol* **14**: 844–854.
- Ouibrahim L, Mazier M, Estevan J, Pagny G, Decroocq V, Desbiez C, Moretti A, Gallois J-L, 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. *PLANT JOURNAL* **79**: 705–716.
- Pagan I, Fraile A, Fernandez-Fueyo E, Montes N, Alonso-Blanco C, Garcia-Arenal F. 2010.** *Arabidopsis thaliana* as a model for the study of plant-virus co-evolution. *Philos Trans R Soc Lond B Biol Sci* **365**: 1983–1995.

- Pagny G, Paulstephenraj PS, Poque S, Sicard O, Cosson P, Eyquard JP, Caballero M, Chague A, Gourdon G, Negrel L, et al. 2012.** Family-based linkage and association mapping reveals novel genes affecting Plum pox virus infection in *Arabidopsis thaliana*. *New Phytol* **196**: 873–886.
- Poque S, Pagny G, Ouibrahim L, Chague A, Eyquard J-P, Caballero M, Candresse T, Caranta C, Mariette S, Decroocq V. 2015.** Allelic variation at the *rpv1* locus controls partial resistance to Plum pox virus infection in *Arabidopsis thaliana*. *BMC PLANT BIOLOGY* **15**.
- Prasch CM, Sonnewald U. 2013.** Simultaneous Application of Heat, Drought, and Virus to *Arabidopsis* Plants Reveals Significant Shifts in Signaling Networks. *PLANT PHYSIOLOGY* **162**: 1849–1866.
- Prasch CM, Sonnewald U. 2015.** Signaling events in plants: Stress factors in combination change the picture. *ENVIRONMENTAL AND EXPERIMENTAL BOTANY* **114**: 4–14.
- Revers F, Nicaise V. 2014.** Plant Resistance to Infection by Viruses. *eLS John Wiley & Sons, Ltd: Chichester*.
- Rizhsky L, Liang HJ, Shuman J, Shulaev V, Davletova S, Mittler R. 2004.** When Defense pathways collide. The response of *Arabidopsis* to a combination of drought and heat stress. *PLANT PHYSIOLOGY* **134**: 1683–1696.
- Rouhier N, Jacquot J-P. 2008.** Getting sick may help plants overcome abiotic stress. *NEW PHYTOLOGIST* **180**: 738–741.
- Sanogo S. 2004.** Response of chile pepper to *Phytophthora capsici* in relation to soil salinity. *PLANT DISEASE* **88**: 205–209.
- Sato M, Nakahara K, Yoshii M, Ishikawa M, Uyeda I. 2005.** Selective involvement of members of the eukaryotic initiation factor 4E family in the infection of *Arabidopsis thaliana* by potyviruses. *FEBS Lett* **579**: 1167–1171.
- Suzuki N, Rivero RM, Shulaev V, Blumwald E, Mittler R. 2014.** Abiotic and biotic stress combinations. *NEW PHYTOLOGIST* **203**: 32–43.
- Tomlinson, A J. 1987.** EPIDEMIOLOGY AND CONTROL OF VIRUS DISEASES OF VEGETABLES. *ANNALS OF APPLIED BIOLOGY* **110**: 661–681.
- Triky-Dotan S, Yermiyahu U, Katan J, Gamliel A. 2005.** Development of crown and root rot disease of tomato under irrigation with saline water. *PHYTOPATHOLOGY* **95**: 1438–1444.
- Walsh JA, Jenner CE. 2002.** Turnip mosaic virus and the quest for durable resistance. *Mol Plant Pathol* **3**: 289–300.
- Wang CC, Cai XH, Zheng Z. 2005.** High humidity represses Cf-4/Avr4- and Cf-9/Avr9-dependent hypersensitive cell death and defense gene expression. *PLANTA* **222**: 947–956.
- Wang A, Krishnaswamy S. 2012.** Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. *MOLECULAR PLANT PATHOLOGY* **13**: 795–803.
- Whitham SA, Anderberg RJ, Chisholm ST, Carrington JC. 2000.** *Arabidopsis* RTM2 gene is necessary for specific restriction of tobacco etch virus and encodes an unusual small heat shock-like protein. *Plant Cell* **12**: 569–582.
- Whitham SA, Yamamoto ML, Carrington JC. 1999.** Selectable viruses and altered susceptibility mutants in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **96**: 772–777.

Wiese J, Kranz T, Schubert S. 2004. Induction of pathogen resistance in barley by abiotic stress. *PLANT BIOLOGY* **6**: 529–536.

Xu P, Chen F, Mannas JP, Feldman T, Sumner LW, Roossinck MJ. 2008. Virus infection improves drought tolerance. *New Phytol* **180**: 911–921.

Yang C, Guo R, Jie F, Nettleton D, Peng J, Carr T, Yeakley JM, Fan JB, Whitham SA. 2007. Spatial analysis of *Arabidopsis thaliana* gene expression in response to Turnip mosaic virus infection. *Mol Plant Microbe Interact* **20**: 358–370.

You MP, Colmer TD, Barbetti MJ. 2011. Salinity drives host reaction in *Phaseolus vulgaris* (common bean) to *Macrophomina phaseolina*. *FUNCTIONAL PLANT BIOLOGY* **38**: 984–992.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *NATURE GENETICS* **44**: 821–U136.

A-2. Vers une validation fonctionnelle de gènes candidats par une analyse de mutants

Introduction

Les analyses de GWA effectuées ont permis d'identifier des régions du génome comprenant de potentiels gènes candidats de la réponse d'*A.thaliana* au TuMV. Parmi ces candidats, trois gènes, ont été sélectionnés afin d'essayer de valider leur implication potentielle dans cette interaction sur la base du phénotypage de mutants T-DNA KO (« Knock-Out ») d'*A. thaliana* (fond génétique Col-0). Les candidats choisis sont : le gène **AT2G12900** qui code pour un facteur de transcription basic-leucine zipper (BZIP33), le gène **AT3G58350** qui code pour la protéine RTM3 (Restricted Tobacco etch potyvirus Movement) et le gène **AT5G18170** qui code pour l'enzyme glutamate déhydrogénase 1 (GDH1) (Tab. A2.1). Les deux premiers candidats ont été choisis sur la base de la valeur élevée de la p-value et sur la connaissance préalable de leur implication dans l'interaction *A. thaliana*/potyvirus pour *RTM3*. Enfin, le dernier gène a été choisi car il est en lien avec une réponse métabolique. Le mutant KO correspondant a par ailleurs été testé, par Fernandez-Calvino *et al.*, 2014, pour l'accumulation du *Tobacco rattle virus* (TRV).

Deux paramètres ont été utilisés, pour évaluer l'impact des mutations d'insertion sur les trois gènes sélectionnés, sur la réponse au TuMV : **i)** des notations de symptômes et **ii)** la charge virale quantifiée par Q-PCR. Les notations de symptômes ont été effectuées à différents stades de l'infection. Ces différentes notations ont permis de faire des calculs d'AUDPC (« Area Under Disease Progress Curve »). La charge virale a été quantifiée à 7 dpi, les premiers symptômes apparaissant à cette date.

Les données obtenues chez les mutants T-DNA KO ont été comparées à celles obtenues chez l'accèsion sauvage Col-0, sensible au virus et chez le mutant *elf(iso)4G* (noté E6) qui est résistant au TuMV.

Le design expérimental de cette analyse est décrit en détail dans la partie Matériels et Méthodes de cette thèse (§ B-2).

Tableau A2.1 | Liste des trois gènes choisis et de leurs mutants T-DNA associés pour la validation

Locus	Nom_Locus	Nom_Mutant	Stock code
AT2G12900	Basic leucine zipper 33 (BZIP33)	N682366	SALK_054897C
AT3G58350	RESTRICTED TEV MOVEMENT (RTM3)	N517845	SALK_017845
AT5G18170	Glutamate dehydrogenase 1 (GDH1)	N860072	SALK_042736

1 Analyse des symptômes

Des notations de symptômes ont été effectuées tout au long de l'infection depuis le début l'apparition des premiers symptômes (7 dpi) jusqu'à la fin de l'essai (14 dpi). Des scores d'AUDPC ont été calculés pour chaque mutant et pour les deux témoins d'infection (Col-0 et E6). Les deux témoins présentent des résultats très contrastés entre eux. Aucun symptôme n'a été observé chez le témoin négatif E6. Chez Col-0, les premiers symptômes étaient clairement visibles à partir de 7 dpi. Aucune différence significative sur le score d'AUDPC n'a été mise en évidence entre les mutants entre eux et entre les mutants et Col-0 (Fig. A2.1)

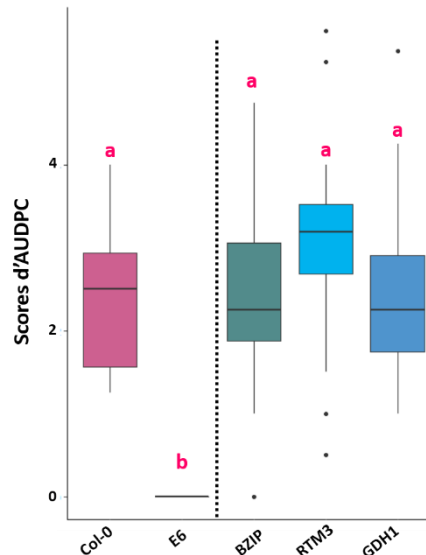


Figure A2.1 | Comparaisons des scores d'AUDPC entre les mutants T-DNA testés et les témoins d'infection
L'ensemble des mutants et les deux témoins ont été répétés 36 fois. Des calculs d'AUDPC ont été effectués pour chacun des réplicats. Des analyses statistiques ont été menées (kruskal-wallis) sur l'ensemble de ces échantillons (Sign.codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1)

2 Analyse de la charge virale à 7 dpi

Des quantifications de la charge virale ont été effectuées par Q-PCR sur les échantillons prélevés à 7 dpi. Cette analyse de la charge virale a été effectuée sur des feuilles inoculées et sur des feuilles en systémie. Les échantillons correspondant au témoin négatif de l'infection (E6), ne présentent pas d'accumulation virale que ce soit en feuille inoculée ou en systémie. En revanche, une accumulation virale est quantifiée en feuilles inoculées mais également en feuilles en systémie chez Col-0. La même observation est réalisée pour l'ensemble des trois mutants que ce soit sur feuilles inoculées que sur les feuilles en systémie (Fig. A2.2). Comme pour les résultats de score d'AUDPC, les mutants T-DNA testés présentent le même phénotype que l'accession Col-0. Il est intéressant aussi de noter que la quantification du virus à 7 dpi est déjà possible en systémie avec des concentrations virales semblables entre feuilles inoculées et feuilles en systémie.

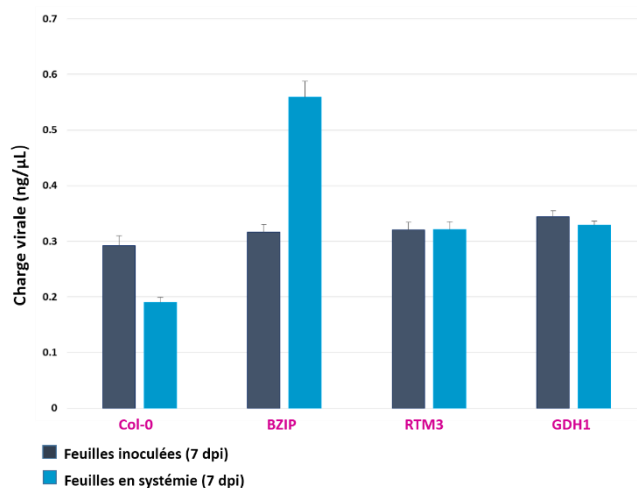


Figure A2.2 | Quantification de la charge virale par Q-PCR à 7 dpi sur des feuilles inoculées et en systémie des mutants T-DNA et du contrôle d'infection positif Col-0

La charge virale à 7 dpi a été quantifiée chez les trois mutants T-DNA et chez le témoin Col-0. La charge virale au niveau des feuilles inoculées (bleu foncé) et en systémie (bleu clair) a été mesurée.

Bilan

Les analyses de génétique d'association sur la population mondiale ont permis de détecter 25 régions du génome retrouvées de façon commune sur les deux années d'étude. L'ensemble de ces loci contiennent des gènes candidats potentiels de la réponse d'*A. thaliana* au TuMV en conditions extérieures. Dans le but de valider leur implication dans cette interaction, des mutants T-DNA de trois gènes candidats ont été testés. Des scores d'AUDPC ont été calculés suite à des notations de symptômes tout au long de l'infection. Ils révèlent qu'il n'existe pas de différences significatives entre les mutants T-DNA et le témoin positif à l'infection, Col-0. De même, la quantification de la charge virale à 7 dpi sur feuilles inoculées et sur les feuilles en systémie montre que le virus s'accumule à des concentrations similaires entre mutants et similaires à celles quantifiées chez Col-0. D'ailleurs, les concentrations virales des feuilles en systémie sont aussi importantes que celles des feuilles inoculées. Sur la base de ces résultats, la conclusion serait de dire que ces gènes ne seraient pas impliqués dans l'interaction *A. thaliana*/TuMV.

La limite majeure de l'analyse des mutants T-DNA KO dont le fond génétique est Col-0 réside dans le fait que Col-0 est sensible au virus. De ce fait, l'effet du KO peut être nul. Ainsi, **i)** si le gène est un gène de résistance, aucun effet n'est attendu, le virus utilisant les facteurs de sensibilité de Col-0, **ii)** si le gène est un gène de sensibilité, dans ce cas une perte de fonction chez Col-0 pourrait éventuellement entraîner une diminution et/ou un retard de l'accumulation virale et des symptômes chez les mutants par rapport à l'accession sauvage Col-0. Au regard des résultats obtenus sur les trois mutants testés, cette stratégie de validation semble complexe à utiliser. La seule stratégie de validation fonctionnelle qui pourrait être mise en œuvre serait une stratégie de complémentation du fond génétique Bay-0 par les allèles de Shahdara et/ou Col-0 (des accessions sensibles au TuMV).

Dans leur revue, Bartoli et Roux (2017) font un recensement des différentes études s'intéressant aux bases génétiques de la réponse des plantes à des agents pathogènes via des approches de génétique d'association. Il en ressort, parmi les 35 études listées, que plus de la moitié (~62.8%) ont été directement menées sur des plantes à intérêt agronomique. L'objectif

majeur de ces études étant de pouvoir accélérer l'identification de marqueurs génétiques pouvant être utilisés comme des marqueurs de sélection. D'ailleurs, ces analyses de GWA sont rarement suivies par des études permettant la validation fonctionnelle de gènes candidats potentiels (Bartoli & Roux, 2017).

Toutes ces données et résultats montrent qu'il n'est effectivement pas toujours facile de valider des candidats mis en évidence sur des loci détectés par GWA. Cette difficulté à sélectionner et valider des gènes candidats suite à des analyses GWA pourrait s'expliquer par la forte densité de gènes dans le génome d'*A. thaliana*. De plus, malgré un LD moyen estimé à ~10kb, il existe des régions où le LD est très étendu ce qui ne facilite pas la sélection des candidats. Enfin, un phénotype peut être obtenu par la combinaison de différents gènes, la diversité génétique explorée en GWA peut donc limiter la détection des gènes impliqués dans la variation du phénotype de part cette hétérogénéité génétique et/ou allélique (Bartoli & Roux, 2017; Fusari *et al.*, 2017). De plus, les loci détectés par ces approches de GWA sont en relation avec des variations alléliques qui ne peuvent être validées par l'utilisation de mutant KO (Bergelson & Roux, 2010).

Dans un certain nombre d'études, les analyses GWA sont complémentaires à d'autres analyses ayant déjà permis de caractériser les bases génétiques associées à un trait complexe. Les analyses de génétique d'association sont donc souvent utilisées dans le but de re-définir avec plus de précision un QTL large déjà identifié. Par exemple, dans leur étude, Pagny et ses collaborateurs (2012), ont mis en place une analyse de génétique d'association en utilisant 147 accessions issues de la population mondiale d'*A. thaliana* dans le but de compléter leur étude génétique sur la résistance d'*A. thaliana* au mouvement en systémie du PPV. Les analyses GWA leur ont ainsi permis de confirmer les loci identifiés par des approches plus classiques de QTL mapping sur des populations bi-parentales tel que le locus *sha3* sur le chromosome 3. La position de ce locus a d'ailleurs été redéfinie et réduite à un intervalle de 13 gènes (Pagny *et al.*, 2012).

Dans notre situation, nous disposons « uniquement » des résultats d'analyses GWA menées en extérieur sur deux années. L'intersection des candidats trouvés pour les deux années d'expérimentation, détectés avec plusieurs variables et en relation avec des données bibliographiques nous permet de dresser une première liste de « top » candidats. Il s'agit à présent d'affiner l'analyse de ces régions afin d'essayer de confirmer l'implication de certains des gènes candidats dans la réponse d'*A. thaliana* au TuMV. A ce stade, un travail bioinformatique semble indispensable afin de pouvoir comparer les haplotypes de génotypes sensibles – telle que l'accession Col-0 – avec les haplotypes des accessions résistantes. Ces analyses comparatives pourraient être réalisées **i)** sur une région génomique tout entière ou **ii)** sur des gènes candidats spécifiques. Cela permettrait de voir s'il est possible de déterminer des groupes sur les bases des comparaisons de séquences qui correspondraient aux groupes de sensibilité permettant de valider un ou plusieurs gènes candidats. Des analyses de réseaux de gènes pourraient également être envisagées afin d'apporter de l'information sur les schémas de réponse (Martínez *et al.*, 2016).

A-3 Réponse d'une population d'*A. thaliana* locale française vis-à-vis du TuMV

Introduction

La caractérisation des bases génétiques de la réponse d'*A. thaliana* au TuMV en conditions de multistress a été réalisée en se basant sur la variabilité génétique d'une population mondiale d'*A. thaliana*. Travailler sur un grand nombre d'accessions avec des origines mondiales diverses a permis d'exploiter cette diversité naturelle afin d'identifier et caractériser des facteurs impliqués dans l'interaction *A. thaliana*/TuMV. Parmi ces génotypes, plusieurs étaient issus de populations françaises dont nous avons pu observer la variabilité de réponse au cours des deux années d'expérimentations à l'extérieur. Ces observations nous ont conduit à explorer la diversité naturelle de populations locales. Les questions qui se posent sont **i) quelle est la réponse de génotypes locaux vis-à-vis d'un virus commun le TuMV ? ii) l'architecture génétique de ces réponses présente-t-elle des loci communs avec ceux observés pour les génotypes mondiaux ?**

Une expérimentation a de ce fait été menée sur une **population d'*A. thaliana* française, la population TOU-A** mise à disposition par F. Roux (LIPM Toulouse, France). Il s'agit d'une population de l'Est de la France (village de Toulon-sur-Arroux, France). Cette population apparaît fortement isolée génétiquement des autres populations locales échantillonnées autour d'un rayon de 1km ce qui rend donc négligeable les effets de migration sur la population TOU-A. Cette population est composée de 195 accessions dont 80 ont été collectées en 2002 (TOU-A1) et 115 en 2010 (TOU-A6). Cette population a fait l'objet d'un séquençage dense de 1.902.203 SNPs. Ce séquençage ainsi que la variabilité phénotypique rendent possible des analyses de génétique d'association sur cette population locale française (Frachon *et al.*, 2017).

Une expérimentation a été conduite à partir de cette population TOU en conditions de multistress sur la base des autres expériences déjà conduites en conditions extérieures (Rubio *et al.*, 2017 – *submitted*). Au-delà des deux questions posées, l'utilisation de cette population échantillonnée à deux périodes différentes (2002 et 2010) permettra de regarder d'éventuelles variations dans la réponse au TuMV sur un pas de temps de 8 générations.

Le phénotypage de la réponse au TuMV a été réalisé par la mesure de six variables phénotypiques mesurées/quantifiées sur des échantillons prélevés 13 jours après l'infection par le TuMV. Ces variables ont été utilisées par la suite dans le cadre d'une analyse de génétique d'association. Les résultats obtenus ont été comparés à ceux déjà décrits pour la population mondiale d'*A. thaliana*.

1. Réponse phénotypique de la population française TOU-A au TuMV

La réponse à l'infection par le TuMV des différentes accessions d'*A. thaliana* issues de la population française TOU a été caractérisée à partir de six traits phénotypiques liés à la maladie et à la production de biomasse. Ainsi, les traits phénotypiques analysés lors de cet essai concernent :

- la **charge virale** mesurée par ELISA semi-quantitatif. Elle correspond de ce fait à une valeur moyenne de la densité optique mesurée à 405 nm pour chacun des échantillons de chaque accession
- la **fréquence d'infection en lien avec la charge virale** (notée Fc_charge virale). Elle correspond au nombre d'échantillons/accession présentant une DO positive en ELISA sur le nombre total d'échantillons/accession
- la **symptomatologie** en lien avec la notation de symptômes sur la base d'une échelle de notation allant de 0, absence de symptômes à 4, symptômes sévères caractérisés par de fortes jaunisses, dessèchement et une forte réduction de la croissance
- la **fréquence d'infection en lien avec les symptômes** (notée Fc_Symptômes). Elle correspond au nombre d'échantillons/accession présentant des symptômes sur le nombre total d'échantillons/accession.
- la **hauteur des hampes** qui a été normalisée par des échantillons inoculés à blanc (mock-inoculés) pour chacune des accessions analysées. Cette normalisation permet de prendre en compte la variabilité liée au génotype en ne considérant que la variation phénotypique en lien avec l'infection virale.
- la **biomasse sèche** également normalisée par des échantillons inoculés à blanc (mock-inoculés) pour chacune des accessions étudiés.

L'ensemble de ces traits phénotypiques ont été transformés en utilisant la fonction log.

L'ensemble de ces variables ont été choisies car elles ont été montrées comme associées à la réponse à la maladie notamment en conditions contrôlées (données chambre de culture non présentées).

La charge virale moyenne sur l'ensemble de la population TOU-A a été estimée par DAS ELISA à une valeur de 0.445 (DO 405 nm). Cette valeur moyenne est supérieure aux DO moyennes obtenues sur les accessions de la population mondiale testées en 2013 (0.249) et en 2014 (0.148) (Rubio et al., 2017 – *submitted*). La quantification de la charge virale de l'inoculum mesurée par Q-PCR est 1.2 fois et 1.4 fois supérieure à celle estimée sur les inocula des essais extérieurs de 2013 et 2014 respectivement (Rubio et al., 2017 – *submitted*). Par rapport aux accessions de la population mondiale testées en 2013 et 2014, 27 accessions font partie de la population TOU-A1. Deux types de comportements sont observés en fonction des trois années d'étude : **i**) 23 accessions qui sont sensibles au TuMV sur les trois essais (en bleu Tab. A3.A1) et **ii**) 4 accessions dont la réponse au TuMV varie en fonction des essais. Il s'agit d'accessions qui en 2013 et 2014 semblaient être résistantes mais qui en 2016 présentent une fréquence d'infection en lien avec la charge virale allant de 44% à 78% (en violet Tab. A3.A1).

La distribution de ces différents traits phénotypiques montre qu'il existe une variabilité de réponse notamment en ce qui concerne la charge virale (Fig. A3.1a), les symptômes (Fig.

A3.1c) et la biomasse sèche (Fig. A3.1f). La distribution de la longueur des hampes est marquée par un nombre important d'échantillons ne présentant pas de hampes. En effet, près de la moitié des génotypes (93/191 soit 48.7 % de génotypes de la population) ne présentait pas de hampes au moment du prélèvement (flèche orange Fig. A3.1e). De même, la fréquence d'infection basée sur la charge virale montre qu'un très faible nombre d'échantillons (< 10%) n'accumule pas de virus au sein de cette population (flèche rose Fig. A3.1b.).

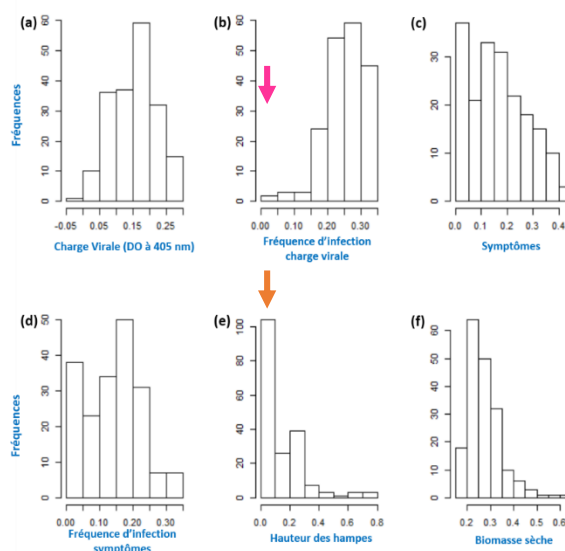


Figure A3.1 | Distribution des différents traits phénotypiques mesurés en réponse au TuMV pour la population française TOU-A d'*A.thaliana* en conditions extérieures

Fréquence des distributions représentée en ordonnée et les différents traits phénotypiques (transformés en utilisant la fonction log) en abscisse : **(a)** charge virale quantifiée par DAS-ELISA (DO mesurée à 405 nm) **(b)** fréquence d'infection en lien avec la charge virale = nombre d'échantillons/accession avec une DO en ELISA positive/nombre total d'échantillons/accession **(c)** symptômes notés à partir d'une grille de notation allant de 0, aucun symptôme à 4, symptômes très sévères **(d)** fréquence d'infection en lien avec les symptômes = nombre d'échantillons/accession avec des symptômes/nombre total d'échantillons/accession **(e)** hauteur des hampes normalisée par les échantillons de chaque génotype non infectés **(f)** biomasse sèche normalisée par les échantillons de chaque génotype non infectés

Les données relatives à la hauteur des hampes et à la biomasse sèche ont été comparées entre échantillons inoculés par le virus et ceux mock-inoculés. Ainsi, il n'existe pas de différences significatives dans la hauteur des hampes entre échantillons inoculés et mock-inoculés (Fig. A3.A1). En revanche, une différence significative a été mise en évidence sur la biomasse sèche (Fig. A3.A1). Cela montre l'effet de l'infection sur la biomasse des plantes. En ce qui concerne la hauteur des hampes, ce trait phénotypique semble être moins clairement associé à l'infection. Cela pourrait s'expliquer par les particularités de réponse déjà observées lors des essais extérieurs réalisés sur la population mondiale (Rubio et *al.*, 2017 – *submitted*). En effet, le phénotype est très différent par rapport aux conditions contrôlées. Par exemple, les notations de symptômes sont beaucoup plus complexes à réaliser ce qui amène le plus souvent à sous-estimer la maladie. La hauteur des hampes serait donc un autre trait phénotypique sur lequel les conditions environnementales semblent avoir un effet important.

Le lien entre les différentes variables caractérisées a été testé dans une analyse des corrélations (Fig. A3.2a) et est représenté sur le graphique des variables (Fig. A3.2b) où trois

groupes de variables peuvent être décrits **i)** le groupe comprenant les données en relation avec la charge virale (en bleu), **ii)** le groupe comprenant les données de symptômes (en vert) et **iii)** le groupe comprenant les données de hauteur de hampe et de biomasse sèche. L'indépendance entre les données de charge virale et la symptomatologie est très surprenante (Fig. A3.2a). La population française TOU-A étant constituée d'accessions échantillonnées à deux périodes différentes (2002 et 2010), des analyses comparatives entre ces deux groupes d'accessions (TOU-A1 et TOU-A6) permettent d'apporter des éléments de réponse à ces observations.

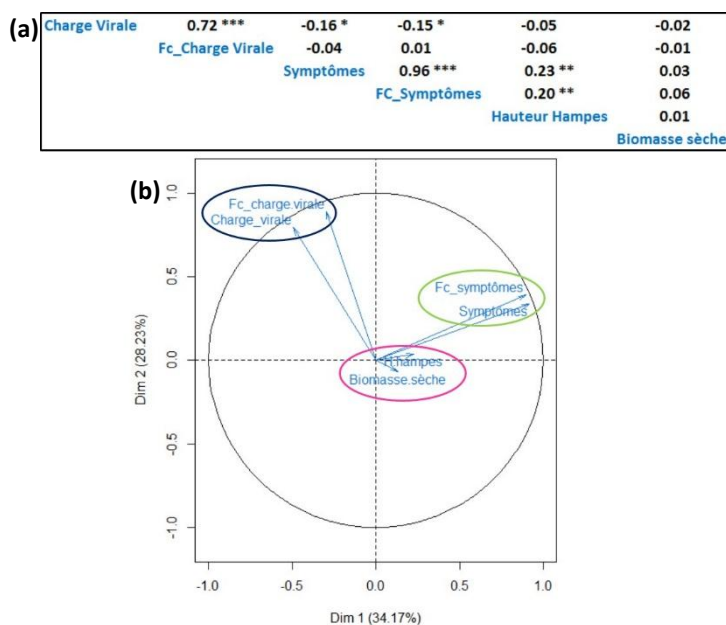


Figure A3.2] Tableau des corrélations (a) et graphique des variables (b) des différents traits phénotypiques mesurés en réponse au TuMV pour la population française TOU-A d'*A.thaliana* en conditions extérieures

La significativité des coefficients de corrélation de Spearman a été testée (Sign.codes : 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1). Les relations entre les différentes variables sont représentées dans le graphique des variables (b).

2. Analyse comparative des sous-groupes de la population française TOU-A

Cette population française d'*A. thaliana* est constituée de deux groupes d'accessions différents : les accessions collectées en 2002 (notées TOU-A1) et celles collectées en 2010 (notées TOU-A6). Chaque accession correspond aux graines collectées sur une plante individuelle selon le protocole d'échantillonnage décrit par Frachon et ses collaborateurs (2017). Les graines ont ensuite été semées en conditions contrôlées afin de diminuer les effets maternels possibles parmi les différentes accessions (Frachon *et al.*, 2017). Le graphique des individus obtenus à partir de l'analyse en composantes principales réalisée à partir de l'ensemble des données phénotypiques ne permet pas de distinguer de façon nette les deux sous-groupes. En revanche, les accessions TOU-A6 semblent être plus regroupées en haut à gauche du graphique alors que les accessions TOU-A1 semblent avoir une répartition plus dispersée. Les ellipses, correspondant à l'intervalle de confiance de l'estimation des coordonnées du barycentre de chaque groupe, se retrouvent côte à côte mais ne se superposent pas (Fig. A3.3). En regardant en parallèle le graphique des variables commenté précédemment (Fig. A3.2b), le regroupement de la sous-population TOU-A6 pourrait être expliqué en particulier par la charge virale. La comparaison des variables entre les accessions

TOU-A1 et TOU-A6, révèle des différences significatives en ce qui concerne la charge virale, les symptômes, la fréquence d'infection en lien avec les symptômes et la hauteur des hampes (Fig. A3.4). Ainsi, une accumulation virale plus importante est mesurée chez les accessions TOU-A6 alors qu'à l'inverse la sévérité et la fréquence des symptômes sont plus importantes chez les accessions TOU-A1. Cela vient corroborer les résultats précédents sur les corrélations entre les variables où charge virale et symptomatologie étaient indépendantes. Des différences significatives sont également observées avec des hauteurs de hampes plus élevées chez les accessions TOU-A1. Aucune différence significative n'est en revanche associée à la biomasse sèche (Fig. A3.4).

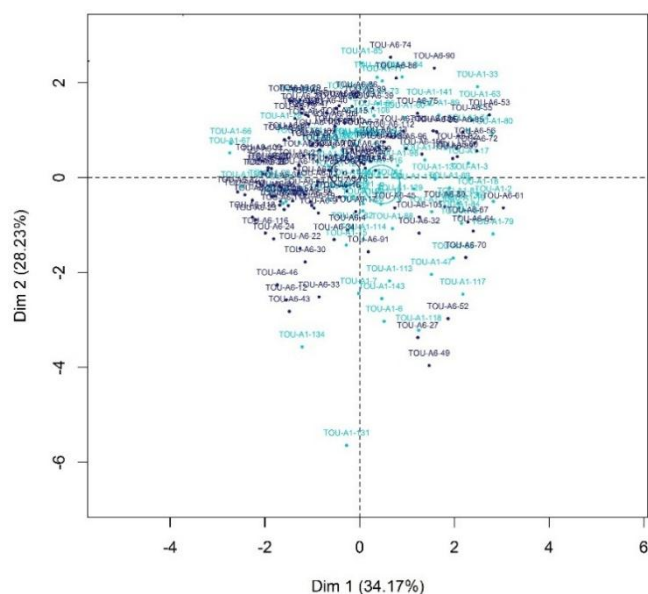


Figure A3.3 | Graphique des individus obtenu par l'analyse en composantes principales sur les données phénotypiques mesurées en réponse au TuMV pour la population française TOU-A d'*A.thaliana*

Les deux sous-populations sont représentées par deux couleurs différentes, bleu clair pour les accessions TOU-A1 et bleu foncé pour les accessions TOU-A6. Les ellipses correspondant à l'intervalle de confiance de l'estimation des coordonnées du barycentre de chaque groupe se retrouvent côte à côte mais ne se superposent pas.

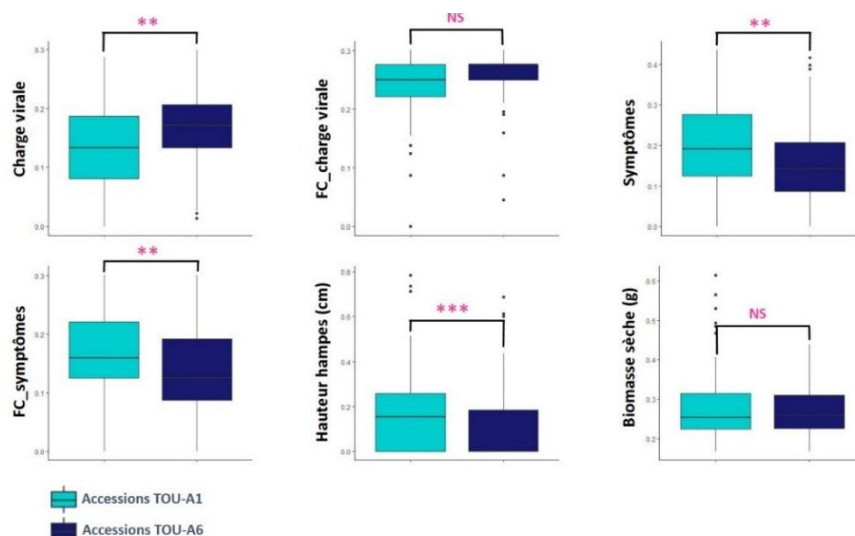


Figure A3.4 | Boxplots des différents traits phénotypiques mesurés chez les accessions TOU-A1 et TOU-A6
 Les comparaisons entre accessions de la population française TOU-A1 et TOU-A6 ont été effectuées pour l'ensemble des traits phénotypiques analysés. Des analyses statistiques ont été menées (test non-paramétrique de Wilcoxon). Les significativités statistiques sont représentées par des astérisques roses (*Signif.codes* : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1). Les différences non statistiquement significatives sont notées NS.

3. Analyse d'association : recherche des déterminants génétiques en réponse au TuMV dans la population française TOU-A

L'ensemble de la population a fait l'objet d'un génotypage dense avec 1.902.203 SNPs. L'analyse GWA a été réalisée avec le modèle mixte linéaire GEMMA déjà utilisé dans le cadre des analyses d'association réalisées sur la population mondiale d'*A. thaliana* (Rubio *et al.*, 2017 – *submitted*). Afin d'identifier de potentiels gènes candidats de la réponse des accessions de la population française TOU-A au TuMV, les 100 meilleurs SNPs de chaque analyse ont été sélectionnés. A partir de ces données, des zones comprenant des SNPs dont la distance qui les sépare est inférieure ou égale à 10kb ont été mises en évidence. Cette fenêtre de 10kb pour déterminer une région génomique d'intérêt a été choisie sur la base du déséquilibre de liaison estimé pour la population mondiale d'*A. thaliana* (Kim *et al.*, 2007). Elle est en ce qui concerne la population TOU-A suffisante puisque le déséquilibre de liaison a été estimé à 18 paires de base (Frachon *et al.*, 2017). Le seuil de p-value 10^5 ($-\log(p\text{-value}) > 5$) a été retenu afin de considérer une région comme potentiellement associée à la réponse au TuMV. Les gènes compris dans ces régions ont alors été identifiés. Une première analyse a été menée en utilisant l'ensemble de la population TOU-A réunie puis des analyses ont été réalisées de façon indépendante sur les deux groupes d'accessions TOU-A1 et TOU-A6.

Ces analyses GWA ont été menées sur les traits phénotypiques en lien avec la charge virale, les symptômes ainsi que leurs fréquences associées et la biomasse sèche. Les héritabilités (H^2) de ces traits phénotypiques ont été calculées (d'après Rubio *et al.*, 2017 – *submitted*). Les héritabilités sont élevées, la charge virale moyenne présente une H^2 de 0.784 et la biomasse sèche une H^2 de 0.721.

Les données relatives à la hauteur des hampes n'ont pas été exploitées en réponse à l'infection puisqu'aucune différence significative n'a été observée entre échantillons inoculés et mock-inoculés par le TuMV. L'effet de l'infection étant bien moins important que l'effet de l'environnement, une analyse GWA sur ces données ne reflèterait pas la réponse génétique d'*A. thaliana* au TuMV. Cela est d'ailleurs confirmé par la comparaison des analyses GWA menées sur les données de hauteur des hampes des échantillons inoculés et mock-inoculés. En effet, une même région est retrouvée, avec de forte p-value, au niveau du chromosome 4. Dans ce locus se trouve en particulier le gène FRIGIDA (AT4G00650) qui est impliqué dans les processus de floraison (Fig. A3.A2).

Les analyses GWA menées avec les 5 traits phénotypiques (à l'exception de la hauteur des hampes) sur l'ensemble des accessions de la population TOU-A, montrent des résultats assez proches entre les données de charges virales et de symptômes. Cela est particulièrement visible sur les manhattan plots au niveau du chromosome 2 où les mêmes pics de SNPs dépassent le seuil de significativité (Flèches noires - Fig. A3.5 a-b-c-d).

En revanche avec les analyses GWA menées sur les données normalisées de biomasse sèche, les p-value obtenues restent relativement faibles. Il n'existe pas – contrairement aux autres traits phénotypiques – de pic d'association clairement différencié (Fig. A3.5e). De plus, l'annotation des gènes identifiés dans les zones génomiques détectées n'ont pas permis de mettre en évidence de potentiels candidats qui nous paraissaient intéressants dans la réponse

au TuMV. Des analyses plus approfondies de ces résultats GWA ont donc été menées sur les traits phénotypiques en lien avec la charge virale et les symptômes.

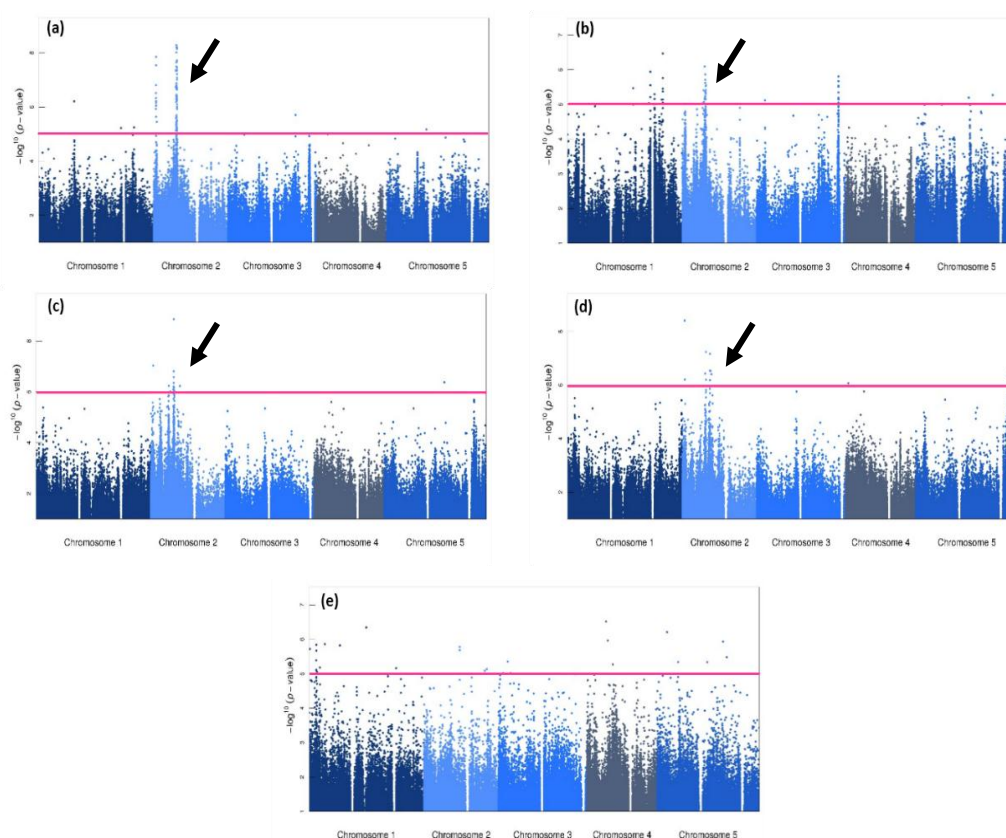


Figure A3.5| Manhattan plot des analyses GWA menées sur les données phénotypiques obtenues sur l'ensemble de la population TOU-A à 13 dpi (days post inoculation) avec le modèle GEMMA

Analyses GWA à partir (a) des valeurs de densité optique (charge virale), (b) des fréquences d'infection estimées à partir du nombre d'échantillons avec une DO positive sur le nombre total d'individus pour chaque génotype, (c) des notations de symptômes, (d) des fréquences estimées à partir du nombre d'échantillons présentant un symptôme sur le nombre total d'individus pour chaque génotype, (e) des données de biomasse sèche normalisées par les échantillons mock-inoculés de chaque génotype. Le trait rose représente le seuil de significativité placé à une p-value 10^5 ($-\log(p\text{-value}) > 5$).

Sur le diagramme de Venn construit à partir des données de charge virale et fréquences calculées ainsi qu'avec les données de symptômes et fréquences associées, deux régions du génome sont retrouvées de façon commune au niveau du chromosome 2 avec de fortes p-value (Fig. A3.6). La première région entre 6.039.018 et 6.057.267 bp retrouvée avec une valeur de $-\log(p\text{-value})$ maximale de 8.85 (obtenue avec les données symptômes) comporte 7 gènes potentiellement impliqués dans la réponse au TuMV. La deuxième région entre 6.178.020 et 6.205.760 bp identifiée avec une valeur de $-\log(p\text{-value})$ maximale de 8.19 (obtenue avec les données de charge virale) comporte 4 gènes potentiellement en lien avec la réponse au TuMV (Tab. A3.1a). D'autres régions sont également retrouvées spécifiquement avec une de ces quatre variables en lien avec la charge virale ou la symptomatologie. Il s'agit de régions retrouvées aussi avec de fortes p-value (Tab. A3.1b). Parmi ces régions, une d'entre elles – identifiées avec les fréquences en lien avec la charge virale – est localisée sur le

chromosome 3 entre 21.594.441 et 21.615.965 bp et comporte 8 gènes dont 7 caractérisés par un domaine TRAF like (Tab. A3.1b)

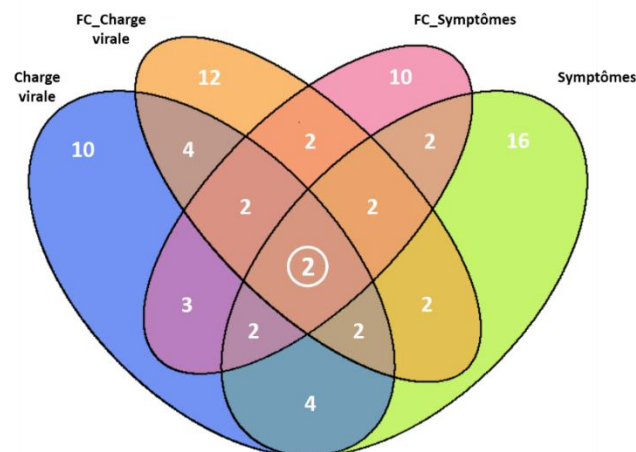


Figure A3.6 | Diagramme de Venn des zones du génome identifiées par les analyses GWA

Les données ont été obtenues suite aux analyses GWA réalisées à partir des données de charge virale (en bleu), de fréquence en lien avec la charge virale (en orange), de symptômes (en vert) et de fréquence en lien avec les symptômes (en rose). Les deux régions du génome retrouvées de façon commune sur l'ensemble des analyses sont entourées en blanc.

Des analyses GWA ont également été menées de façon indépendante sur les données obtenues sur les deux sous-populations TOU-A1 et TOU-A6 avec les variables phénotypiques en lien avec la charge virale, les symptômes et leurs fréquences associées. Les résultats concernant la distribution de chaque trait phénotypique analysé et les manhattan plots obtenus suite aux analyses GWA sont représentés en annexes Fig. A3.A3 et Fig. A3.A4 pour la sous-population TOU-A1 et TOU-A6 respectivement.

Les résultats obtenus sur la sous-population TOU-A1 montrent qu'uniquement 4 régions du génome co-localisent avec les régions identifiées sur l'ensemble de la population TOU-A avec les données en lien avec l'accumulation du virus (charge virale et fréquence associée). Il s'agit de quatre régions localisées au niveau du chromosome 2 (Tab. A3.2a). Aucune autre co-localisation n'est décrite avec les autres données phénotypiques. En revanche, d'autres régions génomiques présentant de fortes p-value sont identifiées (Tab. A3.2b).

Les régions génomiques identifiées avec les analyses GWA sur la sous-population TOU-A6 et qui co-localisent avec les résultats obtenus sur l'ensemble de la population TOU-A sont plus nombreuses par rapport à la sous-population A1, 13 régions au total (Tab. A3.3a). D'autres régions spécifiques sont également identifiées avec de fortes p-value (Tab. A3.3b).

Tableau A3.1 | Zones du génome mises en évidence par les analyses GWA sur la population TOU-A menées sur les différents traits phénotypiques mesurés

Trait phénotypique	Chr	Région Génomique (Début..Fin)	Meilleur SNP	Log10(p-val)	Gènes	Description des gènes
(a)	2	6.039.018..6.057.267	6.044.716	8.85	AT2G14260	Proline iminopeptidase (PIP) - Involved in proteolysis
Charge virale		6.039.018..6.048.803	6.048.025	8.19	AT2G14270	Cytochrome C oxidase polypeptide VIB family protein
Fc_Charge Virale		6.040.919..6.048.025	6.044.598	5.86	AT2G14282	SCR-like 18
Symptomatologie		6.043.395..6.057.267	6.044.716	8.85	AT2G14285	Small nuclear ribonucleoprotein family protein
FC_Symptômes		6.044.716..6.045.068	6.044.716	5.26	AT2G14288	Unknown protein
					AT2G14289	Unknown protein
					AT2G14290	F-box domain, cyclin like - Unknown function
		6.178.020..6.205.760	6.193.069	8.19	AT2G14510	Leucine-rich repeat protein kinase family protein
Charge virale	2	6.178.020..6.198.332	6.193.069	8.19	AT2G14520	Unknown protein
Fc_Charge Virale		6.178.020..6.198.332	6.178.584	5.53	AT2G14530	TRICHOME BIREFRINGENCE LIKE 13
Symptomatologie		6.191.399..6.205.760	6.191.809	6.04	AT2G14540	Serpin 2 (SRP2) - Involved in DNA repair, response to DNA damage stimulus
FC_Symptômes		6.191.393..6.205.760	6.194.962	7.24		
(b)						
Symptomatologie	2	593.084..600.890	595.300	7.03	AT2G02240	Maternal effect embryo arrest 66
					AT2G02250	Phloem protein 2-B2 - N-terminal protein myristoylation
					AT2G02280	Phloem protein 2-B4 - N-terminal protein myristoylation
Charge virale	2	618.712..624.161	619.472	7.85	AT2G02350	SKP1 Interacting partner 3
Fc_Charge Virale		619.292..619.472	619.292	5.01	AT2G02360	Phloem protein 2-B10
					AT2G02370	SNARE associated Golgi protein family
Charge virale	2	5.977.676..6.015.988	5.994.683	8.28	AT2G14170	Aldehyde dehydrogenase 6B2 - response to oxydative stress
FC_Symptômes	2	7.267.648..7.274.139	7.267.678	7.17	AT2G16730	Glycosyl hydrolase family 35 protein - lactose catabolic process
					AT2G16740	Ubiquitin conjugating enzyme 29
					AT2G16750	Protein kinase protein with adenine nucleotide alpha hydrolase like domain
Fc_Charge Virale	3	21.594.441..21.615.965	21.608.485	5.81	AT3G58360	TRAF like family protein
					AT3G58370	TRAF like family protein
					AT3G58380	TRAF like family protein
					AT3G58390	Eucaryotic release factor 1 (eRF1) family protein
					AT3G58400	TRAF like family protein
					AT3G58410	TRAF like family protein
					AT3G58420	TRAF like family protein
					AT3G58430	TRAF like family protein
FC_Symptômes	5	24.378.433..24.383.289	24.383.289	6.41	AT5G60660	Plasma membrane intrinsic protein 2;4 - Hydrogen peroxyde transmembrane transport
					AT5G60670	Ribosomal protein L11 family protein - translation, ribosome biogenesis

(a) Régions génomiques **communes** identifiées à partir des données de charge virale, symptômes et leurs fréquences associées **(b)** Régions génomiques présentant de fortes p-value mais **spécifiques** à un trait phénotypique. Le gène associé au SNP présentant la best p-value est représenté en gras. Lorsque le best SNP ne correspond à aucun gène, le ou les gènes à proximité immédiate sont représentés en gras.

Tableau A3.2 | Zones du génome identifiées à partir des analyses GWA à partir des 4 traits phénotypiques obtenus sur la sous-population TOU-A1

Trait phénotypique	Chr	Région Génomique (Début..Fin)	Meilleur SNP	Log10(p-val)	Gènes	Description des gènes
(a)	Charge Virale	620.778..623.523	620.778	5.9	AT2G02360 AT2G02370	Phloem protein 2-B10 SNARE associated Golgi protein family
	Charge virale	5.998.316..6.021.308	6.016.028	8.53	AT2G14206 AT2G14210	Unknown protein AGAMOUS like 44 - response to nutrient lateral root development
	Charge virale	6.033.712..6.036.291	6.033.712	5.29	AT2G14247 AT2G14255	Unknown protein Ankyrin repeat family protein with DHHC zinc finger domain
	FC_Charge Virale	721.932..746.874	728.431	6.36	AT2G02630 AT2G02640 AT2G02650 AT2G02660 AT2G02680	Cysteine/Histidine rich C1 domain family protein Cysteine/Histidine rich C1 domain family protein Ribonuclease H-like superfamily protein F-box associated ubiquitination effector family protein Cysteine/Histidine rich C1 domain family protein
(b)	FC_Charge Virale	971.831..972.766	971.831	8.35	AT2G03320 AT2G03230	Fucosyltransferase 1 (FT1) - plant type cell wall biogenesis GCK domain containing protein
	Symptomatologie	18.619.756..18.649.707	18.628.872	6.47	AT5G45910 AT5G45920 AT5G45930 AT5G45940 AT5G45950 AT5G45960 AT5G45970 AT5G45973 AT5G45980	GDSL-like Lipase/Acylhydrolase superfamily protein - lipid metabolic process SGNH hydrolase type esterase superfamily protein Magnesium chelatase i2 - chlorophyll biosynthetic process Nudix hydrolase homolog 11 GDSL-like Lipase/Acylhydrolase superfamily protein - lipid metabolic process GDSL-like Lipase/Acylhydrolase superfamily protein - lipid metabolic process RAC like 2 - small GTPase mediated signal transduction Unknown protein WUSCHEL related homeobox 8
	Charge virale	24.366.720..24.388.200	24.378.433	6.56	AT5G60615 AT5G60620 AT5G60630 AT5G60640 AT5G60650 AT5G60660 AT5G60670 AT5G60680	Defensin like (DEFL) family protein glycerol 3 phosphatse acyltransferase 9 Unknown protein PDI like 1-4 Unknown protein Plasma membrane intrinsic protein 2;4 Ribosomal protein L11 family protein Unknown protein

(a) Régions génomiques **communes** à celles identifiées à partir des données obtenues sur l'ensemble de la population TOU-A **(b)** Régions génomiques présentant de fortes p-value mais **spécifiques** aux données des accessions de la sous-population TOU-A1. Le gène associé au SNP présentant la best p-value est représenté en gras. Lorsque le best SNP ne correspond à aucun gène, le ou les gènes à proximité immédiate sont représentés en gras.

Tableau A3.3 | Zones du génome identifiées à partir des analyses GWA à partir des 4 traits phénotypiques obtenus sur la sous-population TOU-A6

Trait phénotypique	Chr	Région Génomique (Début..Fin)	Meilleur SNP	Log10(p-val)	Gènes	Description des gènes	
(a)	FC_Charge Virale	21.942.406..21.945.855	21.944.652	7.31	AT1G59720	Chlororespiratory reduction 28	
					AT1G59722	Unknown protein	
	Charge virale	24.897.578..24.898.968	24.898.968	5.5	AT1G66750	CDK-activating kinase 4 (CAK4)	
	Charge virale Symptomatologie	593.084..600.890	593.084	593.084	6.37	AT2G02240	Maternal effect embryo arrest 66
						AT2G02250	Phloem protein 2-B2
						AT2G02280	Phloem protein 2-B4
	Symptomatologie Charge Virale	6.041.901..6.048.803	6.044.503..6.046.744	6.044.716	6.99	AT2G14260	Proline iminopeptidase (PIP)
						AT2G14270	Cytochrome C oxidase polypeptide VIB family protein
	Charge virale	618.712..619.243	619.243	5.33	AT2G02350	SKP1 interacting partner 3	
					AT2G02360	Phloem protein 2-B10	
	Symptomatologie Charge Virale	6.177.192..6.197.474	6.191.809..6.197.474	6.195.341	6.68	AT2G14520	Unknown protein
						AT2G14530	Trichome birefringence like 13
						AT2G14540	Serpin 2 - DNA repair, response to DNA damage stimulus
FC_Symptômes	24.201.871..24.212.687	24.205.898	6.93	AT5G60100	Pseudo response regulator 3		
				AT5G60110	Pumilio 18		
				AT5G60120	Target of early activation tagged (EAT) 2		
				AT5G60130	AP2/B3-like transcriptional factor family protein		
FC_Symptômes	26.606.394..26.627.149	26.615.875	5.27	AT5G66658	Unknown protein		
				AT5G66660	Unknown protein		
				AT5G66670	Unknown protein		
				AT5G66675	Unknown protein		
				AT5G66680	DEFECTIVE GLYCOSYLATION (DGL1)		
				AT5G66690	UGT72E2		
(b)	Charge virale	24.928.935..24.954.434	24.929.709	7.31	AT1G66820	Glycine rich protein	
					AT1G66830	Leucine rich repeat protein kinase family protein	
					AT1G66840	Unknown protein	
					AT1G66850	Bifunctional inhibitor/lipid transfert protein/seed storage 2S albumin superfamily protein	
					AT1G66852	Carbohydrate binding X8 domain superfamily protein	
					AT1G66855	Carbohydrate binding X8 domain superfamily protein	
					AT1G66860	Class I glutamine amidotransferase like superfamily protein	
					AT1G66870	Carbohydrate binding X8 domain superfamily protein	
					AT1G66880	Protein kinase superfamily protein	
					Symptomatologie	24.198.667..24.219.988	24.206.063
	AT5G60100	Pseudo response regulator 3					
	AT5G60110	Pumilio 18					
	AT5G60120	Target of early activation tagged					
	AT5G60130	AP2/B3 like transcriptional factor family protein					
		AT5G60140	AP2/B3 like transcriptional factor family protein				
	AT5G60142	AP2/B3 like transcriptional factor family protein					
	AT5G60150	Unknown protein					

(a) Régions génomiques **communes** à celles identifiées à partir des données obtenues sur l'ensemble de la population TOU-A (b) Régions génomiques présentant de fortes p-value mais **spécifiques** aux données des accessions de la sous-population TOU-A1. Le gène associé au SNP présentant la best p-value est représenté en gras. Lorsque le best SNP ne correspond à aucun gène, le ou les gènes à proximité immédiate sont

Discussion

Les précédentes analyses s'étaient intéressées à la réponse d'une population mondiale d'*A. thaliana* au TuMV en conditions extérieures (Rubio *et al.*, 2017 – *submitted*). Parmi les diverses accessions étudiées, plusieurs avaient une origine française. Une diversité dans la réponse au TuMV de ces accessions françaises dans ces conditions extérieures avait été alors observée. Il a donc été choisi de travailler avec une population locale française, la population TOU-A. Cette population étant constituée d'accessions ayant été échantillonnées en 2002 (TOU-A1) et en 2010 (TOU-A6). Les objectifs étant d'une part de caractériser la réponse d'accessions locales à une infection par le TuMV et d'autre part de comparer les bases génétiques de la réponse au TuMV entre cette population locale française et la population mondiale étudiée précédemment.

L'analyse de la réponse de la population TOU-A au TuMV a permis de mettre en évidence une diversité de réponses pour l'ensemble des six traits phénotypiques analysés. Cette analyse phénotypique a permis de confirmer que des différences existent entre les études menées dans ces conditions extérieures et celles réalisées en conditions contrôlées. C'est notamment le cas en ce qui concerne la hauteur des hampes qui semble fortement influencée par les conditions environnementales. Cette variable, dans ces conditions extérieures n'est plus aussi bien associée à l'infection par le TuMV contrairement à ce qui était observé en conditions contrôlées (données chambre de culture non présentées). D'ailleurs, aucune différence significative n'a été mise en évidence entre les échantillons inoculés et mock-inoculés par le TuMV. De plus, les analyses GWA menées en parallèle sur ces données de hauteur de hampe mesurée chez les échantillons inoculés et mock-inoculés par le virus, mettent en évidence une réponse génétique commune. Plus précisément, une région sur le chromosome 4 a été identifiée dans les deux analyses avec de fortes p-value. Dans ce locus, se trouve le gène FRIGIDA impliqué dans les phénomènes de floraison. Dans l'étude menée par Brachi et ses collaborateurs (2013) les analyses GWA menées sur des sets d'accessions françaises font toujours ressortir un fort pic d'association en lien avec ce gène (Brachi *et al.*, 2013). Cela renforce le fait que la hauteur des hampes ne constitue pas un bon trait phénotypique pour l'étude des bases génétiques de la réponse d'*A. thaliana* au TuMV en conditions extérieures. Cela souligne l'importance d'avoir des phénotypes robustes associés à la réponse biologique étudiée puisqu'ils constituent la base de toutes les analyses. L'ensemble de ces observations, soulignent l'intérêt et l'apport des essais en conditions extérieures où le comportement et donc les réponses aux stress des plantes peuvent fortement varier par rapport aux conditions contrôlées.

La variabilité phénotypique ainsi que le génotypage dense effectué sur cette population TOU-A a conduit à s'intéresser aux bases génétiques de la réponse au TuMV en utilisant une approche de génétique d'association. Ces analyses de GWA ont été menées sur l'ensemble de la population TOU-A réunie mais également de façon indépendante sur les deux sous-populations.

L'analyse GWA menée sur l'ensemble de la population TOU-A a permis de mettre en évidence des réponses génétiques différentes en fonction des traits phénotypiques étudiés. Ainsi, deux principales régions ont été identifiées sur le chromosome 2 avec de fortes p-value

à partir des analyses GWA menées sur les quatre traits phénotypiques en lien avec la charge virale et la symptomatologie. Parmi ces deux loci, le gène AT2G14540 qui code pour une serpine (SERPIN 2) est un gène candidat intéressant. En effet, les serpines sont des protéines ayant la capacité d'inhiber des protéases à sérine. Les mécanismes de protéolyse étant indispensables dans la croissance, le développement, la réponse à des stress des plantes, les serpines pourraient avoir un rôle crucial sur l'ensemble de ces phénomènes (Roberts & Hejgaard, 2008). Une étude récente a d'ailleurs mis en évidence le rôle des serpines dans le contrôle contre des ravageurs des cultures (Alvarez-Alfageme *et al.*, 2011).

D'autres gènes candidats intéressants – identifiés à partir des données de charge virale et symptomatologie – de la réponse au virus sont les gènes AT2G02360, AT2G02250 et AT2G02280 qui codent pour des protéines phloémiennes PP2 (« Phloem Protein 2 »). Ces gènes ont été identifiés sur deux régions proches du chromosome 2, la zone comprise entre 593.084 et 600.890 bp et celle entre 618.712 et 624.161 bp, mises en évidence avec les données relatives aux symptômes et à la charge virale respectivement. Sur l'ensemble de ces deux régions, un cluster de 8 gènes codant pour des protéines phloémiennes PP2 a été identifié dont font partie les trois gènes décrits précédemment. Ces protéines PP2 sont largement présentes et conservées chez les plantes vasculaires. Ces protéines présentent une activité lectine et d'interaction avec d'autres protéines du phloème (Dinant, 2003). Deux études indépendantes ont montré que les protéines PP2 avaient la capacité de se lier à divers ARN. Cela a permis de mettre en évidence leur rôle dans le mouvement à longue distance de viroïdes chez des cucurbitacées (Owens *et al.*, 2001; Gomez & Pallas, 2004). Dans une étude plus récente, des protéines PP2 chez le concombre interagissent avec un polérovirus, le *Cucurbit aphid borne yellow virus* (CABYV) en stimulant la transmission du virus via des pucerons (Bencharki *et al.*, 2010). Chez *A. thaliana*, les gènes PP2-like constitue une large famille multigénique avec 30 gènes caractérisés. Cela suggère que des duplications sont certainement à l'origine de différents sous-groupes de gènes PP2-like. Ces duplications ont certainement conduit à l'acquisition de différentes fonctions (Dinant, 2003). Ce qui est particulièrement intéressant c'est le mouvement de cellule à cellule à travers les plasmodesmes dont sont capables les protéines PP2. Or, les virus utilisent les plasmodesmes pour pouvoir se propager de cellule à cellule et donc pouvoir se propager à l'ensemble de la plante (Kumar *et al.*, 2015).

Ces protéines phloémiennes identifiées sont d'autant plus intéressantes qu'un projet concernant l'étude de la régulation du transcriptome du phloème lors de l'infection par des potyvirus a été développé au sein de l'Equipe Virologie Végétale de l'UMR BFP par S. German-Retana. Dans le cadre de ce projet, les ARNm en cours de traduction au niveau des cellules compagnes du phloème et de la gaine périvasculaire seront isolés afin de décrypter les modifications du transcriptome du phloème lors d'une infection par le TuMV. Il sera intéressant une fois les résultats de ce projet obtenus de regarder si ces gènes codant pour des protéines PP2 identifiés par GWA sont retrouvés dérégulés dans le phloème lors d'une infection virale.

L'identification, à partir des données de fréquence d'infection basée sur la charge virale, d'une région au niveau du chromosome 3 entre 21.594.441 et 21.615.965 bp est particulièrement intéressante puisque lors des analyses GWA effectuées sur la population

mondiale en conditions extérieures, la zone comprise entre 21.581.213 et 21.592.778 bp a été identifiée avec l'ensemble des traits phénotypiques étudiés. Ces deux régions sont donc situées à moins de 10kb l'une de l'autre. Entre ces deux régions proches se trouve un cluster de 11 gènes codant pour des protéines de la famille des TRAF. Plus particulièrement, la région mise en évidence par GWA sur les données de la population mondiale co-localise avec le gène *RTM3* et le locus *sha3* déjà identifiés comme impliqués dans la réponse à des potyvirus (Whitham *et al.*, 1999; Decroocq *et al.*, 2006; Cosson *et al.*, 2012). La région identifiée sur la population TOU-A est située juste après celle décrite avec la population mondiale. Cela renforce l'idée que chez *A. thaliana* les protéines de la famille des TRAF pourraient avoir un rôle fondamental dans la réponse à des virus. Leur rôle réel dans cette réponse ainsi que leur rôle biologique hors infection virale reste très largement inconnu.

Les **analyses GWA menées de façon spécifique sur chacune des sous-populations TOU-A1 et TOU-A6** ont permis de mettre en évidence des régions communes avec celles identifiées sur l'ensemble de la population. Ainsi, est retrouvée en particulier une zone au niveau du chromosome 2 (données TOU-A1 entre 620.778.623 et 623.523 bp ; données TOU-A6 entre 618.712 et 619.243 bp) où se trouve le gène AT2G0260 qui code d'ailleurs pour une protéine du phloème 2-B10.

Des régions génomiques spécifiquement identifiées à partir des données de chaque sous-population ressortent avec de fortes p-value. C'est le cas notamment d'un locus au niveau du chromosome 5 (données TOU-A6 entre 24.366.720 et 24.388.200 bp) où est retrouvé le gène AT5G60640 qui code pour protéine disulfide isomerase (PDI). Une autre protéine de la même famille, PDI-like 5-1 a été décrite chez l'orge en réponse à deux bymovirus, le *Barley yellow mosaic virus* (BaYMV) et le *Barley mild mosaic virus* (BaMMV) (Yang *et al.*, 2014). Cette famille de protéines pourrait se révéler être très intéressante. Néanmoins, l'évaluation de lignée KO d'*A. thaliana* affectées au locus PDIL5-1 (Protein Disulfide Isomerase Like 5-1, AT1G07960) pour la réponse à cinq potyvirus différents (*Plum pox virus*-GFP, *Turnip mosaic virus*, *Tobacco etch virus*, *Water melon virus* and *Lettuce mosaic virus*) selon deux modes d'inoculation (agro-inoculation et inoculation mécanique) n'a montré aucune différence de sensibilité par rapport à la lignée non-KO (Projet ANR – PLANT-KBEE 2013 – COBRA).

L'ensemble des données phénotypiques et de l'analyse d'association permettent d'illustrer l'intérêt d'étudier une population locale. L'étude de cette population locale française TOU-A, permet de conforter l'implication du locus du chromosome 3 autour de la position 21.608.485. Ce locus voisin du locus *RTM3* déjà largement caractérisé comme impliqué dans la réponse d'*A. thaliana* aux potyvirus en conditions contrôlées et en conditions extérieures – code pour des protéines avec des domaines TRAF. Il semble donc que *RTM3* et/ou plusieurs des protéines à domaine TRAF soient des intervenants majeurs et constants de la réponse d'*A. thaliana* au TuMV mais plus largement à différents potyvirus.

Cette analyse met également en évidence des bases génétiques de la réponse au TuMV en conditions extérieures différentes de celles caractérisées avec la population mondiale. En effet, aucun autre locus caractérisé dans ce travail ne co-localise avec une région du génome identifiée lors de l'étude menée sur la population mondiale (Rubio *et al.*, 2017 – *submitted*). En ce sens, l'étude de cette population locale permet de mettre en évidence de nouvelles

régions du génome avec de potentiels candidats de l'interaction *A. thaliana*/TuMV. D'une certaine façon, ce résultat n'est pas surprenant. En effet, il a été montré que des accessions d'*A. thaliana* collectées dans un environnement proche présentent une empreinte génétique spécifique qui est le reflet des processus locaux à l'origine de divers patterns de structures génétiques observés en fonction de l'échelle d'étude (Bomblies *et al.*, 2010). Une autre étude a été menée par Brachi et ses collaborateurs (2013) sur différents traits phénologiques caractérisés au sein de différentes groupes d'accessions d'*A. thaliana* correspondant à différentes échelles géographiques. Les analyses GWA menées montrent que les régions génomiques identifiées, en lien avec la variabilité naturelle des traits phénotypiques étudiés, dépendent de l'échelle géographique considérée (Brachi *et al.*, 2013).

De plus, la détection de ces nouvelles régions non identifiées lors de l'étude de la population mondiale corrobore l'idée que les analyses GWA menées sur des populations locales permettent de dépasser certaines limites de cette approche génétique en lien notamment avec la structuration des populations, la détection des allèles rares et l'hétérogénéité allélique (Bergelson & Roux, 2010; Brachi *et al.*, 2013).

Lors de l'analyse phénotypique des différences significatives sur les traits analysés (à l'exception de la biomasse sèche) ont été mises en évidence entre les accessions des deux sous-populations (TOU-A1 et TOU-A6). Ainsi, les accessions TOU-A6 accumulent plus de virus (charge virale moyenne plus élevée) alors que les accessions TOU-A1 expriment des symptômes liés à l'infection plus sévères. Cette différence de réponse pose la question de l'adaptation de cette population au cours du temps, les deux sous-populations ayant été échantillonnées à huit ans d'intervalle. Au cours de cette période, cette population naturelle a été exposée à différentes contraintes biotiques et abiotiques. L'architecture génétique a pu être modifiée à l'origine d'une évolution phénotypique rapide. Ces aspects d'adaptations rapides aux contraintes extérieures sont des questions largement traitées en écologie évolutive. Cela est d'autant plus vrai dans le contexte de changement climatique actuel, ce sont donc le plus souvent des adaptations à des contraintes abiotiques qui sont étudiées (Bradshaw & Holzapfel, 2006; Hoffmann & Sgrò, 2011; Kopp & Matuszewski, 2014). Dans le cadre de cette étude, il est difficile de pouvoir aborder la notion d'adaptation vis-à-vis de l'infection virale. En effet, pour qu'il puisse y avoir une adaptation, les plantes doivent être soumises à une contrainte permanente. Or, dans le cas de la population TOU-A, aucune information n'est disponible sur l'éventuelle présence de pressions biotiques et en particulier de pressions virales. Pour cela, il aurait fallu mettre en place un dispositif d'échantillonnage tous les ans par exemple, avec une analyse de détection des virus afin d'évaluer une éventuelle pression virale.

Ces différences phénotypiques observées et en particulier la sensibilité plus grande des accessions TOU-A6 vis-à-vis du TuMV pourrait être le reflet des effets de l'histoire démographique de cette population TOU-A entre 2002 et 2010. Ainsi, des adaptations à d'autres contraintes subies par cette population, auraient pu avoir des conséquences sur la sensibilité au TuMV. Cette hypothèse permet de faire référence à la notion de « cost of complexity » en lien notamment avec l'effet pléiotropique des gènes (Tyler *et al.*, 2009; Wagner & Zhang, 2011). Cela voudrait donc dire que les accessions se sont adaptées aux contraintes rencontrées au cours des 8 générations. Les mutations avantageuses vis-à-vis de

ces contraintes ont peut-être une conséquence dans la réponse aux virus, aboutissant dans ce cas à une augmentation de la sensibilité des accessions.

Néanmoins, cette notion d'adaptation est dans le cadre de cette étude à manipuler avec précaution. Cela est d'autant plus vrai que la réponse analysée est en relation avec un stress appliqué de façon ponctuelle sur des accessions dont l'histoire vis-à-vis des infections virales n'a pas été étudiée.

L'ensemble de ce travail mené sur la population locale TOU-A permet de compléter le travail de caractérisation des bases génétiques de la réponse d'*A. thaliana* au TuMV en conditions extérieures commencé par l'étude d'une population mondiale. La combinaison d'analyses GWA menées sur des accessions à différentes échelles géographiques permet d'identifier de potentielles nouvelles sources de résistance chez *A. thaliana* en réponse au TuMV en conditions extérieures. Cette étude a permis de montrer les avantages de travailler avec des populations locales dans la recherche des bases génétiques d'une réponse biologique. L'objectif étant de pouvoir par la suite compiler l'ensemble des résultats obtenus afin de faire ressortir les candidats les plus intéressants pour une analyse permettant dans le meilleur des cas de valider de nouveaux facteurs de l'hôte impliqués dans l'interaction avec des virus.

ANNEXES

Tableau A3.A1 Liste des accessions française TOU-A1 communes entre les différents essais extérieurs (2013-2014-2016)

Accessions TOU-A1	Charge Virale (DO 405 nm)			Fc_charge virale		
	2013	2014	2016	2013	2014	2016
TOU-A1-111	0.3612	0.1301	0.4673	100.00	66.67	77.78
TOU-A1-114	0.0612	0.0428	0.2177	33.33	33.33	66.67
TOU-A1-115	0.0985	0.4031	0.1847	66.67	66.67	66.67
TOU-A1-116	0.0650	0.1466	0.1961	66.67	66.67	100.00
TOU-A1-120	0.3254	0.3359	0.5219	66.67	100.00	88.89
TOU-A1-125	0.4469	0.1505	0.7819	100.00	33.33	100.00
TOU-A1-128	0.3032	0.0753	0.4275	100.00	50.00	66.67
TOU-A1-129	0.1653	0.0396	0.2356	33.33	33.33	75.00
TOU-A1-133	0.1084	0.1642	0.2300	66.67	66.67	55.56
TOU-A1-143	0.2698	0.0541	0.1001	66.67	33.33	42.86
TOU-A1-23	0.4235	0.1183	0.6253	100.00	66.67	77.78
TOU-A1-37	0.3436	0.0348	0.4987	100.00	33.33	100.00
TOU-A1-39	0.4539	0.0417	0.4623	100.00	33.33	100.00
TOU-A1-41	0.5018	0.4510	0.5274	100.00	100.00	88.89
TOU-A1-60	0.3056	0.0673	0.5385	100.00	33.33	100.00
TOU-A1-61	0.2781	0.1467	0.1954	100.00	66.67	55.56
TOU-A1-62	0.2093	0.0476	0.4620	66.67	33.33	100.00
TOU-A1-63	0.1300	0.0745	0.4476	51.16	30.00	88.89
TOU-A1-66	0.2179	0.2505	0.7946	100.00	50.00	100.00
TOU-A1-73	0.2971	0.1408	0.7295	100.00	66.67	88.89
TOU-A1-74	0.1675	0.0518	0.3115	100.00	33.33	88.89
TOU-A1-77	0.2700	0.1323	0.8491	100.00	33.33	100.00
TOU-A1-98	0.2525	0.0646	0.4009	66.67	33.33	77.78
TOU-A1-18	0.0327	0.0304	0.1765	0.00	0.00	66.67
TOU-A1-65	0.0473	0.0317	0.1227	0.00	0.00	44.44
TOU-A1-69	0.0089	-0.0016	0.1821	0.00	0.00	77.78
TOU-A1-84	0.0306	0.0276	0.1493	0.00	0.00	66.67

27 accessions appartenant à la sous-population TOU-A1 ont été testées au cours des trois essais extérieurs en 2013, 2014 et 2016. La charge virale (DO à 405 nm) ainsi que la fréquence d'infection en lien avec la charge virale (Fc_charge virale) ont été notées pour chacune des accessions sur les trois années d'étude. Les accessions en bleus sont celles avec une réponse au TuMV similaire sur les trois ans (il s'agit d'accessions sensibles au TuMV). Les accessions en violet sont celles avec une réponse qui varie en fonction de l'essai.

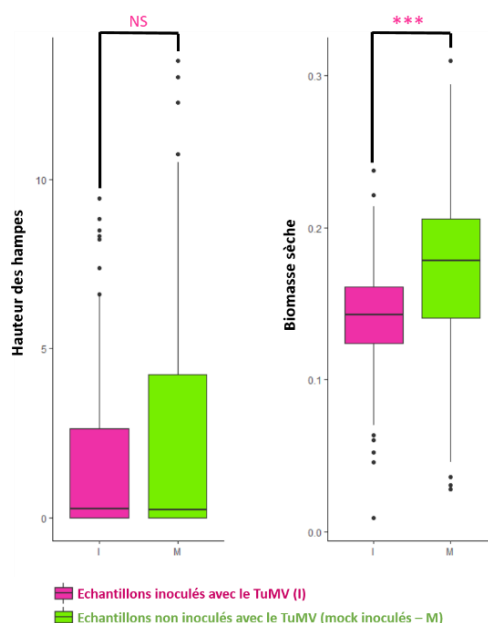


Figure A3.A1 Comparaison de la hauteur des hampes et de la biomasse sèche entre échantillons inoculés (I) et non inoculés (M) par le TuMV en conditions extérieures

Sign.codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

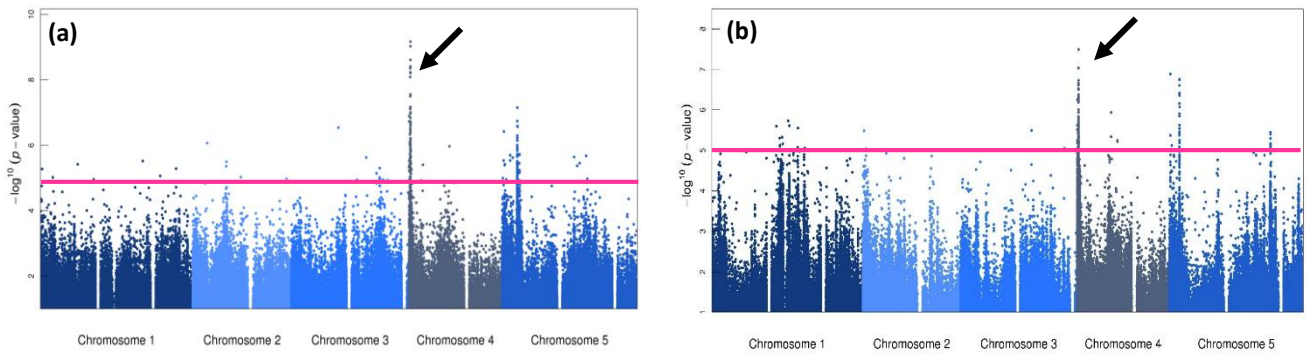


Figure A3.A2 Manhattan plot obtenu avec les analyses GWA sur les données de hauteur de hampe des échantillons mock-inoculés (a) et inoculés (b) avec le TuMV

Analyses GWA menées en parallèle chez les échantillons mock-inoculés (a) et inoculés (b) avec le TuMV sur les données de hauteur de hampe. Mise en évidence d'une zone commune notamment au niveau du chromosome 4 (flèches noires) où se trouve le gène FRIGIDA impliqué dans la floraison. La ligne rose représente de le seuil de significativité fixé à $-\log(p\text{-value}) = 5$

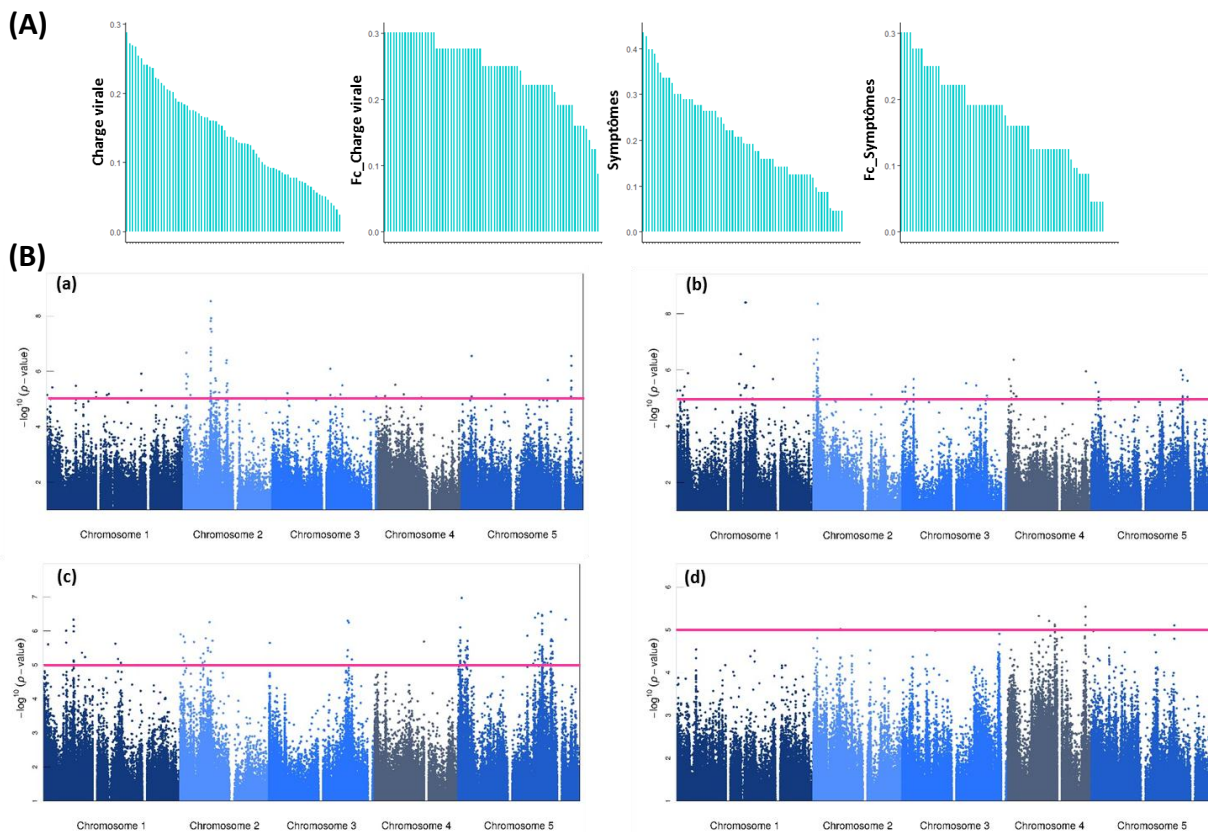


Figure A3.A3 Données phénotypiques et *Manhattan plot* concernant la sous-population TOU-A1

(A) Distribution des différentes données phénotypiques mesurées à 13 dpi (B) Résultats des analyses GWA sous la forme de *Manhattan plot*. Le seuil de significativité placé à $-\log(p\text{-value}) = 5$ est représenté par le trait rose. Les analyses GWA ont été menées sur les données de charge virale (a), les fréquences associées à la charge virale (b), les symptômes (c), les fréquences associées aux symptômes (d).

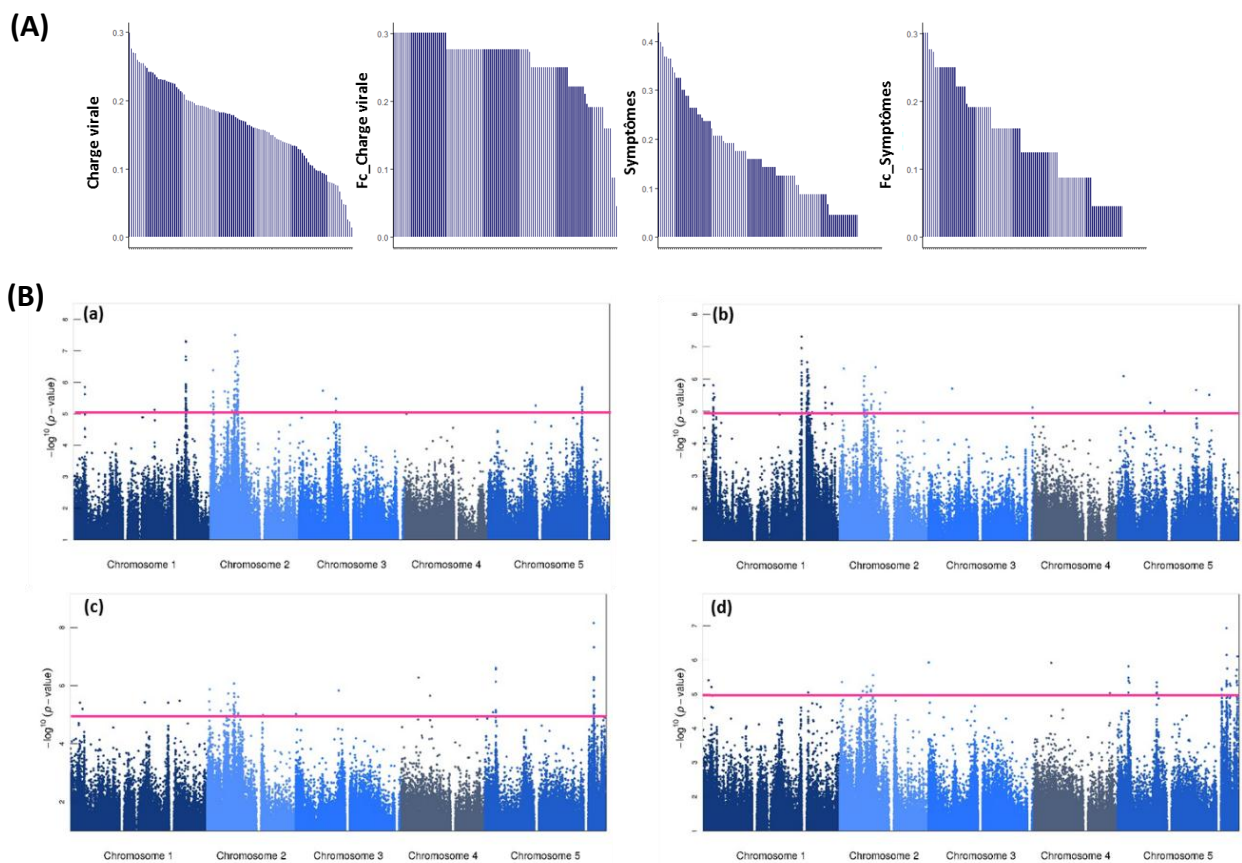


Figure A3.A4 Données phénotypiques et *Manhattan plot* concernant la sous-population TOU-A6

(A) Distribution des différentes données phénotypiques mesurées à 13 dpi **(B)** Résultats des analyses GWA sous la forme de *Manhattan plot*. Le seuil de significativité placé à $-\log(p\text{-value}) = 5$ est représenté par le trait rose. Les analyses GWA ont été menées sur les données de charge virale **(a)**, les fréquences associées à la charge virale **(b)**, les symptômes **(c)**, les fréquences associées aux symptômes **(d)**

Chapitre 1 - PARTIE B

**Phénotypage métabolique fin de la réponse
au TuMV en conditions extérieures**

B-1. Le métabolisme primaire et secondaire discrimine la sensibilité d'*A. thaliana* au TuMV en conditions extérieures

Cette partie est présentée sous la forme d'un article scientifique

Primary and secondary metabolic content discriminate *Arabidopsis thaliana* susceptibility to *Turnip mosaic virus* under field conditions

Bernadette Rubio¹, Olivier Fernandez¹, Patrick Cosson¹, Thierry Berton¹, Mélodie Caballero¹, Fabrice Roux², Joy Bergelson³, Yves Gibon¹, Valérie Schurdi-Levraud^{1*}

¹INRA – Université de Bordeaux, UMR 1332 Biologie du Fruit et Pathologie, Villenave D'Ornon, France

²CNRS, INRA, Université de Toulouse, LIPM, Castanet-Tolosan, France.

³Ecology & Evolution, University of Chicago, 1101 E 57th St Chicago IL 60637, USA.

*** Corresponding author**

Mailing address : INRA, UMR 1332 Biologie du Fruit et Pathologie

Phone : +33 (0)5 57 12 23 77

E-mail : valerie.schurdi-levraud@inra.fr

Keywords: *Arabidopsis thaliana*, *Turnip mosaic virus*, multi-stress conditions, primary and secondary metabolism

Introduction

In a scenario of global change, plant viruses appear as one of the most alarming biotic threat for their immobile hosts. Climate change is likely to modify viral epidemics with modifying vector areas and time and spatial redistribution of viruses (Jones, 2016). With no sustainable fighting available once the disease occurs and a high potential of mutations plant viruses are pathogens able to contribute to emerging diseases (Becker *et al.*, 2016). This threat may occur in wild plants but crops with their uniformity are the first concerned (Roossinck & García-Arenal, 2015).

As obligate intracellular organisms, viruses require host cell components to complete their cycle. Studies of plant-virus interactions usually focus on finding the players of this interaction and most attention has been paid to the genetic basis of these interactions, especially in terms of innate immunity, resistance and defense (Revers & Nicaise, 2014). In contrary to bacteria and fungi (Berger *et al.*, 2007), experiments and information on metabolic perturbations, as an outcome of this interaction, are more scarce. Nevertheless, viruses deeply alter host cellular metabolism as they demand energy and metabolites to build their own blocks (Biemelt & Sonnewald, 2006). Viruses also exploit the assimilate transport system for their long-distance transport to invade cells in systemic tissues at a distance away from the initial site of infection. To facilitate their transport viral proteins interact with host factors and components of the long distance transport machinery like plasmodesmata or phloem proteins. In a compatible interaction where the process is fully completed the outcome is the symptom expressed as the stop of cell growth, stunting, chlorosis or necrosis depending on the pathosystem. In incompatible interactions, plant will deploy strategies of resistance which through downstream cascading pathways are energetically costly and compete with growth. In both cases, this stress situation enhances defense signaling involving the action of antimicrobial components and specific defense proteins (Piasecka *et al.*, 2015).

Plant primary metabolism response to viruses has often been studied in terms of respiration, photosynthetic efficiency (Kangasjärvi *et al.*, 2012) and carbon partitioning. In a susceptible interaction between tobacco and *Potato virus Y* (PVY), Herbers *et al.*, (---2000) showed the increase of soluble carbohydrates and decrease of photosynthesis 4 days after inoculation (dai). These observations were confirmed by Stare *et al.*, 2015 and Kogovšek *et al.*, 2016 who showed in the pathosystem Potato/PVY that at very early time point of the infection (1 dai) sugar decreased in leaves but an increase could be observed some days after in infected and systemic leaves. A photosynthetic shift was also observed. In leaves of susceptible plants directly inoculated with *Cucumber mosaic virus* (CMV), the alterations in carbohydrate metabolism frequently result in lower starch accumulation together with altered starch metabolism at viral infection sites (Teci *et al.*, 1996; Shalitin *et al.*, 2002) and accumulation of this storage compound in systemically infected leaves. Starch accumulation was also observed in *Arabidopsis thaliana*/*Cauliflower mosaic virus* (CaMV) interaction (Handford & Carr, 2007). Modifications in free amino acids contents were also considered. In different pathosystems, increase of total or individual amino acids was observed as well as polyamines (review in (Llave, 2016) (Pilar Lopez-Gresa *et al.*, 2012; Fernandez-Calvino *et al.*, 2014)).

Secondary metabolism also participate in such a response (Piasecka *et al.*, 2015). Numerous compounds including hormones have been found to be involved in the defense (Verma *et al.*,

2016; Palukaitis *et al.*, 2017). Among them, glucosinolates and phytoalexins greatly participate to *Arabidopsis thaliana* defense against a range of pathogens (Kliebenstein, 2004).

Whatever the interaction – compatible or incompatible – massive reprogramming of metabolism has to meet the increased demand for energy needed to sustain viral multiplication and systemic spread and defense responses against viruses (Llave, 2016).

Most of the experiments on metabolism perturbations related to virus infection have been mainly conducted in controlled condition and on a very narrow genetic basis- often one genotype. Single genotype experiments cannot grasp the complexity and the variability of metabolic response when exposed to biotic stresses.

To decipher the metabolic landscape of plant virus interaction, we chose to work on the natural *Arabidopsis thaliana*/*Turnip mosaic virus* (TuMV) pathosystem in a two years common garden experiment. In this complex situation where plants have to regulate multiple pathways related to biotic and abiotic variations, we believe the response pattern will better mimic natural conditions. We also explored the natural variability of *Arabidopsis thaliana*. Targeted and untargeted metabolic phenotyping was conducted on respectively 130 and 29 accessions, taking into account the quantitative response spanning from high susceptibility to complete resistance. Metabolic phenotyping was considered in the systemic part of the plants. This work aimed at identifying discriminant biomarkers of susceptibility/resistance to viruses in plants (Fernandez *et al.*, 2016).

Material and Method

Plant Material

Arabidopsis thaliana natural worldwide accessions from the Hapmap population (http://bergelson.uchicago.edu/?page_id=790#sets) were provided by F. Roux (LIPM, Toulouse, France). 130 accessions were used during the first experiment (2014) then a subset of 29 accessions were selected for a second experiment (2015) according to their susceptibility to TuMV on the basis of the OD value (see description of “Quantification viral titre”). To prevent from potential false negative due to a lack of inoculation, resistant accessions have been selected as they present a resistance phenotype in all our experiments conducted in the common garden. The accessions used in both experiments are listed in the [Dataset1.1](#).

Virus Material

Turnip mosaic virus isolate UK1 (Jenner & Walsh, 1996) was routinely propagated by mechanical inoculations on turnip plants *Brassica rapa* L.ssp *rapa* NA FR 490001 provided by the BraCySol germplasm center (Ploudaniel, France). To prepare the inoculum, three weeks old turnip plants were mechanically inoculated. Two weeks later symptoms appeared. Young symptomatic leaves were then collected to produce the inoculum.

Experimental Design and Growth conditions

Two common garden experiments were lead in 2014 and 2015 with 130 and 29 *A. thaliana* accessions respectively. The first common garden experiment (130 accessions) was organized in three blocks, each block being an independent randomization of one replicate per accession. The second common garden experiment was composed by four randomized blocks with four replicates per accession. Seedling trays of 40 wells were used. Seeds were sown on March 2014 and 2015 in a professional horticultural soil (106S cope – PletrAcom) in a non-heated and non-lighted glasshouse. At three weeks of age, plants were acclimatized under an

opened tunnel before their transfer to the common garden. Plants were inoculated during the acclimation step before the transfer to the common garden. Soil of the common garden had been tilled so that seedling trays could be slightly buried. Because the bottom of the wells was pierced, roots were able to reach the soil.

TuMV inoculation procedure

Inoculum was prepared by collecting young symptomatic leaves of 5 weeks old turnip. 1 g of fresh leaves was grinded in three volumes of sodium dihydrogen phosphate dodecahydrate ($\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$) 30mM and 0.2 % of diethyldithiocarbamic acid (DIECA). Inoculum was clarified through 10 min centrifugation at 13000 g. Supernatant was recovered and maintained at 4°C till the inoculation. *Arabidopsis* inoculation occurred on 4 weeks-old plants at 8-10 leaves corresponding to 1.09, 1.10 Boyes stage (Boyes *et al.*, 2001). Four young expanded leaves of each plants were mechanically inoculated: 20µL of inoculum were deposit on each leave and slightly scrubbed with carborandum. Ten minutes after inoculation, plants were rinsed with water. 2014 and 2015 assays were doubled with mock-inoculated plants treated exactly as inoculated plants except the virus. Viral concentration of the TuMV inoculum was quantified after inoculation by quantitative PCR. There is no significant differences between the viral concentration of the inocula of 2014 (0.1458 ng/µL) and 2015 (0.1345 ng/µL) common garden experiments.

Quantification of viral titre

The viral titre was estimated for each individual plant using 100 mg of *A. thaliana* powder in a semi-quantitative double antibody sandwich assay (DAS-ELISA) with a commercial anti-potyvirus monoclonal antibody kit (Agdia-Biofords – France). The reaction of the substrate (p-nitrophenyl phosphate) was followed at 405 nm. Optical densities (OD) were normalized using *A. thaliana* (Col-0 accession) positive control deposited on every ELISA plate of an assay. The values of optical density obtained correspond to the viral titre of each sample. For each accession, the final value retained correspond to an average of the OD obtained for each replicate. This value makes it possible to classify the accessions into two categories: resistant and susceptible accessions to TuMV infection. Resistant accessions do not accumulate virus and present an OD equal or under the mean of the OD of the non-infected *A. thaliana* control. Susceptible accessions have a superior OD value to this healthy sample. The category to which the accessions used in both experiments belong is listed in [Dataset1.1](#) as well as the OD values measured during the two common garden experiments. Mock-inoculated plants were controlled to be virus-free.

Harvest and Sample processing

According to previous experiments conducted in 2013, samples were collected 13 days after inoculation (dpi) at the onset of the very first symptoms. The harvest was done in the morning, at the same time for the two years of experiment. All the rosette leaves above the inoculated leaves (systemic leaves) were collected in vials from Zinsser Analytic[®]. Samples were ground and stored at -80°C until metabolite measurements. In 2014 three biological replicates were considered and in 2015, four biological replicates were prepared composed by a pooling of 4 plants maximum according to the viral charge measured by DAS ELISA. Aliquots of about 17 mg of fresh powder were weighted in 1.1-mL Micronic tubes were used in the targeted metabolite analysis. Samples were then lyophilized. Dry biomass was recorded by using an analysis balance. Aliquots of about 10 mg were weighted in 1.1-mL Micronic tubes and used in the untargeted metabolite analysis.

Targeted Metabolite analysis

Metabolites were extracted in a final volume of 650 μ L, twice with 80 % (v/v) of ethanol – HEPES/KOH 10mM (pH6) and once with 50 % (v/v) ethanol – HEPES/KOH 10mM (pH6) (Geigenberger *et al.*, 1996). Chlorophyll content was determined immediately after the extraction (Arnon, 1949). Glucose, fructose, sucrose (Stitt & Quick, 1989) malate, fumarate (Nunes-Nesi *et al.*, 2007), glutamate (Zhang *et al.*, 2015) and amino acids (Bantan-Polak *et al.*, 2001) were determined in the ethanolic supernatant. Starch (Hendriks, 2003) and protein (Bradford, 1976) contents were determined on the pellet resuspended in 100 mM NaOH. Analyses were realized in 96-well microplates using Starlet pipetting robots (Hamilton), and absorbance was read in MP96 microplate readers (SAFAS). Glutathione and ascorbate total content were also measured from 30 mg of sample powder extracted with 600 μ L of 6% trichloroacetic acid (TCA) according to Stevens *et al.*, 2006 and Griffith, 1980 respectively. **Data treatments.** Metabolite datas were normalized by dividing each measure by the value of a biological standard sample corresponding to a non-inoculated *A. thaliana* Col-0 sample harvested at the same time as the other samples. The final concentration value of each metabolite is the average of the replicates of each accession. **Statistical analysis.** All principal component analysis (PCA), parametric and nonparametric statistical tests were performed using R version 3.4.0. Statistical significance was set at $P < 0.05$. PLS and OPLS-DA were performed using the R packages mixOmics (Rohart *et al.*, 2017) and PLS (Wehrens & Mevik, 2007).

Untargeted Metabolic analysis. Untargeted metabolite measurements were conducted on the subset of 29 accessions in the second common garden experiment in 2015. Extraction was realized using Starlet pipetting robot (Hamilton) with an extraction buffer composed of 80 % (v/v) of ethanol and 0.1 % (v/v) of formic acid and methyl vanillate as an internal standard (50 μ g/mL). All the samples extracted were filtrated through a Multiscreen Solvintert 96-well filter plate (Merck Millipore). **Quality control (QC).** 25 μ L of each sample are mixed together to generate a pooled quality control sample (QC). QC samples were analyzed every 10 injections to monitor and correct changes in the instrument response. Solvent blank samples (80% ethanol in water – 0.1 % formic acid) were also analyzed in-between the other samples. **Liquid chromatography.** Liquid chromatography was performed on a Dionex UHPLC Ultimate 3000 (Thermo Scientific). Chromatographic separation was carried out in reversed-phase mode on a Gemini C18 column (2 x 150 mm; 3 μ m, Phenomenex) equipped with a Gemini C18 guard column (2 x 4 mm, Phenomenex). The mobile phase was composed of milliQ water with 0.1% formic acid (solvent A), and 100% Acetonitrile (solvent B) for a total run time of 18 min. The flow rate was 0.3 mL.min⁻¹ and the column was heated at 30°C. The autosampler temperature was maintained at 4°C and the injection volume was 5 μ L. **Mass spectrometry.** The UHPLC system was coupled with a LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific). A heated electrospray interface was used and analyses were performed in positive mode. Acquisition was performed in full scan mode with a resolution of 240 000 FWHM in the scan range of m/z 50-1000. Data were recorded using Xcalibur software (Thermo Scientific) and extracted with XCMS. Data processing and statistical analysis. Data were converted to mzXML file format. Peak picking and alignment were performed using XCMS in R (Smith *et al.*, 2006). XCMS-parameters were optimized as described by Patti and collaborators (Patti *et al.*, 2012). XCMS-data processing results in a data matrix which contains peak intensities which are a unique combination of retention-time and median m/z ratio. To exclude system-peaks (impurities in the measurement-system, visible in blanks) as well as poorly detected metabolic features, filter steps were performed based on QCs and blanks. All statistical analysis were

done in R version 3.4.0 and also using the web application BioStatFlow version 2.7.7. Thus, orthogonal partial least squares – discriminant analysis (OPLS-DA) were performed using the 505 features of the untargeted metabolic analysis combined with the 12 primary metabolites assayed by targeted methods.

Results

Strong perturbations of primary metabolism in susceptible accessions to TuMV

Experiments were performed in two different common garden experiments on 130 and on a subset of 29 “extreme” resistant and susceptible accessions (see Material and Method description). The analysis of primary metabolism focused on key 10 metabolic traits, chlorophyll a, fructose, glucose, sucrose, fumarate, malate, glutamate, starch, total amino acids and protein content. Two additional metabolites, ascorbate and glutathione, were analyzed on the 29 accessions for which the important roles towards environmental stresses were already described. In a first step, mock-inoculated samples (inoculation free of virus) were analyzed. In both years, variable graphs (Fig. S1B and S2B) show the same pattern. Correlations between dry biomass and primary metabolites showed positive and significant correlations with malate, fumarate, starch and glucose (Tables S1). Moreover, in both years dry biomass was negatively correlated with protein, amino acids and glutamate contents (Tables S1).

The 130 accessions were characterized by their quantitative response to TuMV through OD classification. Some genotypes were found to be highly susceptible whereas others were resistant (Fig. S3). Among the 130 accessions, 20 were considered fully resistant (Dataset1.1). The graphs representing the individuals (Fig. S1A and S2A) show that for mock-inoculated plants, there was no difference in terms of biomass or metabolic composition between susceptible and resistant accessions for the two years and whatever the subset. In TuMV-infected plants, the pattern of correlation was modified (Fig. 1A and B). A significant decrease in dry biomass was observed for both years on TuMV-inoculated samples compared to mock-inoculated samples (Fig. S4). Moreover, the positive correlation between dry biomass and metabolites was lost (Tables S2). Opposite to mock plants, viral infection conditions clearly distinguish two groups of accessions according to their susceptibility to TuMV among the 130 accessions analyzed in 2014 (Fig. 1A). This result is even clearer and confirmed in the 2015 common garden experiment performed on the subset of 29 accessions (Fig. 1B). The infection provoked a significant accumulation of the 10 in 2014 and of the 12 in 2015 primary metabolism compounds in susceptible accessions compared to resistant ones. Even the contents of the two additional metabolites analyzed in 2015, ascorbate and glutathione are higher in susceptible samples compared to resistant genotypes (Fig. S5 and S6). These observations were confirmed by the positive and significant correlations found between the viral titre and the amounts of primary metabolic traits (Table S2). Concerning the dry biomass there was a negative correlation with the viral titre, which was particularly significant in the 2015 common garden experiment (Table S2). In contrast, dry biomass was maintained in resistant accessions (Fig. S.7), indicating that they were able to defend themselves at low metabolic costs. Strikingly and at the contrary of what has been reported in many studies, there was no decrease in chlorophyll a content in susceptible infected accessions. All these results, confirmed in the second year experiment, suggest that the viral infection provoked the accumulation of major primary metabolites, probably as a result of an imbalance between photosynthesis and growth.

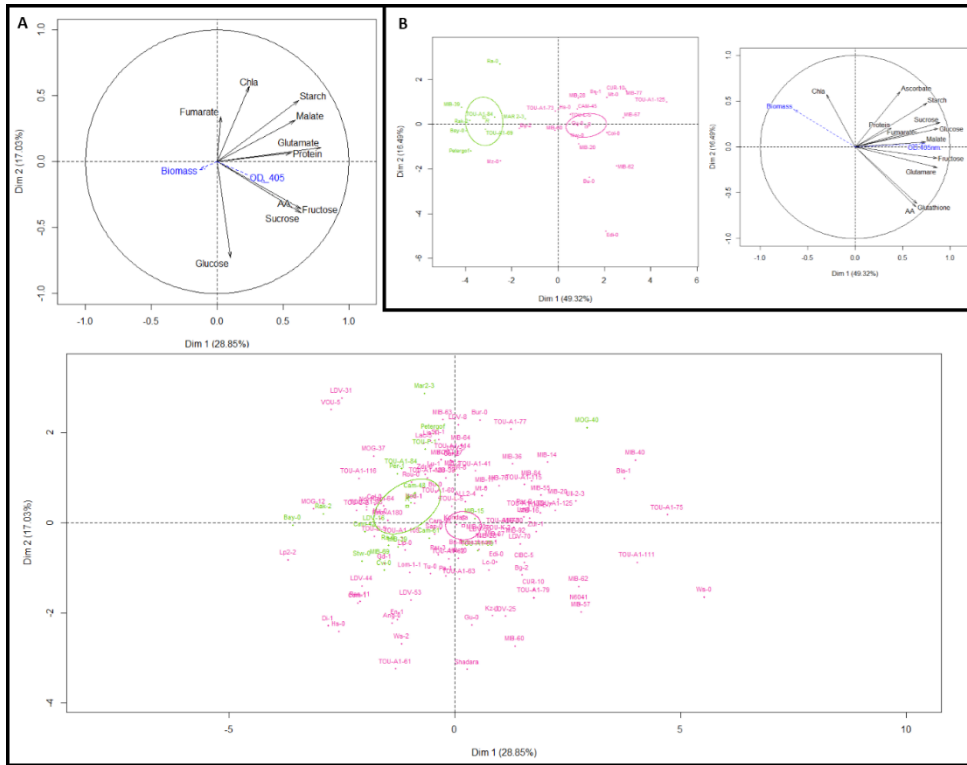


Figure 1. Principal component (PC) analysis of the 130 and 29 *A. thaliana* accessions inoculated by the TuMV in the 2014 and 2015 common garden experiments

A. PCA performed with 10 primary metabolic traits measured at 13 days after TuMV inoculation on the 130 *A. thaliana* accessions in 2014. The two major PC that accounted for 45.88% of the variance have been plotted. The dry biomass (Biomass) and the viral accumulation (OD_405nm) were considered as explanatory traits (in blue on the variable factor map). **B.** PCA performed with 12 primary metabolic traits measured at 13 days after TuMV inoculation on the 29 *A. thaliana* accessions in 2015. The two major PC that accounted for 65.81% of the variance have been plotted. The dry biomass (Biomass) and the viral accumulation (OD_405nm) were considered as explanatory traits (in blue on the variable factor map).

The metabolic content discriminates *A. thaliana* genotypes according to their susceptibility to the TuMV

To further analyze the pattern of infection on a larger set of metabolites and to decipher the secondary metabolism response, untargeted metabolic analysis was conducted. Phenotyping through UHPLC-LTQ Orbitrap was realized on the set of 29 accessions showing a contrasted response to TuMV. After data analysis, using the R package xcms, this metabolic phenotyping revealed a total of 505 workable metabolic signatures (m/z). As already observed with primary metabolism, there was no difference between susceptible and resistant accessions when they were mock-inoculated (Fig. S8A). In contrast, when the accessions were inoculated, the principal component analysis performed with the metabolic variables clearly separated two groups according to the susceptibility to the TuMV (Fig. S8B).

An Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) was then performed to maximize the variation between the two groups of accessions and determine the most significant variables contributing to this variation. OPLS-DA analysis was carried out using the primary and secondary metabolite data. The quality of the model was validated by the Q^2 parameter (goodness-of-prediction parameter) with a value of 0.846 showing high predictive capabilities (Fig. S9). As already observed with the analysis of the primary metabolic variables, susceptible accessions accumulated significantly a large number of secondary compounds in comparison with resistant accessions (Fig. S9A). To examine which metabolic compounds/signatures strongly contribute to the OPLS-DA model, they were ranked

according to their “Variable Importance in the Projection (VIP)” values (see [Dataset1.2](#) for the complete list of VIPs). This confirmed the major role of sucrose, glucose, fructose, glutathione, glutamate, amino acids and fumarate in the response of susceptible accessions. Among the other VIP values, some seem to accumulate in large proportions in susceptible or resistant accessions which leads to high fold changes between them. Nevertheless, among these VIP values, only 15% (21/140 VIPs) accumulate significantly more in resistant accessions, which implies that a large majority were accumulated in susceptible accessions ([Dataset1.2](#)).

A complementary analysis was realized without the *a priori* groups of accessions according to the susceptibility to TuMV. To test whether viral accumulation can be predicted from metabolic data, a Partial Least Square (PLS) analysis was performed on the optical density values corresponding to the viral titre measured by DAS ELISA for each accession. The PLS coefficient estimated in the training data set, shows after cross-validation a correlation between predicted and true viral accumulation of 0.61 confirming the high predictive power of metabolic composition for TuMV susceptibility ([Fig. S10](#)). Among the best 50 VIP-PLS values, 44 were found together with the VIP values obtained with the OPLS-DA analysis. Only 6 of them concern metabolites mostly accumulated among the resistant accessions. In both analyses, sucrose corresponded to one of the best VIP values. Sucrose concentrations were significantly higher in susceptible accessions which suggests that it is not involved in defense mechanisms, contrary to what is widely accepted in the topic of plant/pathogen interactions (Rojas *et al.*, 2014) (see [Dataset1.3](#) for the complete list of the common VIPs between OPLS-DA and PLS analysis). Finally, OPLS-DA analysis was performed between the mock- and TuMV-inoculated samples giving also VIP values which were compared to the VIP values obtained with the OPLS-DA analysis already described on resistant and susceptible accessions. There are 63 common metabolic signatures between the two analyses most of which correspond to more accumulated variables in TuMV-inoculated samples (58/63). This shows well the strong metabolic disturbances under viral infection. In addition, most of these variables correspond to metabolic traits found in higher concentrations in susceptible accessions ([See Dataset 1.4](#)).

Discussion

In this study we characterized the metabolic response of *A. thaliana* to its natural viral pathogen, the *Turnip mosaic virus* (TuMV) (Pagan *et al.*, 2010). Natural diversity of *A. thaliana* was explored in order to answer the question whether metabolism products could be biomarkers of the susceptibility versus resistance. This study was carried out on a large set of *A. thaliana* accessions which present a quantitative response to the virus from high susceptibility to resistance. Moreover, in order to be as close as possible to reality, this study was conducted for two consecutive years in common garden conditions, in which plants are submitted to daily and complex environmental changes. It has indeed been shown that the responses of plants submitted to multiple stresses are sometimes very different to those observed under single stress conditions (Rasmussen *et al.*, 2013). In common garden experiments, quality and intensity of light, temperature, humidity and many other environmental factors can vary considerably more than in greenhouses or growth cabinets, suggesting that plants grown under controlled conditions probably express only a fraction of their capacity for metabolic responses. This provides strong arguments for studying plants in conditions close to the reality of the field (Jänkänpää *et al.*, 2012; Annunziata *et al.*, 2017). Primary and secondary metabolites analysis were performed on a large set of resistant and susceptible *A. thaliana* accessions.

Over the two years of study, it was not possible to distinguish TuMV-susceptible from TuMV-resistant accessions based on the metabolic traits measured in mock-inoculated plants. It suggests that there are no constitutive metabolic patterns associated with resistance or susceptibility.

On the other hand, these mock-inoculated samples, made it possible to describe elements of primary metabolism in *A. thaliana* under common garden conditions. Dry biomass was significantly correlated with fumarate concentrations. Fumarate can accumulate to high levels in *A. thaliana* in comparison to other plant species suggesting that it likely constitutes a significant fraction of the fixed carbon in *A. thaliana* rosette leaves (Araújo *et al.*, 2011). The amount of carbon stored in fumarate can be similar to that accumulated in starch. Relatively fast-growing plant species such as *A. thaliana* contain significantly higher concentrations of organic acids, but not of soluble sugars, compared to slow-growing plant species (Chia *et al.*, 2000; Araújo *et al.*, 2011) especially under high light intensity conditions, which is the case in these common garden experiments (Chia *et al.*, 2000). There was also a positive significant correlation between dry biomass and glucose. Glucose but also fumarate reached levels corresponding to 22.65 mM and 18.88 mM respectively in fast growing accessions, suggesting that they might be involved in turgor dynamics and thus growth by cellular expansion (Pritchard, 2001). On the contrary, the negative correlation found between dry biomass and protein could be linked with the fact that less protein synthesis contributes to increase the efficiency of use of carbon, as protein synthesis is a costly process. This negative relationship is in agreement with what had been observed in a collection of *A. thaliana* accessions grown under controlled conditions (Sulpice *et al.*, 2013).

Upon viral infection, the susceptible accessions were clearly distinguished from the resistant ones by the accumulation of primary metabolites observed over the two years of study. This is in agreement with several studies that have shown the massive reprogramming of the plants that takes place in response to pathogens (Rojas *et al.*, 2012). In a large study on alteration of primary metabolism to *Tobacco rattle virus* (TRV) in *Arabidopsis*, Fernandez-Calvino and collaborators (Fernandez-Calvino *et al.*, 2014) showed on Col-0 accession and in controlled conditions that sucrose was significantly accumulated at 8 days post inoculation. Amino acids were also globally accumulated in infected plants compared to mock plants whereas neither starch nor fumarate were found over accumulated. Differences in the virus species, the number of accessions observed and conditions of experiments could explain these different results. In a controlled condition multistress experiment including TuMV, drought and heat, Prasad and Sonnewald (2013) also observed accumulation of soluble sugars in multistressed Col-0 plants. The high glutathione content which had an antioxidant role minimizing oxidative stress illustrates the important oxidative stress of susceptible accessions. This reflects the strong metabolic disturbances generated in susceptible plants by TuMV infection. Indeed, resistant accessions seem to be able to ensure their defense but also to ensure the maintenance of their growth which results in a significantly higher dry biomass.

Metabolites are often classified as being either primary or secondary. Primary metabolism includes essential metabolites whereas the secondary metabolism is often connected to interactions with environmental cues (Fernie & Klee, 2011). Despite this seemingly very clear distinction, there is a strong link between central and secondary metabolism (Fritz *et al.*, 2006). Thus, in view of the strong perturbations observed on the primary metabolic content, it seemed interesting to analyze the impact of the viral infection on secondary metabolism

using a non-targeted approach. Strong perturbations were also observed in susceptible accessions. In contrast, in the resistant accessions, a small group of molecules were found at significantly higher concentrations compared to the susceptible ones. Furthermore, in susceptible accessions, the synthesis of numerous secondary molecules might be expensive in energy and thus compete with growth or simply have toxic effects, thus resulting in decrease biomass. On the contrary, in two complementary studies working on metabolic response to *Tomato yellow leaf curl virus* (TYLCV) of resistant and susceptible lines of tomato, a stronger response in resistant lines was showed mostly linked to secondary metabolism (Sade *et al.*, 2015).

Our study conducted on a high number of accessions and in common garden conditions clearly highlights new results on *Arabidopsis*/virus interaction. We have shown that the primary and secondary metabolic contents allow the discrimination of accessions according to their susceptibility. Among these metabolites, seven primary metabolites highly contribute to predict the response in terms of viral titre. Among secondary metabolites, quantitative and qualitative differences between susceptible and resistant group could be observed. Some compounds which presented very high fold change in resistant accessions compared to susceptible ones are of great interest. The next step in this work will be to characterize these secondary compounds and their biosynthetic pathways and then to test their involvement in resistance by using pharmacological and/or genetic approaches. This would allow for better understanding the complexity of the underlying mechanisms involved in plant's response to viruses.

Supporting Information

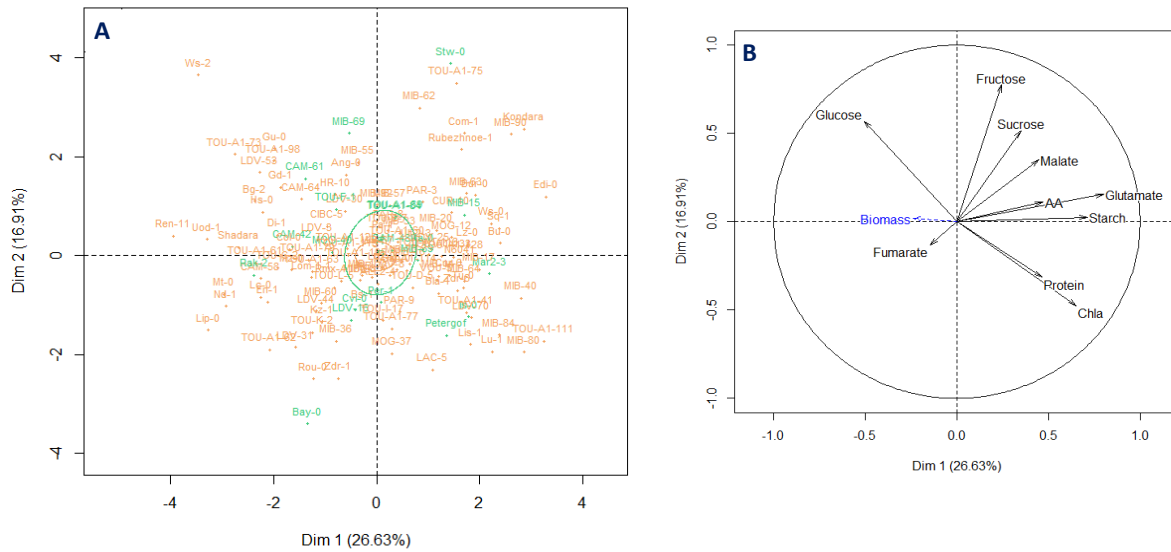


Fig. S 1. Principal component (PC) analysis of the 130 mock-inoculated accessions of the 2014 common garden experiment

PCA was performed with the 10 primary metabolic traits measured at 13 days after inoculation on the 130 *A. thaliana* accessions of the 2014 common garden experiment. The two major PC that accounted for 43.51% of the variance have been plotted. **A.** Score plot in which resistant (in green) and susceptible (in orange) accessions were represented. The confidence ellipses around the barycentre of individuals are also represented. **B.** Variable factor map of the 10 primary metabolic traits analyzed; AA = Amino Acids, Chla = Chlorophylls a. The dry biomass (Biomass) is considered as an explanatory trait (in blue).

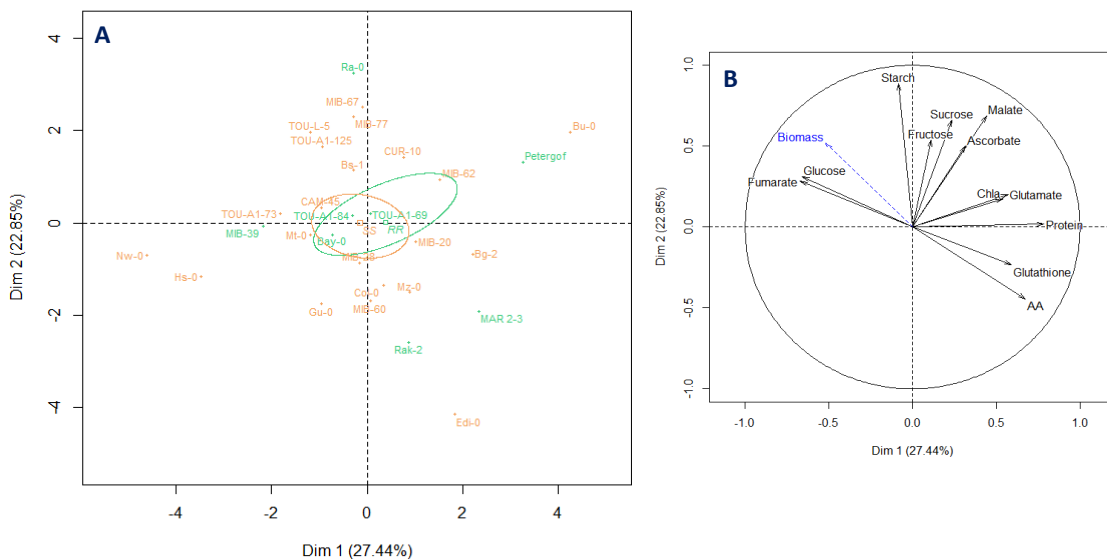


Fig. S 2. Principal component (PC) analysis of the 29 mock-inoculated accessions of the 2015 common garden experiment

PCA was performed with the 12 primary metabolic traits measured at 13 days after inoculation on the 29 *A. thaliana* accessions of the 2015 common garden experiment. The two major PC that accounted for 50.29% of the variance have been plotted. **A.** Score plot in which resistant (in green) and susceptible (in orange) accessions are represented. The confidence ellipses around the barycentre of individual are also represented. **B.** Variable factor map of the 12 primary metabolites analyzed – AA = Amino Acids; Chla = Chlorophyll a. The dry biomass (Biomass) was taking into account as an explanatory trait (in blue).

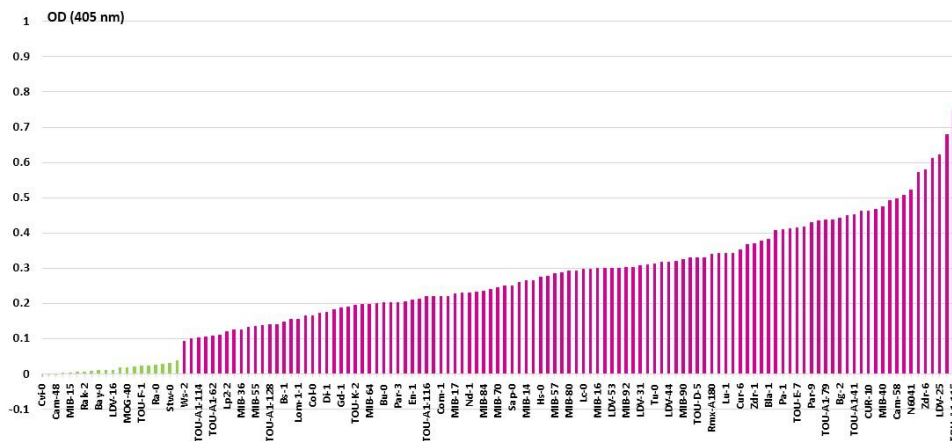


Fig. S 3 : Distribution of the 130 *A. thaliana* accessions of the 2014 common garden experiment according to their OD value

Each accession is represented by a bar corresponding to its OD value measured by DAS ELISA. The resistant accessions are represented by the green bars. These accessions do not accumulate virus. They have an OD value equal or under the mean OD of the non-infected *A. thaliana* control. The susceptible accessions are represented by the pink bars. These accessions have an OD value higher than OD value of the healthy control *A. thaliana* sample.

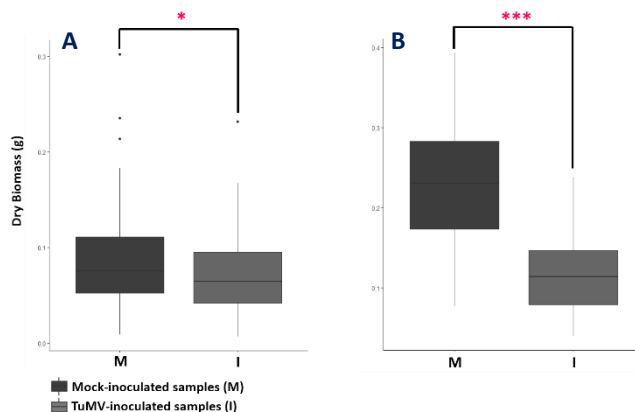


Fig. S 4 Dry biomass of TuMV- and mock-inoculated samples in 2014 (A) and 2015 (B) common garden experiments
 Statistical comparisons on dry biomass were performed between TuMV and mock-inoculated samples on both common garden experiments. Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.' 1

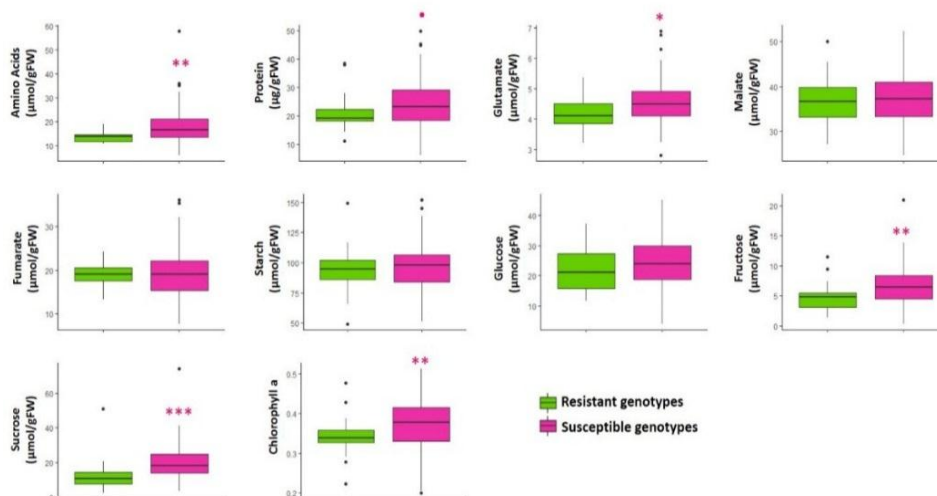


Fig. S 5. Comparisons of primary metabolic content in TuMV-inoculated resistant and susceptible 130 *A. thaliana* accessions of 2014 common garden experiment

Resistant accessions were represented by green boxplot and susceptible accessions by pink boxplots. Statistical analyzes were performed on each of the 10 primary metabolites (Wicoxon-Mann Whitney or Student test according to the normality of the data). Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.' 1

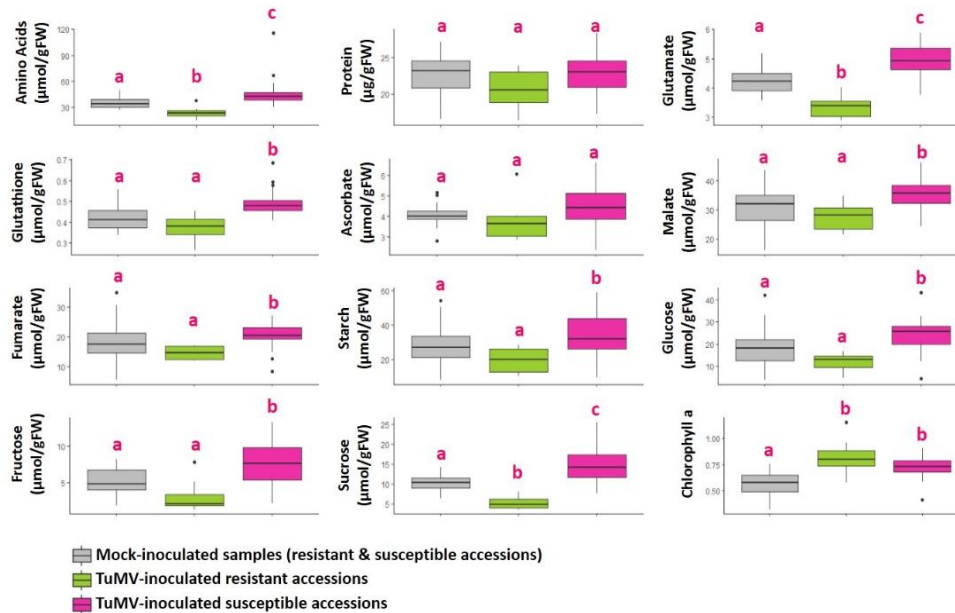


Fig. S 6. Comparisons of primary metabolic content in Mock-inoculated and TuMV-inoculated resistant and susceptible 29 *A. thaliana* accessions of 2015 common garden experiment

Resistant accessions were represented by green boxplot and susceptible accessions by pink boxplots. Statistical analyzes were performed on each of the 10 primary metabolites (Wicoxon-Mann Whitney or Student test according to the normality of the data). Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1

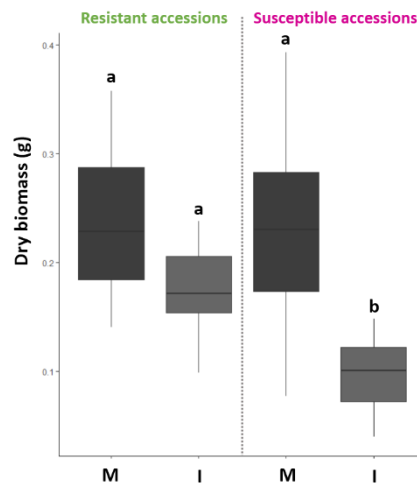


Fig. S 7. Dry biomass on TuMV- and mock-inoculated samples of resistant and susceptible accessions in the 2015 common garden experiment

Statistical comparisons (Kruskal-Wallis test) on dry biomass were performed between the different categories of samples. Letters refer to groups with statistically significant differences between them.

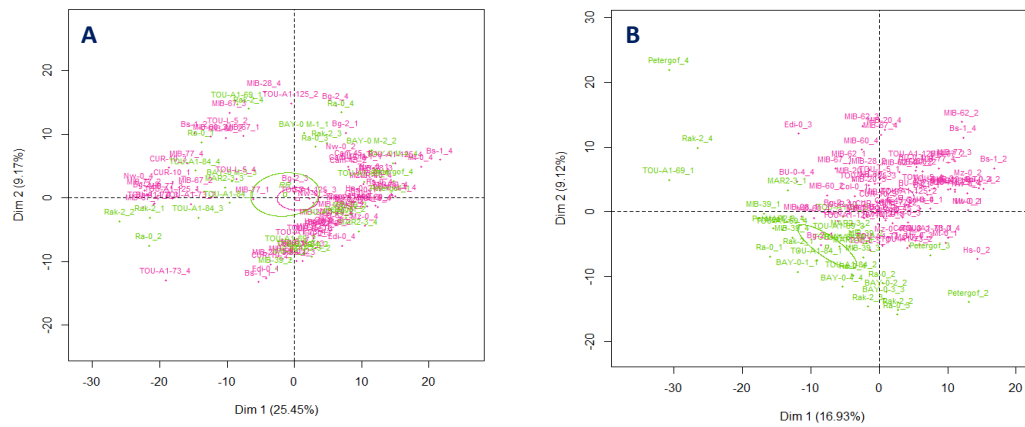


Fig. S 8. Principal component analysis performed on the 505 metabolic signatures (m/z) measured on the 29 *A. thaliana* accessions of the 2015 common garden experiment

A. PCA carried out using the 505 metabolic signatures (m/z) measured by UHPLC-LTQ Orbitrap on the 29 *A. thaliana* **mock-inoculated** accessions. The two major PC that accounted for 34.62% of the variance. **B.** PCA carried out using the 505 metabolic signatures (m/z) measured by UHPLC-LTQ Orbitrap on the 29 *A. thaliana* **TuMV-inoculated** accessions. The two major PC accounted for 26.05% of the variance.

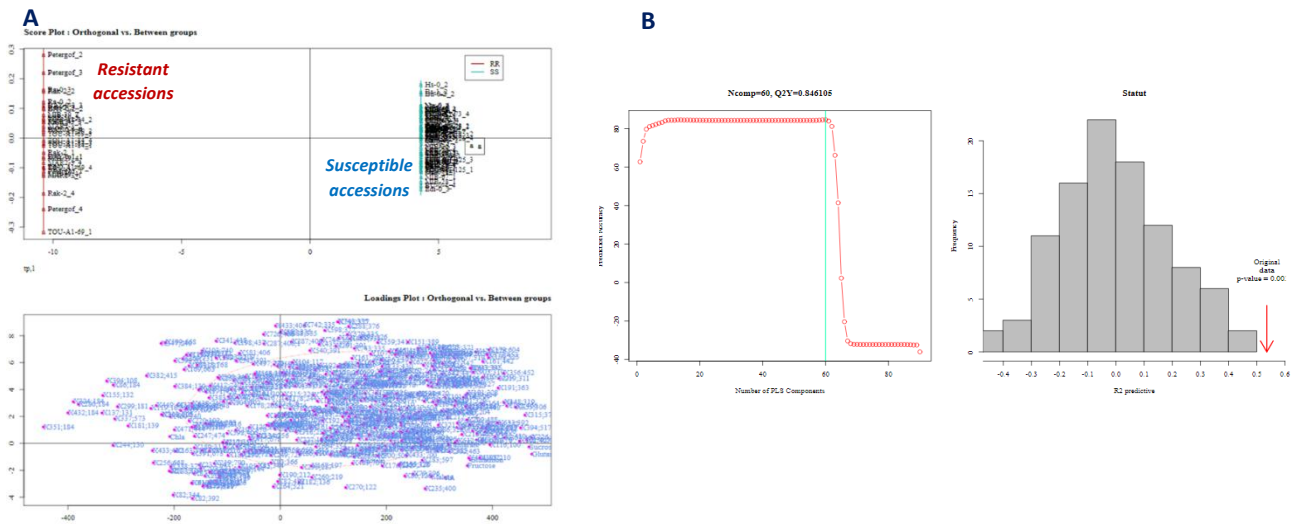


Fig. S 9. OPLS-DA analysis and its parameters of validation for the TuMV-inoculated resistant and susceptible 29 accessions of the 2015 common garden experiment

A. OPLS-DA results with the score plot and the loading plot containing all the metabolic variables tested. The resistant (in red) and susceptible (in blue) accessions were represented. **B.** Parameters for validation of the OPLS-DA analysis.

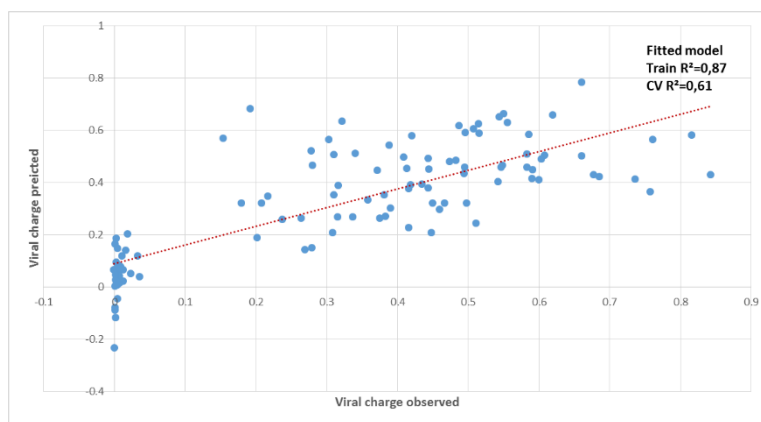


Fig. S 10. Prediction of the viral accumulation by the metabolite matrix
The metabolite matrix is composed by 12 primary metabolic traits and 505 metabolic signatures (m/z). The 29 *A. thaliana* accessions replicates are represented by the blue points. The straight red line represents the exact prediction.

Table S 1 . Spearman correlations between dry biomass and the primary metabolites measured on the 130 (A) and 29 (B) *A. thaliana* accessions of the 2014 and 2015 common garden experiments

A											
	AA	Protein	Glutamate	Malate	Fumarate	Starch	Glucose	Fructose	Sucrose	Chla	
Biomass	-0.02	-0.27 **	-0.24 **	0.23 **	0.42 ***	0.05	0.17 -	0	-0.21 *	-0.22 *	

B												
	AA	Protein	Glutamate	Glutathione	Ascorbate	Malate	Fumarate	Starch	Glucose	Fructose	Sucrose	Chla
Biomass	-0.66 ***	-0.42 *	-0.45 *	-0.31 .	-0.12	0.38 *	0.36 .	0.48 **	0.65 ***	0.18	0.3	-0.12

All the significant correlations were showed by asterisks and values (Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.' 1)

Table S 2. Spearman correlations between dry biomass and viral charge and the primary metabolites content in the two common garden experiments

A											
	AA	Protein	Glutamate	Malate	Fumarate	Starch	Glucose	Fructose	Sucrose	Chla	
Biomass	-0.11	-0.04	-0.14	-0.22 *	0.14	0.08	-0.06	0.15 .	-0.12	-0.16 .	-0.17 *
OD		0.36 ***	0.20 *	0.29 ***	0.05	0	0.23 **	0.21 *	0.32 ***	0.42 ***	0.14

B													
	AA	Protein	Glutamate	Glutathione	Ascorbate	Malate	Fumarate	Starch	Glucose	Fructose	Sucrose	Chla	
Biomass	-0.62 ***	-0.75 ***	-0.27	-0.71 ***	-0.62 ***	0.03	-0.44 *	-0.22	-0.3	-0.38 *	-0.61 ***	-0.52 ***	0.23
OD		0.49 **	0.35 .	0.63 ***	0.47 *	0.3	0.22	0.44 *	0.58 **	0.55 ***	0.42 *	0.70 ***	-0.08

All the significant correlations were showed by asterisks and values (Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.' 1)

Other Supporting Information Files

[Dataset S1 \(XLSX\)](#)

Dataset 1.1 : List of the *A. thaliana* accessions with their geographic position and their susceptible status to TuMV infection confirmed by OD values in both experiments

Genotype	ID	Latitude	Longitude	Country	Suceptibility Groups	2014 - OD	2015 - OD
Bay-0	6899	49	11	GER	R	0.009	0.004
Cam-42	45	48.2667	-4.58333	FRA	R	0.023	-
Cam-48	51	48.2667	-4.58333	FRA	R	-0.001	-
Cam-61	66	48.2667	-4.58333	FRA	R	0.004	-
Cvi-0	6911	15.1111	-23.6167	CPV	R	-0.009	-
Is-0	8312	50.5	7.5	GER	R	0.016	-
LDV-16	106	48.5167	-4.06667	FRA	R	0.011	-
Mar2-3	159	47.35	3.93333	FRA	R	0.002	0.005
MIB-15	166	47.3833	5.31667	FRA	R	0.002	-
MIB-39	190	47.3833	5.31667	FRA	R	0.011	0.007
MIB-69	212	47.3833	5.31667	FRA	R	0.020	-
MOG-40	244	48.6667	-4.06667	FRA	R	0.017	-
Per-1	8354	58	56.3167	RUS	R	0.037	-
Petergof	7296	59	29	RUS	R	0.006	0.011
Ra-0	6958	46	3.3	FRA	R	0.024	0.003
Rak-2	8365	49	16	CZE	R	0.004	0.006
Stw-0	8388	52	36	RUS	R	0.029	-
TOU-A1-69	335	46.6667	4.11667	FRA	R	-0.002	0.015
TOU-A1-84	348	46.6667	4.11667	FRA	R	0.028	0.007
TOU-F-1	369	46.6667	4.11667	FRA	R	0.023	-
ALL2-4	12	45.2667	1.48333	FRA	S	0.462	-
Ang-0	6992	50.3	5.3	BEL	S	0.316	-
Bg-2	6709	47.6479	-122.305	USA	S	0.441	0.324
Bla-1	7015	41.6833	2.8	ESP	S	0.381	-
Bs-1	8270	47.5	7.5	SUI	S	0.147	0.66
Bu-0	8271	50.5	9.5	GER	S	0.201	0.4
Bur-0	5719	54.1	-6.2	IRL	S	0.239	-
Cam-58	62	48.2667	-4.58333	FRA	S	0.496	-
Cam-64	69	48.2667	-4.58333	FRA	S	0.407	-
CIBC-5	6730	51.4083	-0.6383	UK	S	0.287	-
Col-0	6909	38.3	-92.3	USA	S	0.165	0.391
Com-1	7092	49.416	2.823	FRA	S	0.220	-
CUR-10	79	45	1.75	FRA	S	0.463	0.576
Cur-6	84	45	1.75	FRA	S	0.352	-
Di-1	7098	47	5	FRA	S	0.175	-
Edi-0	6914	56	-3	UK	S	0.438	0.258
En-1	8290	50	8.5	GER	S	0.210	-
Gd-1	8296	53.5	10.5	GER	S	0.186	-
Gu-0	6922	50.3	8	GER	S	0.298	0.389
HR-10	6923	51.4083	-0.6383	UK	S	0.343	-
Hs-0	8310	52.24	9.44	GER	S	0.274	0.46
Kondara	6929	38.48	68.49	TJK	S	0.110	-
Kz-1	6930	49.5	73.1	KAZ	S	0.377	-
Lac-5	96	47.7	6.81667	FRA	S	0.368	-
Lc-0	8323	57	-4	UK	S	0.296	-

LDV-25	116	48.5167	-4.06667	FRA	S	0.622	-
LDV-3	121	48.5167	-4.06667	FRA	S	0.140	-
LDV-30	122	48.5167	-4.06667	FRA	S	0.492	-
LDV-31	123	48.5167	-4.06667	FRA	S	0.306	-
LDV-44	137	48.5167	-4.06667	FRA	S	0.318	-
LDV-53	146	48.5167	-4.06667	FRA	S	0.300	-
LDV-70	156	48.5167	-4.06667	FRA	S	0.165	-
LDV-8	157	48.5167	-4.06667	FRA	S	0.308	-
Lip-0	8325	50	19.3	POL	S	0.330	-
Lis-1	8326	56	14.7	SWE	S	0.679	-
Lom-1-1	6042	56.09	13.9	SWE	S	0.155	-
Lp2-2	7520	49.38	16.81	CZE	S	0.119	-
Lu-1	8334	55.71	13.2	SWE	S	0.341	-
Lz-0	6936	46	3.3	FRA	S	0.341	-
MIB-1	160	47.3833	5.31667	FRA	S	0.138	-
MIB-14	165	47.3833	5.31667	FRA	S	0.265	-
MIB-16	167	47.3833	5.31667	FRA	S	0.298	-
MIB-17	168	47.3833	5.31667	FRA	S	0.227	-
MIB-20	171	47.3833	5.31667	FRA	S	0.202	0.525
MIB-28	178	47.3833	5.31667	FRA	S	0.211	0.428
MIB-33	184	47.3833	5.31667	FRA	S	0.197	-
MIB-36	187	47.3833	5.31667	FRA	S	0.125	-
MIB-40	191	47.3833	5.31667	FRA	S	0.474	-
MIB-55	201	47.3833	5.31667	FRA	S	0.134	-
MIB-57	202	47.3833	5.31667	FRA	S	0.284	-
MIB-60	204	47.3833	5.31667	FRA	S	0.896	0.425
MIB-62	206	47.3833	5.31667	FRA	S	0.507	0.346
MIB-63	207	47.3833	5.31667	FRA	S	0.182	-
MIB-64	208	47.3833	5.31667	FRA	S	0.198	-
MIB-67	210	47.3833	5.31667	FRA	S	0.434	0.549
MIB-70	213	47.3833	5.31667	FRA	S	0.244	-
MIB-76	216	47.3833	5.31667	FRA	S	0.411	-
MIB-80	219	47.3833	5.31667	FRA	S	0.292	-
MIB-84	223	47.3833	5.31667	FRA	S	0.234	-
MIB-90	229	47.3833	5.31667	FRA	S	0.325	-
MIB-92	230	47.3833	5.31667	FRA	S	0.302	-
MOG-12	237	48.6667	-4.06667	FRA	S	0.105	-
MOG-37	242	48.6667	-4.06667	FRA	S	0.124	-
Mt-0	6939	32.34	22.46	LIB	S	0.450	0.467
Mz-0	6940	50.3	8.3	GER	S	0.190	0.59
N6041	8420	50.0667	8.5333	GER	S	0.521	-
Nd-1	6942	50	10	SUI	S	0.229	-
Pa-1	8353	38.07	13.22	ITA	S	0.409	-
Par-3	258	46.65	-0.25	FRA	S	0.202	-
PAR-8	262	46.65	-0.25	FRA	S	0.198	-
Par-9	263	46.65	-0.25	FRA	S	0.429	-
Ren-11	6960	48.5	-1.41	FRA	S	0.172	-
Rmx-A180	7525	42.036	-86.511	USA	S	0.339	-
Rou-0	7320	49.4424	1.09849	FRA	S	0.611	-

Rubezhnoe-1	7323	49	38.28	UKR	S	0.300	-
Sap-0	8378	49.49	14.24	CZE	S	0.250	-
Shahdara	6962	38.35	68.48	TJK	S	0.467	-
Sq-1	6966	51.4083	-0.6383	UK	S	0.220	-
TOU-A1-105	273	46.6667	4.11667	FRA	S	0.233	-
TOU-A1-111	277	46.6667	4.11667	FRA	S	0.276	-
TOU-A1-114	280	46.6667	4.11667	FRA	S	0.101	-
TOU-A1-115	281	46.6667	4.11667	FRA	S	0.750	-
TOU-A1-116	282	46.6667	4.11667	FRA	S	0.219	-
TOU-A1-125	291	46.6667	4.11667	FRA	S	0.418	0.403
TOU-A1-128	292	46.6667	4.11667	FRA	S	0.140	-
TOU-A1-133	296	46.6667	4.11667	FRA	S	0.228	-
TOU-A1-41	320	46.6667	4.11667	FRA	S	0.451	-
TOU-A1-60	326	46.6667	4.11667	FRA	S	0.131	-
TOU-A1-61	327	46.6667	4.11667	FRA	S	0.219	-
TOU-A1-62	328	46.6667	4.11667	FRA	S	0.108	-
TOU-A1-63	329	46.6667	4.11667	FRA	S	0.249	-
TOU-A1-73	338	46.6667	4.11667	FRA	S	0.205	0.55
TOU-A1-75	340	46.6667	4.11667	FRA	S	0.299	-
TOU-A1-77	341	46.6667	4.11667	FRA	S	0.328	-
TOU-A1-79	343	46.6667	4.11667	FRA	S	0.436	-
TOU-A1-98	359	46.6667	4.11667	FRA	S	0.101	-
TOU-C-2	361	46.6667	4.11667	FRA	S	0.319	-
TOU-D-5	364	46.6667	4.11667	FRA	S	0.329	-
TOU-E-7	368	46.6667	4.11667	FRA	S	0.414	-
TOU-I-17	378	46.6667	4.11667	FRA	S	0.302	-
TOU-K-2	385	46.6667	4.11667	FRA	S	0.194	-
TOU-L-5	389	46.6667	4.11667	FRA	S	0.570	0.407
Tu-0	8395	45	7.5	ITA	S	0.312	-
Ull-2-3	6973	56.0648	13.9707	SWE	S	0.292	-
Uod-1	6975	48.3	14.45	AUT	S	0.260	-
VOU-5	394	46.65	0.166667	FRA	S	0.154	-
Ws-0	6980	52.3	30	RUS	S	0.265	-
Ws-2	6981	52.3	30	RUS	S	0.092	-
Zdr-1	6984	49.3853	16.2544	CZE	S	0.369	-
Zdr-6	6985	49.3853	16.2544	CZE	S	0.578	-

Dataset1.2: List of the VIPs identify by OPLS-DA analysis performed on TuMV-inoculated resistant and susceptible accessions. For each VIP, comparisons between resistant and susceptible accessions were done. The fold change was also calculated for each VIP.

VIP OPLS-DA	mz	rt	VIP values	R vs. S metabolic contents	Fold change
Sucrose	NA	NA	2.213	S > R ***	X 2.89
m351r184	351.006	184.222	2.047	R > S ***	X 2.88
m303r463	303.133	463.270	1.973	S > R ***	X 3.69
m356r452	356.120	452.352	1.958	S > R ***	X 2.17
m219r311	219.101	311.134	1.940	S > R ***	X 4.81
m533r392	533.155	391.765	1.926	S > R ***	X 4.98
m332r430	332.132	430.481	1.898	S > R ***	X 3.78
m116r100	116.070	100.163	1.867	S > R ***	X 6.05
m205r242	205.097	242.121	1.837	S > R ***	X 2.51
m432r184	431.971	183.776	1.831	R > S ***	X 3.44
m103r242	103.041	242.089	1.799	S > R ***	X 2.86
m302r506	302.102	506.314	1.796	S > R ***	X 5.76
m255r544	255.112	543.799	1.769	S > R ***	X 9.96
m221r216	221.092	215.744	1.748	S > R ***	X 3.13
m385r210	385.106	210.083	1.738	S > R ***	X 45.77
Glucose	NA	NA	1.734	S > R ***	X 2.49
m175r402	175.148	401.831	1.713	S > R ***	X 3.77
m903r319	903.277	318.889	1.710	S > R ***	X 8.48
m503r391	503.190	391.457	1.695	S > R ***	X 2.51
m343r345	343.117	345.319	1.693	S > R ***	X 2.08
Fructose	NA	NA	1.683	S > R ***	X 2.49
m474r107	474.218	107.2975	1.675	S > R ***	X 2.68
m315r448	315.133	447.821	1.674	S > R ***	X 2.39
m209r488	209.153	488.350	1.669	S > R ***	X 2.70
m209r414	209.153	413.756	1.663	S > R ***	X 2.14
Glutathione	NA	NA	1.660	S > R ***	X 1.32
m212r284	211.559	283.891	1.657	S > R ***	X 1.56
m543r99	543.132	98.522	1.645	S > R ***	X 5.84
m124r346	124.075	346.354	1.638	S > R ***	X 1.92
m137r131	136.931	131.31	1.618	R > S ***	X 1.39
m370r302	370.149	301.519	1.610	S > R ***	X 28.49
m203r204	203.084	203.918	1.591	S > R ***	X 8.24
Glutamate	NA	NA	1.590	S > R ***	X 1.47
m430r333	430.170	332.878	1.581	S > R ***	X 3.30
m331r329	331.117	328.817	1.578	S > R ***	X 1.58
m512r491	512.127	491.367	1.563	R > S ***	X 2.54
m226r406	226.107	406.369	1.554	S > R ***	X 7.97
m151r359	151.075	359.377	1.549	S > R ***	X 1.78
m394r108	394.200	108.419	1.548	R > S ***	X 1.93
m449r319	449.106	318.949	1.545	S > R ***	X 1.75
m757r319	757.217	319.130	1.537	S > R ***	X 1.69
m270r587	270.133	587.091	1.534	S > R ***	X 43.68
m315r370	315.133	370.476	1.524	S > R ***	X 3.93
m162r402	162.055	402.281	1.511	S > R ***	X 1.90

m337r573	337.292	572.917	1.510	R > S ***	X 1.64
m221r230	221.121	230.244	1.503	S > R ***	X 1.74
m394r517	394.204	517.072	1.490	S > R ***	X 17.28
m302r407	302.101	406.697	1.486	S > R ***	X 7.16
m482r101	482.107	101.274	1.475	S > R ***	X 1.79
Amino Acids	NA	NA	1.462	S > R ***	X 1.88
m355r306	355.102	306.288	1.451	S > R ***	X 3.29
m191r416	191.143	416.467	1.448	S > R ***	X 2.49
m193r324	193.125	323.817	1.445	S > R ***	X 1.46
m149r360	149.096	360.180	1.437	S > R ***	X 1.66
m642r445	642.254	445.078	1.435	S > R ***	X 2.00
m321r760	321.114	760.442	1.419	S > R ***	X 3.70
m182r466	182.081	465.796	1.415	S > R ***	X 4.21
m305r210	305.086	210.166	1.408	S > R ***	X 2.24
m348r319	348.274	318.945	1.396	S > R ***	X 3.61
m299r181	299.098	181.363	1.396	R > S ***	X 22.3
m164r496	164.070	496.307	1.375	S > R ***	X 2.22
m165r424	165.127	424.140	1.372	S > R ***	X 1.54
m317r464	317.101	463.650	1.370	S > R ***	X 1.25
m191r363	191.070	362.643	1.360	S > R ***	X 2.76
m219r486	219.101	486.477	1.340	S > R ***	X 2.70
m105r700	105.069	700.287	1.330	S > R ***	X 1.59
m367r358	367.101	358.491	1.323	S > R ***	X 1.69
m201r344	201.054	343.832	1.319	S > R ***	X 1.71
m189r325	189.127	324.754	1.312	S > R ***	X 1.74
m146r242	146.060	241.843	1.311	S > R ***	X 2.31
m133r296	133.064	295.680	1.307	S > R ***	X 2.65
m179r604	179.106	603.961	1.306	S > R ***	X 1.64
m374r322	374.144	321.703	1.305	S > R ***	X 1.77
m181r139	181.053	138.934	1.293	R > S ***	X 1.94
m373r453	373.127	452.600	1.290	S > R ***	X 1.81
m302r382	302.041	381.707	1.282	S > R ***	X 1.54
m161r442	161.096	441.797	1.274	S > R ***	X 1.79
m164r375	164.070	375.012	1.269	S > R ***	X 2.95
m109r359	109.064	359.340	1.267	S > R ***	X 1.92
m324r184	323.989	183.797	1.263	R > S ***	X 106.8
m386r278	386.220	277.549	1.256	S > R ***	X 3.48
m391r774	391.245	774.477	1.255	S > R ***	X 2.29
m409r491	409.169	490.545	1.251	S > R ***	X 27.12
m291r259	291.181	259.063	1.241	S > R ***	X 4.78
m280r184	280.084	183.768	1.229	R > S ***	X 5.90
m192r462	192.041	462.370	1.227	S > R ***	X 1.27
m107r370	107.085	370.001	1.219	S > R ***	X 1.39
m361r491	361.092	490.564	1.218	R > S ***	X 2.07
m420r276	419.695	275.797	1.217	S > R ***	X 1.77
m193r373	193.086	373.266	1.189	S > R ***	X 2.04
m379r402	379.095	402.311	1.179	S > R ***	X 2.88
m181r464	181.086	463.916	1.176	S > R ***	X 1.42
m181r328	181.086	327.894	1.170	S > R ***	X 1.68

m80r395	80.049	395.289	1.168	S > R ***	X 1.57
m373r344	373.127	343.531	1.160	S > R ***	X 1.54
m627r299	627.155	299.369	1.156	S > R ***	X 2.80
m96r389	96.080	389.004	1.155	S > R ***	X 1.74
m195r453	195.065	452.614	1.154	S > R ***	X 1.55
m79r396	79.041	395.754	1.152	S > R ***	X 1.52
m105r327	105.069	326.898	1.150	S > R ***	X 1.72
m210r465	210.112	464.799	1.147	S > R ***	X 2.44
m335r506	335.127	506.441	1.143	S > R ***	X 3.71
m155r132	154.941	131.541	1.141	R > S ***	X 1.33
m86r126	86.096	126.246	1.114	S > R ***	X 2.09
m433r482	433.112	481.588	1.114	R > S ***	X 244
m433r700	433.241	700.261	1.109	R > S ***	X 1.82
m396r257	396.185	256.510	1.106	S > R ***	X 1.43
m611r354	611.158	353.603	1.102	S > R ***	X 1.34
m178r191	178.089	191.013	1.102	S > R ***	X 3.43
m169r496	169.049	495.500	1.101	S > R ***	X 3.67
m162r249	162.055	249.101	1.098	S > R ***	X 1.50
m396r348	396.115	347.739	1.092	S > R ***	X 4.14
m64r395	63.934	395.047	1.091	R > S ***	X 3.60
m256r668	256.080	667.874	1.089	R > S ***	X 2.12
m103r389	103.054	389.444	1.086	S > R ***	X 2
m521r325	521.201	325.328	1.084	S > R ***	X 2.52
m460r740	460.269	740.005	1.076	R > S **	X 1.42
m133r126	133.105	126.347	1.072	S > R **	X 2.32
m209r426	209.153	425.939	1.071	S > R ***	X 1.98
m162r357	162.055	356.879	1.069	S > R ***	X 1.74
m449r442	449.107	442.311	1.069	R > S ***	X 6.17
m201r451	201.054	451.123	1.063	S > R ***	X 1.75
m527r460	527.103	460.095	1.056	S > R ***	X 2.02
Fumarate	NA	NA	1.054	S > R ***	X 1.38
m393r416	393.188	415.816	1.052	S > R ***	X 1.71
m212r774	212.094	774.047	1.052	S > R ***	X 2.76
m244r130	243.942	130.167	1.049	R > S ***	X 1.35
m402r420	402.162	419.597	1.048	S > R **	X 1.87
m347r240	347.159	239.659	1.039	R > S **	X 1.48
m404r388	404.227	388.155	1.038	S > R **	X 1.31
m464r367	464.248	366.989	1.035	S > R ***	X 1.97
m225r371	225.148	370.798	1.034	S > R ***	X 1.57
m103r284	103.054	283.823	1.033	S > R ***	X 3.26
m227r789	227.163	788.506	1.023	S > R ***	X 1.61
m222r216	221.602	215.822	1.016	S > R **	X 1.33
m86r184	86.059	184.278	1.015	R > S ***	X 1.87
m367r342	367.153	341.908	1.011	S > R ***	X 1.96
m302r463	302.102	463.078	1.006	S > R ***	X 1.49
m309r325	309.116	324.659	1.003	S > R **	X 1.44
m244r325	244.096	324.677	1.001	S > R **	X 1.41

Dataset1.3 : Comparisons between the VIPs identified by OPLS-DA analysis and those identified by PLS analysis performed with TuMV-inoculated resistant and susceptible accessions. The fold changes were represented for each VIP

VIP PLS	mz	rt	VIP-PLS value	VIP-OPLS-DA value	R vs. S met. contents	Fold change
m356r452	356.120	452.352	2.104	1.958	S > R ***	X 2.17
Glutamate	NA	NA	2.085	1.590	S > R ***	X 1.47
Sucrose	NA	NA	1.994	2.213	S > R ***	X 2.89
m219r311	219.101	311.134	1.975	1.940	S > R ***	X 4.81
m385r210	385.106	210.083	1.952	1.738	S > R ***	X 45.77
m373r453	373.127	452.600	1.906	1.290	S > R ***	X 1.81
m270r587	270.133	587.091	1.853	1.534	S > R ***	X 43.68
m315r370	315.133	370.476	1.828	1.524	S > R ***	X 3.93
m348r319	348.274	318.945	1.819	1.396	S > R ***	X 3.61
m394r517	394.204	517.072	1.810	1.490	S > R ***	X 17.28
m191r363	191.070	362.643	1.796	1.360	S > R ***	X 2.76
m394r108	394.200	108.419	1.794	1.548	R > S ***	X 1.93
m474r107	474.218	107.298	1.789	1.675	S > R ***	X 2.68
m226r406	226.107	406.369	1.771	1.554	S > R ***	X 7.97
m351r184	351.006	184.222	1.748	2.047	R > S ***	X 2.88
m324r184	323.989	183.797	1.738	1.263	R > S ***	X 106.8
m420r276	419.695	275.797	1.729	1.217	S > R ***	X 1.77
m367r342	367.153	341.908	1.721	1.011	S > R ***	X 1.96
m280r184	280.084	183.768	1.706	1.229	R > S ***	X 5.90
m302r407	302.101	406.697	1.692	1.486	S > R ***	X 7.16
m503r391	503.190	391.457	1.688	1.695	S > R ***	X 2.51
m116r100	116.070	100.163	1.679	1.867	S > R ***	X 6.05
m396r257	396.185	256.510	1.673	1.106	S > R ***	X 1.43
m303r463	303.133	463.270	1.658	1.973	S > R ***	X 3.69
m432r184	431.971	183.776	1.649	1.831	R > S ***	X 3.44
m355r306	355.102	306.288	1.641	1.451	S > R ***	X 3.29
m370r302	370.149	301.519	1.634	1.610	S > R ***	X 28.49
m86r184	86.059	184.278	1.611	1.015	R > S ***	X 1.87
m343r345	343.117	345.319	1.610	1.693	S > R ***	X 2.08
m533r392	533.155	391.765	1.600	1.926	S > R ***	X 4.98
m201r451	201.054	451.123	1.599	1.063	S > R ***	X 1.75
m757r319	757.217	319.130	1.594	1.537	S > R ***	X 1.69
m386r278	386.220	277.549	1.590	1.256	S > R ***	X 3.48
m903r319	903.277	318.889	1.587	1.710	S > R ***	X 8.48
m449r319	449.106	318.949	1.582	1.545	S > R ***	X 1.75
m182r466	182.081	465.796	1.565	1.415	S > R ***	X 4.21
m175r402	175.148	401.831	1.541	1.713	S > R ***	X 3.77
m305r210	305.086	210.166	1.535	1.408	S > R ***	X 2.24
m195r453	195.065	452.614	1.535	1.154	S > R ***	X 1.55
m221r216	221.092	215.744	1.535	1.748	S > R ***	X 3.13
m374r322	374.144	321.703	1.526	1.305	S > R ***	X 1.77
Glucose	NA	NA	1.521	1.734	S > R ***	X 2.49
m179r604	179.106	603.961	1.521	1.306	S > R ***	X 1.64
m161r442	161.096	441.797	1.518	1.274	S > R ***	X 1.79

Dataset1.4 : Common VIPs obtained with the OPLS-DA performed on TuMV-inoculated resistant vs. susceptible accessions AND the one performed on mock- vs. TuMV-inoculated samples. For each VIP, the comparison between the metabolic content in mock and TuMV-inoculated samples was done and the fold change was represented. The comparison, for each VIP, of metabolic content between resistant and susceptible accessions was also represented

VIP-OPLS-DA	mz	rt	VIP values	Mock vs. TuMV ₁	Fold change ₁	R vs. S
m64r395	63.934	395.047	1.669	I > M	X 1.25	R > S ***
m512r491	512.127	491.367	1.813	I > M	X 21.08	R > S ***
m433r700	433.241	700.261	2.254	I > M	X 3	R > S ***
m361r491	361.092	490.564	1.914	I > M	X 4.84	R > S ***
m222r216	221.602	215.822	1.351	I > M	X 2.46	S > R **
m133r126	133.105	126.347	1.067	I > M	X 2.62	S > R **
m543r99	543.132	98.522	1.122	I > M	X 2.62	S > R ***
m503r391	503.190	391.457	1.479	I > M	X 2.24	S > R ***
m482r101	482.107	101.274	2.029	I > M	X 3.24	S > R ***
m474r107	474.218	107.298	2.100	I > M	X 7.12	S > R ***
m409r491	409.169	490.545	1.067	I > M	X 1.46	S > R ***
m396r348	396.115	347.739	1.126	I > M	X 7.35	S > R ***
m394r517	394.204	517.072	1.126	I > M	X 4.36	S > R ***
m391r774	391.245	774.477	1.397	I > M	X 2.36	S > R ***
m386r278	386.220	277.549	1.385	I > M	X 5.51	S > R ***
m379r402	379.095	402.311	1.091	I > M	X 2.47	S > R ***
m374r322	374.144	321.703	1.445	I > M	X 1.88	S > R ***
m373r344	373.127	343.531	1.457	I > M	X 1.78	S > R ***
m370r302	370.149	301.519	1.331	I > M	X 5.59	S > R ***
m367r358	367.101	358.491	1.251	I > M	X 1.62	S > R ***
m367r342	367.153	341.908	1.200	I > M	X 2.40	S > R ***
m356r452	356.120	452.352	1.358	I > M	X 1.69	S > R ***
m355r306	355.102	306.288	1.595	I > M	X 8.11	S > R ***
m348r319	348.274	318.945	1.376	I > M	X 4.20	S > R ***
m343r345	343.117	345.319	1.793	I > M	X 2.41	S > R ***
m332r430	332.132	430.481	1.416	I > M	X 2.32	S > R ***
m331r329	331.117	328.817	2.177	I > M	X 2.55	S > R ***
m315r448	315.133	447.821	1.412	I > M	X 2.05	S > R ***
m315r370	315.133	370.476	1.331	I > M	X 3.58	S > R ***
m305r210	305.086	210.166	1.576	I > M	X 2.69	S > R ***
m303r463	303.133	463.270	1.683	I > M	X 2.90	S > R ***
m302r407	302.101	406.697	1.222	I > M	X 4.30	S > R ***
m291r259	291.181	259.063	1.165	I > M	X 3.68	S > R ***
m227r789	227.163	788.506	1.024	I > M	X 1.71	S > R ***
m221r230	221.121	230.244	1.412	I > M	X 1.67	S > R ***
m221r216	221.092	215.744	1.494	I > M	X 2.46	S > R ***
m219r311	219.101	311.134	1.641	I > M	X 3.44	S > R ***
m212r774	212.094	774.047	1.113	I > M	X 3.14	S > R ***
m212r284	211.559	283.891	1.812	I > M	X 1.97	S > R ***
m210r465	210.112	464.799	1.097	I > M	X 2.27	S > R ***
m209r488	209.153	488.350	1.895	I > M	X 3.87	S > R ***
m205r242	205.097	242.121	1.955	I > M	X 3.16	S > R ***
m201r344	201.054	343.832	1.355	I > M	X 1.71	S > R ***

m193r324	193.125	323.817	2.201	I > M	X 4.54	S > R ***
m192r462	192.041	462.370	1.000	I > M	X 1.31	S > R ***
m191r363	191.070	362.643	1.219	I > M	X 2.37	S > R ***
m182r466	182.081	465.796	1.244	I > M	X 3.65	S > R ***
m178r191	178.089	191.013	1.128	I > M	X 3.11	S > R ***
m175r402	175.148	401.831	1.423	I > M	X 2.52	S > R ***
m164r375	164.070	375.012	1.107	I > M	X 2.38	S > R ***
m162r402	162.055	402.281	1.032	I > M	X 1.41	S > R ***
m162r357	162.055	356.879	1.767	I > M	X 3.10	S > R ***
m162r249	162.055	249.101	1.432	I > M	X 1.81	S > R ***
m151r359	151.075	359.377	1.455	I > M	X 1.28	S > R ***
m133r296	133.064	295.680	1.574	I > M	X 3.79	S > R ***
m124r346	124.075	346.354	1.493	I > M	X 1.68	S > R ***
m116r100	116.070	100.163	1.779	I > M	X 6.88	S > R ***
m109r359	109.064	359.340	1.036	I > M	X 1.67	S > R ***
m449r442	449.107	442.311	1.102	M > I	X 4.26	R > S ***
m137r131	136.931	131.310	1.036	M > I	X 1.35	R > S ***
m79r396	79.041	395.754	2.194	M > I	X 1.85	S > R ***
m169r496	169.049	495.500	1.288	M > I	X 1.91	S > R ***
m105r700	105.069	700.287	1.816	M > I	X 1.88	S > R ***

References

- Annunziata MG, Apelt F, Carillo P, Krause U, Feil R, Mengin V, Lauxmann MA, Köhl K, Nikoloski Z, Stitt M, et al. 2017.** Getting back to nature: a reality check for experiments in controlled environments. *Journal of Experimental Botany* **68**: 4463–4477.
- Araújo WL, Nunes-Nesi A, Fernie AR. 2011.** Fumarate: Multiple functions of a simple metabolite. *Phytochemistry* **72**: 838–843.
- Arnon DI. 1949.** Copper enzymes in isolated chloroplasts. Polyphenoloxidase in *Beta vulgaris*. *Plant Physiology* **24**: 1-15.
- Bantan-Polak T, Kassai M, Grant KB. 2001.** A Comparison of Fluorescamine and Naphthalene-2,3-dicarboxaldehyde Fluorogenic Reagents for Microplate-Based Detection of Amino Acids. *Analytical Biochemistry* **297**: 128–136.
- Becker MM, Smith EC, Denison MR. 2016.** Emerging Viral Infections. In: Weaver, SC and Denison, M and Roossinck, M and Vignuzzi, M, ed. *Virus Evolution: Current Research and Future Directions*. 329–365.
- Berger S, Sinha AK, Roitsch T. 2007.** Plant physiology meets phytopathology: plant primary metabolism and plant-pathogen interactions. *Journal of Experimental Botany* **58**: 4019–4026.
- Biemelt S, Sonnewald U. 2006.** Plant-microbe interactions to probe regulation of plant carbon metabolism. *Journal of Plant Physiology* **163**: 307–318.
- Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Gorchach J. 2001.** Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* **13**: 1499–1510.
- Bradford MM. 1976.** A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**: 248–254.
- Chia DW, Yoder TJ, Reiter W-D, Gibson SI. 2000.** Fumaric acid: an overlooked form of fixed carbon in *Arabidopsis* and other plant species. *Planta* **211**: 743–751.
- Fernandez O, Urrutia M, Bernillon S, Giauffret C, Tardieu F, Le Gouis J, Langlade N, Charcosset A, Moing A, Gibon Y. 2016.** Fortune telling: metabolic markers of plant performance. *Metabolomics* **12**: 158-171.
- Fernandez-Calvino L, Osorio S, Hernandez ML, Hamada IB, Del Toro FJ, Donaire L, Yu A, Bustos R, Fernie AR, Martinez-Rivas JM, et al. 2014.** Virus-induced alterations in primary metabolism modulate susceptibility to *Tobacco rattle virus* in *Arabidopsis*. *Plant Physiology* **4**: 1821-1838.
- Fernie AR, Klee HJ. 2011.** The use of natural genetic diversity in the understanding of metabolic organization and regulation. *Frontiers in Plant Science* **2**: 59.
- Fritz C, Palacios-Rojas N, Feil R, Stitt M. 2006.** Regulation of secondary metabolism by the carbon-nitrogen status in tobacco: nitrate inhibits large sectors of phenylpropanoid metabolism. *The Plant Journal* **46**: 533–548.

Geigenberger P, Lerchi J, Stitt M, Sonnewald U. 1996. Phloem-specific expression of pyrophosphatase inhibits long distance transport of carbohydrates and amino acids in tobacco plants. *Plant, Cell & Environment* **19**: 43–55.

Griffith OW. 1980. Determination of glutathione and glutathione disulfide using glutathione reductase and 2-vinylpyridine. *Analytical Biochemistry* **106**: 207–212.

Handford MG, Carr JP. 2007. A defect in carbohydrate metabolism ameliorates symptom severity in virus-infected *Arabidopsis thaliana*. *Journal of General Virology* **88**: 337–341.

Hendriks JHM. 2003. ADP-glucose pyrophosphorylase is activated by posttranslational redox-modification in response to light and to sugars in leaves of *Arabidopsis* and other plant species. *Plant Physiology* **133**: 838–849.

Herbers K, Takahata Y, Melzer M, Mock HP, Hajirezaei M, Sonnewald U. 2000. Regulation of carbohydrate partitioning during the interaction of *Potato virus Y* with tobacco. *Molecular Plant Pathology* **1**: 51–59.

Jänkänpää HJ, Mishra Y, Schröder WP, Jansson S. 2012. Metabolic profiling reveals metabolic shifts in *Arabidopsis* plants grown under different light conditions: Metabolic profiling under different light regime. *Plant, Cell & Environment* **35**: 1824–1836.

Jenner CE, Walsh JA. 1996. Pathotypic variation in *Turnip mosaic virus* with special reference to European isolates. *Plant Pathology* **45**: 848–856.

Jones RAC. 2016. Future scenarios for plant virus pathogens as climate change progresses. In: Kielian, M and Maramorosch, K and Mettenleiter, TC, ed. *Advances in Virus Research*. *Advances in Virus Research* **95**: 87–147.

Kangasjärvi S, Neukermans J, Li S, Aro E-M, Noctor G. 2012. Photosynthesis, photorespiration, and light signalling in defence responses. *Journal of Experimental Botany* **63**: 1619–1636.

Kliebenstein DJ. 2004. Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. *Plant, Cell & Environment* **27**: 675–684.

Kogovšek P, Pompe-Novak M, Petek M, Fragner L, Weckwerth W, Gruden K. 2016. Primary metabolism, phenylpropanoids and antioxidant pathways are regulated in potato as a response to *Potato virus Y* infection (K Wu, Ed.). *PLoS One* **11**: e0146135.

Llave C. 2016. Dynamic cross-talk between host primary metabolism and viruses during infections in plants. *Current Opinion in Virology* **19**: 50–55.

Nunes-Nesi A, Carrari F, Gibon Y, Sulpice R, Lytovchenko A, Fisahn J, Graham J, Ratcliffe RG, Sweetlove LJ, Fernie AR. 2007. Deficiency of mitochondrial fumarate hydratase activity in tomato plants impairs photosynthesis via an effect on stomatal function: Inhibition of fumarate hydratase impairs photosynthesis. *The Plant Journal* **50**: 1093–1106.

Pagan I, Fraile A, Fernandez-Fueyo E, Montes N, Alonso-Blanco C, Garcia-Arenal F. 2010. *Arabidopsis thaliana* as a model for the study of plant-virus co-evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1983–1995.

- Palukaitis P, Yoon J-Y, Choi S-K, Carr JP. 2017.** Manipulation of induced resistance to viruses. *Current Opinion in Virology* **26**: 141–148.
- Patti GJ, Tautenhahn R, Siuzdak G. 2012.** Meta-analysis of untargeted metabolomic data from multiple profiling experiments. *Nature Protocols* **7**: 508–516.
- Piasecka A, Jedrzejczak-Rey N, Bednarek P. 2015.** Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytologist* **206**: 948–964.
- Pilar Lopez-Gresa M, Lison P, Kim HK, Choi YH, Verpoorte R, Rodrigo I, Conejero V, Maria Belles J. 2012.** Metabolic fingerprinting of *Tomato Mosaic Virus* infected *Solanum lycopersicum*. *Journal of Plant Physiology* **169**: 1586–1596.
- Prasch CM, Sonnewald U. 2013.** Simultaneous application of heat, drought, and virus to *Arabidopsis* plants reveals significant shifts in signaling networks. *Plant Physiology* **162**: 1849–1866.
- Pritchard J. 2001.** Turgor pressure. *eLS*.
- Rasmussen S, Barah P, Suarez-Rodriguez MC, Bressendorff S, Friis P, Costantino P, Bones AM, Nielsen HB, Mundy J. 2013.** Transcriptome responses to combinations of stresses in *Arabidopsis*. *Plant Physiology* **161**: 1783–1794.
- Revers F, Nicaise V. 2014.** Plant Resistance to Infection by Viruses. *eLS John Wiley & Sons, Ltd: Chichester*.
- Rohart F, Gautier B, Singh A, Le Cao K-A. 2017.** MixOmics: an R package for 9omics feature selection and multiple data integration. *bioRxiv*: 108597.
- Rojas CM, Senthil-Kumar M, Tzin V, Mysore KS. 2014.** Regulation of primary plant metabolism during plant-pathogen interactions and its contribution to plant defense. *Frontiers in Plant Science* **5**: 17-28.
- Rojas CM, Senthil-Kumar M, Wang K, Ryu CM, Kaundal A, Mysore KS. 2012.** Glycolate oxidase modulates reactive oxygen species-mediated signal transduction during nonhost resistance in *Nicotiana benthamiana* and *Arabidopsis*. *Plant Cell* **24**: 336–352.
- Roossinck MJ, García-Arenal F. 2015.** Ecosystem simplification, biodiversity loss and plant virus emergence. *Current Opinion in Virology* **10**: 56–62.
- Sade D, Shriki O, Cuadros-Inostroza A, Tohge T, Semel Y, Haviv Y, Willmitzer L, Fernie AR, Czosnek H, Brotman Y. 2015.** Comparative metabolomics and transcriptomics of plant response to *Tomato yellow leaf curl virus* infection in resistant and susceptible tomato cultivars. *Metabolomics* **11**: 81–97.
- Shalitin D, Wang Y, Omid A, Gal-On A, Wolf S. 2002.** *Cucumber mosaic virus* movement protein affects sugar metabolism and transport in tobacco and melon plants. *Plant Cell and Environment* **25**: 989–997.
- Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. 2006.** XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry* **78**: 779–787.

- Stare T, Ramšak Ž, Blejec A, Stare K, Turnšek N, Weckwerth W, Wienkoop S, Vodnik D, Gruden K. 2015.** Bimodal dynamics of primary metabolism-related responses in tolerant potato-*Potato virus Y* interaction. *BMC Genomics* **16**: 716-732.
- Stevens R, Buret M, Garchery C, Carretero Y, Causse M. 2006.** Technique for rapid, small-scale analysis of vitamin C levels in fruit and application to a tomato mutant collection. *Journal of Agricultural and Food Chemistry* **54**: 6159–6165.
- Stitt M, Quick P. 1989.** Photosynthetic carbon partitioning: its regulation and possibilities for manipulation. *Physiologia Plantarum* **77**: 633–641.
- Sulpice R, Nikoloski Z, Tschoep H, Antonio C, Kleessen S, Larhlimi A, Selbig J, Ishihara H, Gibon Y, Fernie AR, et al. 2013.** Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of *Arabidopsis* accessions. *Plant Physiology* **162**: 347–363.
- Tecsi LI, Smith AM, Maule AJ, Leegood RC. 1996.** A spatial analysis of physiological changes associated with infection of cotyledons of marrow plants with *Cucumber mosaic Virus*. *Plant Physiology* **111**: 975–985.
- Verma V, Ravindran P, Kumar PP. 2016.** Plant hormone-mediated regulation of stress responses. *BMC Plant Biology* **16**: 86-95.
- Wehrens R, Mevik B-H. 2007.** The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software* **18**: 2.
- Zhang N, Gibon Y, Wallace JG, Lepak N, Li P, Dedow L, Chen C, So Y-S, Kremling K, Bradbury PJ, et al. 2015.** Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *Plant Physiology* **168**: 575–583.

B-2. Cinétique de la réponse d'*Arabidopsis thaliana* au TuMV en conditions extérieures : évolution des variables maladies et métaboliques pour des génotypes contrastés

Introduction

Les analyses liées au métabolisme primaire ont été effectuées dans le but d'élargir le phénotypage de la réponse d'*A. thaliana* au TuMV et sa caractérisation en utilisant des variables supplémentaires. Il s'agissait également de pouvoir mettre en évidence de potentiels marqueurs métaboliques de résistance/sensibilité. La richesse de ces analyses est liée aux conditions extérieures, dans lesquelles la réponse est caractérisée et au grand nombre d'accessions d'*A. thaliana* mises en jeu. Des analyses métaboliques ciblées et non ciblées ont ainsi été menées sur deux années différentes en conditions extérieures. L'ensemble des réponses métaboliques même si très riche en information est caractéristique uniquement d'un moment en particulier de l'infection puisque les échantillons ont été prélevés à une seule date après inoculation (13 dpi). Or, dans les quelques études menées sur la réponse métabolique des plantes face à des stress biotiques et en particulier aux virus, l'aspect très dynamique de cette réponse est souligné (Fernandez-Calvino *et al.*, 2014; Sade *et al.*, 2015; Stare *et al.*, 2015; Kogovšek *et al.*, 2016). Afin de compléter les résultats obtenus à un temps précis, un essai complémentaire a été mené dans les mêmes conditions extérieures que les essais précédents. Un nombre plus limité d'accessions d'*A. thaliana* a été retenu afin de pouvoir évaluer le contenu en 10 métabolites primaires à 5, 7, 9 et 13 jours après infection. La concentration en métabolites a également été quantifiée avant inoculation (T0) afin de mettre en évidence de potentielles différences entre accessions avant l'inoculation. Le dosage des 10 variables métaboliques (acides aminés, protéines, glutamate, malate, fumarate, glucose, fructose, sucrose, amidon, chlorophylle a) a été réalisé sur la plateforme métabolique de Bordeaux.

Cet essai a donc été mené sur 11 accessions différentes d'*A. thaliana* choisies de par leur faible (Bay-0, Ra-0, TOU-A1-84, Rak-2, Petergof) ou leur forte sensibilité (Bu-0, Bs-1, Mt-0, Col-0, Cur-10, Shahdara) au TuMV. Ces accessions font partie de la liste des 29 accessions analysées dans les précédentes études (Rubio *et al.*, 2017 – *in preparation*). Deux témoins de l'infection ont été placés dans l'expérimentation, le génotype C24 (sensible au TuMV) et un mutant EMS (fond génétique C24 – résistant au TuMV : mutant *elf(iso)4G*). La sensibilité de ces accessions au TuMV a été vérifiée par la quantification de la charge virale à 5-7-9 et 13 dpi par ELISA semi-quantitatif. Les informations de biomasse fraîche pour chaque accession à chacune des dates d'analyse ont également été utilisées.

1 « Temps zéro » : empreinte métabolique des accessions avant inoculation

Le contenu en métabolites primaires a été analysé pour les accessions avant l'inoculation. Il s'agit de plantes de 4 semaines au stade 8-10 feuilles (Boyes *et al.*, 2001). Il s'agit de caractériser l'empreinte métabolique de ces accessions au « temps zéro ». L'analyse en composantes principales réalisée à partir des 10 variables métaboliques montre que le génotype Bay-0 semble se détacher des autres accessions qui sont plus regroupées entre elles (Fig. B2.1). Cette distinction semblerait liée aux concentrations en glucose, fructose, sucrose et acides aminés (graphique des variables en haut à droite Fig. B2.1). Des analyses statistiques (ANOVA ou Kruskal-Wallis en fonction du respect de la normalité des variables analysées) permettant de comparer les concentrations métaboliques entre les différentes accessions n'ont pas mis en évidence de différences statistiquement significatives malgré le fait que l'accession Bay-0 présente de fortes concentrations notamment en glucose, fructose, sucrose et acides aminés (Fig. B2.2).

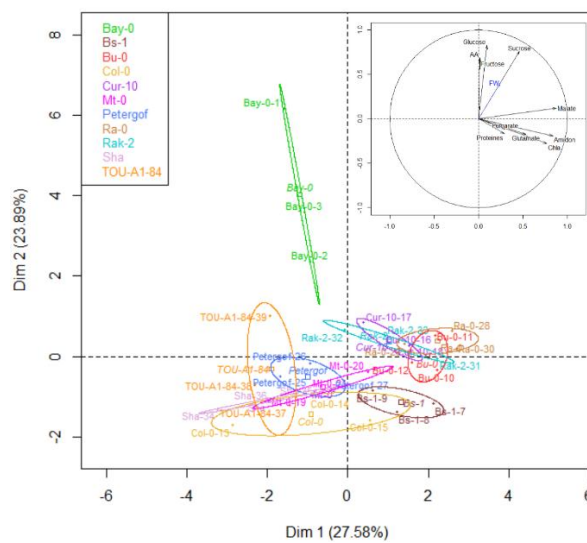


Figure B2.1 | Analyse en composante principale à partir des concentrations en 10 métabolites primaires
ACP réalisée à partir des concentrations en 10 métabolites primaires des échantillons des 11 accessions prélevés avant l'inoculation. Chaque génotype présente 3 réplicats (chaque réplicat est composé d'un pool de trois plantes). Le graphique des variables est représenté en haut à droite du graphique des individus. Les deux premières composantes représentent 51.47 % de la variance.

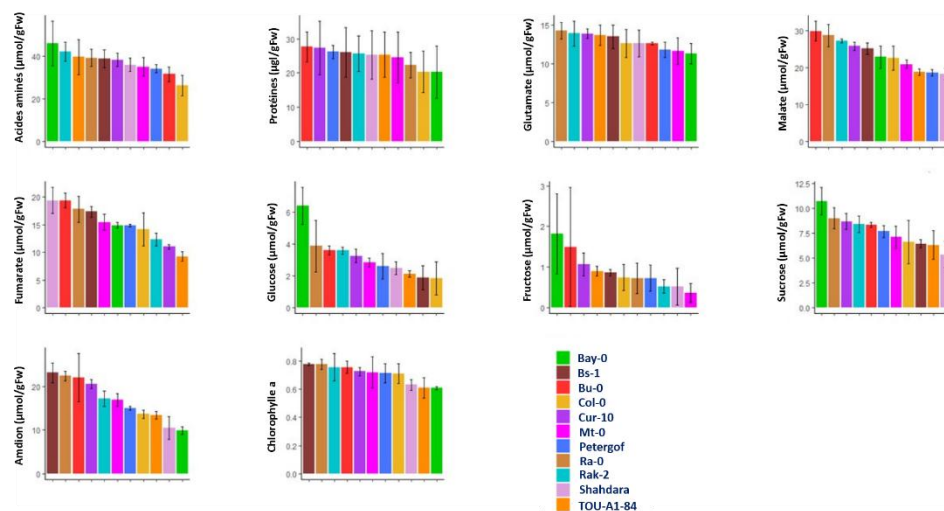


Figure B2.2 | Concentrations des 10 métabolites primaires dosées chez les 11 accessions d'*A. thaliana*
Chaque accession est représentée par une barre de couleur différente. Chaque graphique représente la comparaison du contenu en un métabolite en particulier. Les accessions sont classées par valeur des concentrations en métabolites décroissante. Concentrations exprimées en grammes de matière fraîche (gFW).

Les coefficients de corrélations entre la biomasse fraîche et chacun des métabolites ont été calculés (Tab. B2.1). Il existe une corrélation positive et significative entre la biomasse fraîche, le fumarate, le glucose et le sucrose. Ce résultat est finalement en accord avec ce qui avait été observé dans les essais précédents. En effet, la biomasse (sèche) apparaissait fortement corrélée au fumarate et au glucose dans les échantillons non inoculés avec le TuMV.

Tableau B2.1| Coefficients de corrélations de Spearman calculés pour les 10 métabolites primaires par rapport à la biomasse fraîche

Biomasse F.	0.08	0.1	-0.15	0.25	0.39 *	0.07	0.47 **	0.05	0.52 **	0.03
	Acides Aminés	Protéines	Glutamate	Malate	Fumarate	Amidon	Glucose	Fructose	Sucrose	Chla

Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1

2 Evolution de la charge virale et du contenu en métabolites primaires au cours du temps

2.1 Cinétique du contenu métabolique chez les échantillons non infectés

Les données concernant la biomasse fraîche (en g) montrent une augmentation constante du poids des rosettes au cours du temps. Les profils cinétiques des concentrations en acides aminés ou encore en glutamate se superposent presque entre les différentes accessions. Les concentrations de certains métabolites tels que le malate, le fumarate, l'amidon et les chlorophylles semblent plutôt constantes au cours du temps. Au contraire, pour d'autres métabolites tels que les acides aminés, le glutamate ou encore le fructose des variations plus marquées existent entre chaque date marquées notamment par une augmentation des concentrations à 9 dpi suivie d'une diminution à 13 dpi (Fig. B2.3).

Sur l'ensemble des 10 métabolites primaires dosés sur les échantillons mock-inoculés, les profils métaboliques sont similaires entre les 11 accessions d'*A. thaliana* testées. Ce résultat est en accord avec les données métaboliques obtenues sur les précédents essais (2014 et 2015) où aucune discrimination entre les accessions par rapport à leur sensibilité au TuMV ne pouvait être effectuée avant infection.

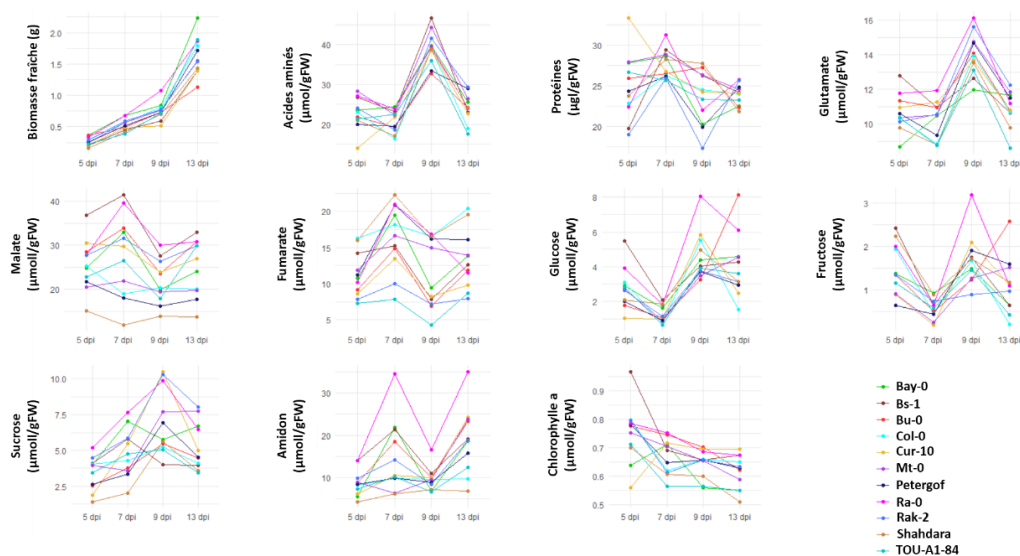


Figure B2.3| Evolution de la biomasse fraîche et des concentrations des 10 métabolites primaires des 11 accessions d'*A.thaliana* testées à 5-7-9 et 13 dpi (days post inoculation) en conditions extérieures

Le premier graphique comprend les données de biomasse fraîche les autres représentent l'évolution des concentrations d'un métabolite donné. Chaque accession est représentée par un trait de couleur différente.

2.2 Cinétique métabolique lors de l'infection par le TuMV

2.2.1 Vérification de la sensibilité au TuMV des accessions

Les 11 accessions d'*A. thaliana* ont été choisies parmi le lot des 29 accessions considérées comme présentant une forte ou faible sensibilité au TuMV et dont une analyse détaillée du métabolisme primaire et secondaire a été effectuée (Rubio et al., 2017 – *in preparation*). Le comportement vis-à-vis du virus a d'ailleurs été vérifié sur deux autres années d'étude en conditions extérieures (Rubio et al., 2017 – *submitted*). Dans un premier temps, l'évolution de l'accumulation du virus a été observée pour chacune de ces 11 accessions testées. La charge virale de chaque accession a été quantifiée par ELISA semi-quantitatif. Il est intéressant de noter que la présence du virus dans les échantillons a été possible uniquement à 13 dpi (Fig. B2.4). Pour toutes les autres dates (5-7-9 dpi), l'ensemble des densités optiques (DO) obtenues sont en dessous du seuil de détection au-delà duquel un échantillon est considéré comme accumulant du virus. Ce seuil est déterminé par la DO moyenne obtenue pour des échantillons d'*A. thaliana* non infectés (trait plein « DO témoin sain » Fig. B2.4). L'évolution de l'infection par le TuMV en conditions extérieures est donc plus lente par rapport à ce qui est normalement observé en conditions contrôlées. En effet, en conditions contrôlées, la présence du TuMV est possible à partir de 7 dpi pour les accessions les plus sensibles comme l'accession Shahdara (données non montrées).

Les différences de sensibilité au TuMV attendues entre les différentes accessions testées sont donc visibles à 13 dpi. Les témoins de l'infection présentent la réponse au TuMV attendue avec C24 qui est une accession sensible et le mutant EMS qui est résistant au TuMV (flèches noires Fig. B2.4). Les accessions attendues sensibles (Bs-1, Bu-0, Col-0, Mt-0, Shahdara et Cur-10) accumulent effectivement le virus et sont donc largement au-dessus des seuils fixés. En ce qui concerne les accessions attendues résistantes (Bay-0, Petergof, Ra-0, Rak-2 et TOU-A1-84), deux accessions (Petergof et Rak-2) présentent des valeurs de DO qui dépassent la DO du témoin sain même si elles restent en-dessous du seuil correspondant à deux fois la valeur de DO du témoin sain qui correspond au seuil au-dessus duquel la sensibilité au virus est très forte (flèche rose Fig. B2.4). Ces accessions accumulent du virus même si les valeurs de DO restent inférieures à celles obtenues chez les accessions sensibles. Aucune détection du virus n'est en revanche relevée chez Bay-0, Ra-0 et TOU-A1-84 qui présentent donc une réponse au TuMV similaire de celle du mutant EMS.

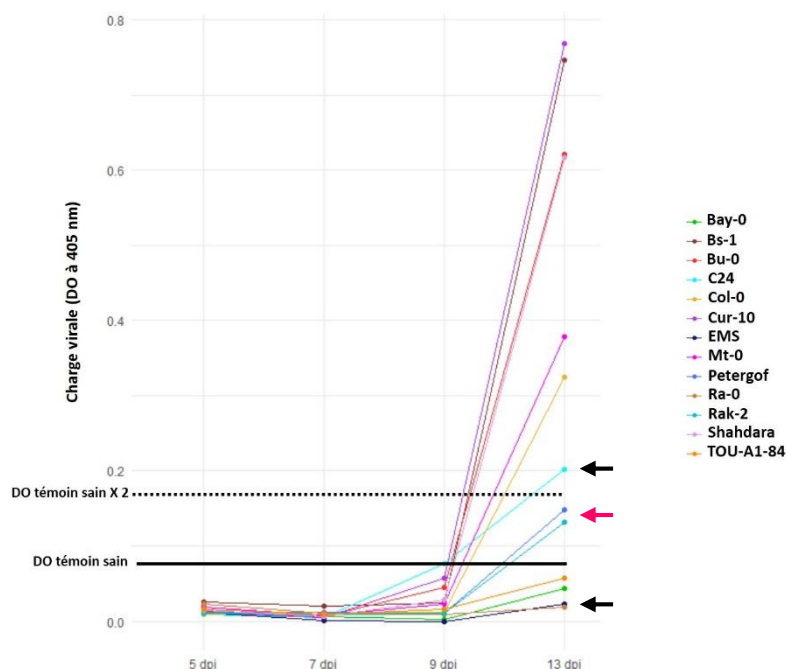


Figure B2.4 | Cinétique de l'accumulation de la charge virale pour les différentes accessions d'*A.thaliana* testées en conditions extérieures

11 accessions ont été testées ainsi que deux témoins de l'infection : l'accession C24, sensible et son mutant EMS (mutant *elf(iso)4G*), résistant. La charge virale a été quantifiée par DAS ELISA à 5-7-9 et 13 dpi (days post infection). Deux seuils sont déterminés à partir de la valeur de DO obtenue pour un échantillon d'*A.thaliana* sain i) un premier seuil correspondant à cette valeur de DO (trait noir plein) en dessous duquel les échantillons sont négatifs à l'ELISA ii) un deuxième seuil correspondant à 2 fois la valeur de DO du témoin sain (trait noir pointillé) au-dessus duquel les échantillons sont très sensibles au TuMV.

Les pourcentages d'infection ont été calculés à 13 dpi pour chacune des accessions. Ils permettent de montrer que pour les accessions sensibles entre 89 et 100% des plantes sont positives en ELISA. En ce qui concerne les accessions résistantes les pourcentages d'infection sont en-dessous de 25%. Bay-0 qui est l'accession résistante de référence sur l'ensemble des essais extérieurs présente un pourcentage d'infection de l'ordre de 10% similaire à ce qui a déjà été observé sur les autres essais. En revanche, les accessions Petergof et Rak-2 présentent un pourcentage d'infection au-dessus de 50% confirmant leur sensibilité intermédiaire observée (Tab. B2.2).

Le statut sensible ou résistant des accessions normalement attendu en réponse au TuMV est donc vérifié pour 9 des 11 accessions testées. C'est donc sur ces neuf accessions que les analyses de l'évolution du métabolisme primaire seront menées.

Tableau B2.2| Pourcentages d'infection calculés pour chaque accession à 13 dpi

Accession	Nbre échantillons positifs à ELISA	Nbre Total Echantillons	% Infection
Ra-0	0	9	0
Bay-0	1	9	11
TOU-A1-84	2	9	22
Rak-2	5	9	56
Petergof	5	9	56
Bu-0	8	9	89
Bs-1	8	9	89
Mt-0	8	9	89
Col-0	8	9	89
Shahdara	9	9	100
Cur-10	9	9	100

2.2.2 Cinétique du contenu en métabolites primaires

Les mesures de charge virale ont été effectuées de manière individuelle sur chaque plante de chaque accession. Afin d'augmenter la robustesse des dosages métaboliques en limitant les variations intra-génotype, des pools de plantes sont effectués pour chacune des accessions. A 5, 7 et 9 dpi, un pool de 3 plantes/accession constitue une répétition. Il y a au total 3 répétitions/accession. A 13 dpi, ce système de pool doit tenir compte des charges virales mesurées. Ainsi, en ce qui concerne les génotypes sensibles, les pourcentages d'infection étant élevés (entre 89 et 100%) seuls les quelques échantillons négatifs en ELISA sont écartés. Les échantillons négatifs sont donc considérés comme des échappés à l'infection. En ce qui concerne les échantillons négatifs, pour les accessions Ra-0, Bay-0 et TOU-A1-84 dont les pourcentages d'infection sont inférieurs à 25%, les échantillons positifs sont écartés de l'analyse.

Dans un premier temps, une analyse comparative est effectuée entre échantillons mock-inoculés et ceux inoculés avec le TuMV afin de mettre en évidence l'effet de l'infection sans tenir compte des différences de sensibilité entre les accessions. Dans un deuxième temps, une comparaison de la réponse métabolique entre les accessions sensibles et résistantes a été menée en condition d'infection par le TuMV.

2.2.2.1 Effet de l'infection sur l'évolution du contenu métabolique

Des comparaisons globales à 5-7-9 et 13 dpi ont été effectuées entre échantillons infectés par le TuMV et ceux mock-inoculés sans tenir compte des différences de sensibilité au TuMV entre les différentes accessions. Elles montrent que des différences significatives existent et ce dès les temps précoces de l'infection. D'ailleurs c'est à 5 dpi que le nombre de métabolites dont le contenu diffère entre échantillons inoculés par le TuMV et ceux mock-inoculés est le plus grand : 7 métabolites sur 10 sont concernés par ces différences significatives (Tab. B2.3). Ces différences sont liées à des concentrations plus élevées en métabolites chez les échantillons inoculés par le TuMV. Le malate, est le seul métabolite qui ne présente aucune différence entre échantillons inoculés par le TuMV et les mock-inoculés quel que soit le temps après inoculation. En ce qui concerne les teneurs en chlorophylles a elles restent toujours significativement supérieures chez les échantillons mock-inoculés (Tab. B2.3).

Tableau B2.3| Comparaisons du contenu en métabolites primaires entre échantillons infectés (I) et non infectés (M) par le TuMV

	Acides Aminés	Protéines	Glutamate	Malate	Fumarate	Glucose	Fructose	Saccharose	Amidon	Chlorophylle a
5 dpi	I > M	I > M	I > M	NS	I > M	NS	NS	I > M	I > M	M > I
7 dpi	M > I	NS	NS	NS	M > I	NS	NS	NS	M > I	M > I
9 dpi	M > I	NS	NS	NS	NS	NS	M > I	NS	NS	M > I
13 dpi	NS	I > M	NS	NS	I > M	NS	NS	NS	NS	M > I

Des comparaisons statistiques (Test de Student ou test de Wilcoxon-Mann-Whitney selon le respect de la normalité des variables étudiées avec n = 9) ont été effectuées pour les concentrations en 10 métabolites primaires entre les échantillons infectés par le virus (notés « I ») et les échantillons mock-inoculés, sans virus (notés « M »). Lorsque des différences statistiquement significatives existaient le sens de la différence a été noté **i**) I > M lorsque la concentration du métabolite est supérieure dans les échantillons infectés par le TuMV ou **ii**) M > I, lorsque la concentration du métabolite est supérieure dans les échantillons mock-inoculés. L'absence de différence statistiquement significative est notée « NS ». Le seuil de significativité statistique a été fixé à 5% ($\alpha = 0.05$).

L'ensemble de ces observations permettent de montrer que dès les premières étapes de l'infection virale, des modifications significatives dans le contenu métabolique ont lieu. En revanche, très peu de différences sont mises en évidence notamment à 13 dpi alors que les analyses déjà effectuées à ce même moment de l'infection montraient pourtant de fortes perturbations du métabolisme primaires se traduisant par une discrimination des accessions en fonction de leur sensibilité au virus. Cela pourrait s'expliquer par le fait d'avoir effectué cette analyse comparative en ne tenant pas compte des différences de sensibilité au virus des accessions testées. En effet, parmi les neuf accessions analysées, 3 accessions sont considérées comme résistantes au TuMV alors que 6 accessions sont considérées sensibles au virus. Une autre analyse a donc été effectuée en séparant – dans les échantillons inoculés par le TuMV – les accessions sensibles et celles résistantes au virus. Aucune distinction n'est faite pour les échantillons mock-inoculés puisqu'aucune différence significative du contenu en métabolites primaires n'a été observée entre les accessions en absence de virus. De nouvelles comparaisons du contenu métabolique ont donc été menées (Fig. B2.5). Cela permet de montrer qu'à un temps précoce, 5 dpi, le contenu métabolique des accessions inoculées avec le TuMV qu'elles soient résistantes ou sensibles est identique pour la plupart des métabolites primaires testés (Fig. B2.5). Il n'y a que le fumarate qui semble s'accumuler majoritairement chez les sensibles sans que cette différence soit statistiquement significative. Les différences décrites précédemment entre échantillons mock-inoculés et inoculés avec le TuMV (Tab. B2.3) sont clairement retrouvées pour les teneurs en acides aminés, protéines, glutamate, fumarate, sucrose, amidon et chlorophylle a (Fig. B2.5).

C'est à partir de 9 dpi que les premières différences (même si pas toujours statistiquement significatives – Tests de Student) entre les accessions sensibles et résistantes inoculées par le TuMV sont mises en évidence. Par exemple, la concentration en acides aminés a tendance à être supérieure chez les accessions sensibles (35.60 $\mu\text{mol/g}$ de matière fraîche) par rapport aux accessions résistantes (30.46 $\mu\text{mol/g}$ de matière fraîche). Des différences significatives dans la concentration en glucose existent entre accessions sensibles (4.28 $\mu\text{mol/g}$ de matière fraîche) et résistantes (2.34 $\mu\text{mol/g}$ de matière fraîche). A 13 dpi, les différences entre accessions sensibles et résistantes deviennent statistiquement significatives pour les concentrations en acides aminés, protéines, glutamate, glucose, saccharose et amidon. Le contenu en ces métabolites primaires étant significativement supérieur chez les accessions sensibles par rapport aux accessions résistantes (Fig. B2.5).

Il semblerait donc qu'à des temps précoces de l'infection, des modifications métaboliques existent en condition d'infection par rapport à des échantillons mock-inoculés. Cette réponse métabolique chez les échantillons inoculés avec le TuMV est identique entre accessions sensibles et résistantes. En revanche, à partir de 9 dpi, des réponses métaboliques différentes sont mises en place en fonction de la sensibilité des accessions au TuMV. Des différences significatives entre accessions sensibles et résistantes sont d'ailleurs clairement mises en évidence à 13 dpi.

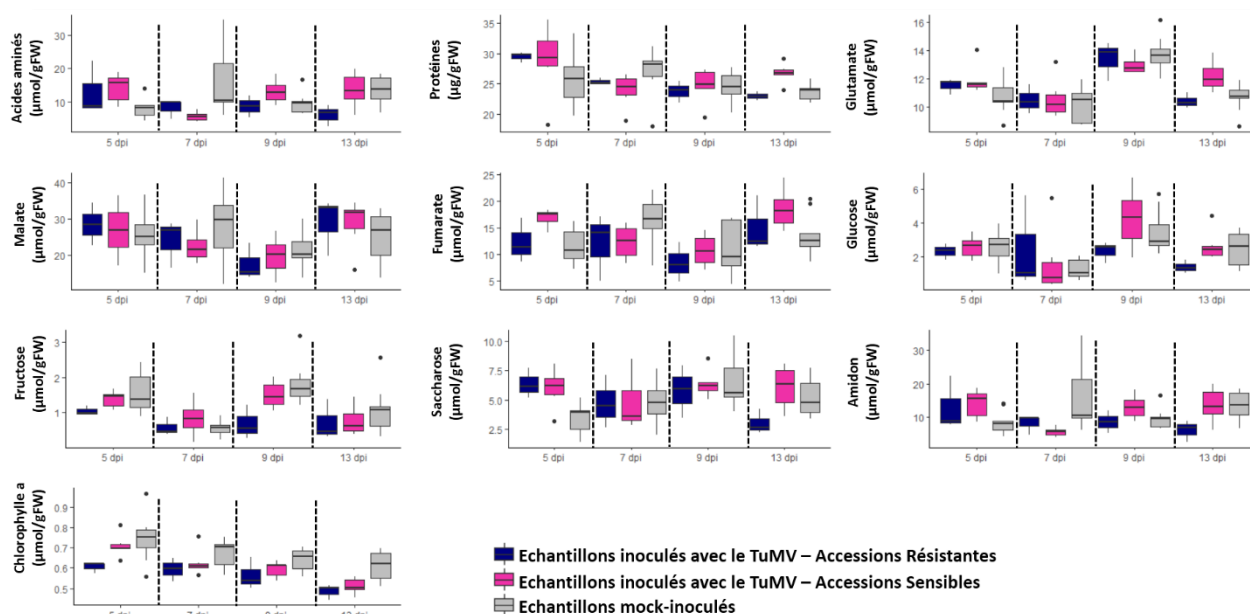


Figure B2.5 | Evolution du contenu en 10 métabolites primaires chez les échantillons mock-inoculés et chez les échantillons inoculés avec le TuMV

Chaque graphique représente l'évolution du contenu en un métabolite pour les quatre dates d'analyse, 5-7-9 et 13 dpi. Les échantillons inoculés par le TuMV ont été séparés en deux groupes avec d'une part les accessions résistantes (en bleu) et d'autres par les accessions sensibles (en rose) au TuMV. Les échantillons mock-inoculés correspondent aux boxplots gris.

2.2.2.2 Evolution du contenu en métabolites primaires en fonction de la sensibilité des accessions au TuMV en condition d'infection

Au vu des résultats précédents, l'infection au TuMV a un effet sur le contenu en métabolites primaires et ce dès les premières étapes de la maladie. Les analyses précédentes ont également montré que les différences dans le contenu métabolique commencent à être visibles à partir de 9 dpi entre les accessions sensibles et résistantes. L'objectif est donc de

faire des analyses complémentaires sur les échantillons infectés par le TuMV afin de voir comment et à partir de quel moment des variations métaboliques en lien avec l'infection peuvent être décrites. Il s'agit également de vérifier s'il est possible de discriminer les accessions en fonction de leur sensibilité au TuMV comme cela a déjà été montré sur les autres essais extérieurs (Rubio et al., 2017 – *in preparation*). Des analyses en composantes principales ont été effectuées à partir des concentrations des 10 métabolites primaires obtenues à 5-7-9-13 dpi. Elles montrent qu'aux temps précoces de l'infection – 5 et 7 dpi – aucune variable métabolique ne semble permettre de séparer les deux groupes d'accessions (Fig. B2.6). Ces accessions sont retrouvées d'ailleurs au centre du graphique des individus (Fig. B2.6a). A partir de 9 dpi, les accessions résistantes sont retrouvées en bas à gauche du graphique des individus alors que les accessions sensibles ont tendance à être retrouvées en haut à droite (Fig. B2.6a). En regardant le graphique des variables, ce seraient les concentrations en glucose, fructose et protéines qui permettraient d'expliquer la distribution des accessions sur le graphique des individus. D'ailleurs, la variable de biomasse fraîche est anti-corrélée à ces concentrations métaboliques et va dans les sens des accessions résistantes (Fig. B2.6b). La séparation entre les accessions sensibles et résistantes est clairement visible à 13 dpi (Fig. B2.6a). Cette distinction semble être particulièrement expliquée par les concentrations en acides aminés, protéine, glutamate et sucrose qui s'accumulent chez les accessions sensibles (Fig. B2.6b). Une corrélation positive et significative est d'ailleurs mise en évidence entre la charge virale (DO) et les concentrations en ces métabolites. Des tendances de corrélation positives entre la charge virale, le glucose et l'amidon sont aussi à noter (Tab. B2.4). De plus, la charge virale est négativement corrélée à la biomasse fraîche ce qui avait déjà été montré lors des essais extérieurs précédents (Tab. B2.4). Finalement cette distinction métabolique entre accessions sensibles et résistantes au TuMV est retrouvée au même moment où la détection du TuMV par ELISA semi-quantitatif devient possible.

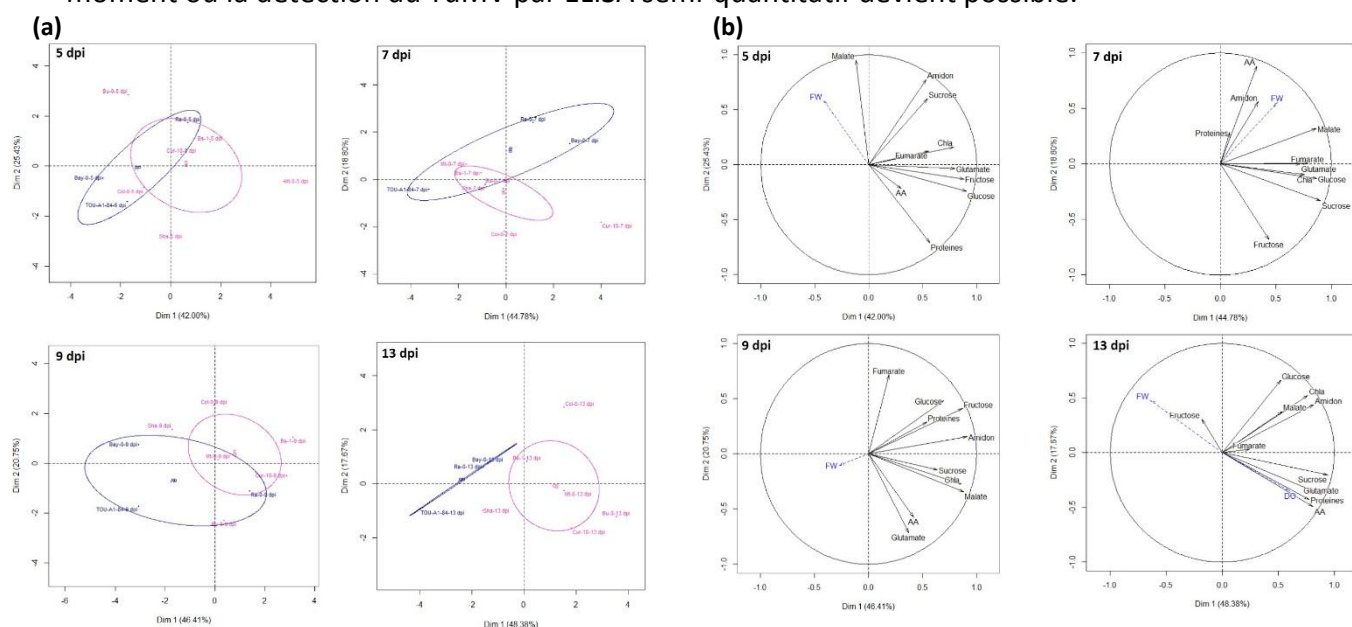


Figure B2.6 | Analyses en composantes principales réalisées à 5, 7, 9 et 13 dpi en utilisant les concentrations en métabolites primaires des échantillons inoculés par le TuMV

Les analyses en composantes principales ont été effectuées en utilisant les concentrations des 10 métabolites primaires obtenues à 5, 7, 9 et 13 dpi. Les graphiques des individus (a) ainsi que les graphiques des variables (b) ont été analysés. Dans le graphique des individus les accessions sensibles sont représentées en rose et les accessions résistantes en bleu. Sur le graphique des variables, la biomasse fraîche est représentée comme une variable descriptive à 5-7-9 et 13 dpi (FW en bleu). A 13 dpi, la charge virale (DO en bleu) est également représentée comme variable descriptive.

Tableau B2.4 | Coefficient de corrélation de Spearman calculés entre les 10 métabolites primaires, la biomasse fraîche et la charge virale chez les échantillons inoculés par le TuMV

Biomasse	-0.88 **	-0.82 *	-0.87 **	-0.88 **	-0.13	-0.02	-0.43	-0.47	-0.15	-0.68 -	-0.13
Charge Virale		0.77 *	0.75 *	0.80 *	-0.15	0.07	0.30	0.50	-0.07	0.67 -	-0.07
		Acides Aminés	Protéines	Glutamate	Malate	Fumarate	Amidon	Glucose	Fructose	Sucrose	Chla

Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Discussion

Cette étude avait pour objectif de compléter l'étude sur la réponse métabolique d'*A. thaliana* au TuMV. En effet, si une caractérisation fine de cette réponse a déjà été menée sur un large panel d'accessions, toutes les analyses ont été menées à un moment précis de l'infection, plus précisément au moment où les premiers symptômes étaient visibles chez les accessions les plus sensibles. Or, tout comme le cycle viral, la réponse métabolique pouvait être dynamique. En fonction des métabolites, de fortes variations sont décrites au cours de l'infection. L'objectif de cette étude était donc de regarder l'évolution du contenu en métabolites primaires à différentes étapes de l'infection par le TuMV, toujours en conditions extérieures. Pour cela 11 accessions d'*A. thaliana* ont été sélectionnées parmi les 29 accessions présentant une forte ou faible sensibilité au TuMV dont les analyses fines du métabolisme primaire et secondaire ont été effectuées au cours des analyses décrites précédemment. L'analyse du contenu en 10 métabolites primaires a été menée à 5, 7, 9 et 13 jours après inoculation (dpi). Ces dates de prélèvement ont été déterminées de façon à pouvoir analyser la réponse métabolique à des temps précoces de l'infection (5-7 dpi) mais également à partir du moment où le développement de l'infection devient important. De plus, ont aussi été prises en compte pour le choix de ces dates de prélèvements, les connaissances en termes d'évolution de la maladie au cours du temps en conditions contrôlées.

Une autre information qui manquait à cette analyse métabolique était celle du profil métabolique des accessions avant toute inoculation que ce soit avec le virus ou sans (échantillons mock-inoculés). Des analyses métaboliques ont donc été effectuées sur des échantillons prélevés avant l'inoculation, au « temps zéro (T0) ». Cette analyse à T0, n'a pas permis de mettre en évidence de différences particulièrement significatives entre les accessions testées. En revanche, en utilisant les données de biomasse fraîche, il a été possible de mettre en évidence la corrélation significative et positive de la biomasse avec les concentrations en fumarate. Ce résultat ayant déjà été montré lors des précédentes études. Cela permet de montrer à nouveau que de fortes concentrations en fumarate sont retrouvées chez *A. thaliana*. Cela confirme l'idée que le fumarate constitue un fraction significative du carbone fixé dans les feuilles chez *A. thaliana* (Chia *et al.*, 2000).

Ainsi, au temps T0, il n'y a donc pas de différences significatives dans le contenu en métabolites primaires des accessions testées qui auraient pu être reliées à la sensibilité ou résistance au TuMV. Aucune distinction entre les accessions, en fonction de leur sensibilité au TuMV ne peut être faite sur la base du contenu en métabolites primaires mais aussi en métabolites secondaires. En effet, l'analyse menée sur les 29 accessions d'*A. thaliana* n'a pas permis de distinguer avant infection les accessions à forte et faible sensibilité au TUMV alors que 505 variables du métabolisme secondaire ont été analysées (Rubio *et al.*, 2017 – *in preparation*). Ce même constat a également été fait dans l'étude de Sade et ses collaborateurs

(2015) sur des lignées de tomate résistantes et sensibles au *Tomato yellow leaf curl virus* (TYLCV). Les lignées résistantes et sensibles montraient des patterns métaboliques similaires avant l'infection (Sade *et al.*, 2015).

S'il n'est pas possible de déterminer la sensibilité des accessions au TuMV à partir des concentrations en métabolites primaires avant infection, d'importantes perturbations du métabolisme primaire et secondaire avaient déjà été mises en évidence dans les précédentes études de cette thèse. Il s'agissait donc **i)** de caractériser l'évolution du contenu en 10 métabolites primaires en condition d'infection mais aussi **ii)** de voir à partir de quel moment au cours de l'infection, la réponse métabolique variait entre les accessions sensibles et résistantes. Ainsi, les analyses montrent que le métabolisme primaire est perturbé dès les temps précoces de l'infection. En effet, des différences significatives entre échantillons mock-inoculés et inoculés par le TuMV ont été identifiées dès 5 dpi. En revanche, à ces temps précoces les accessions sensibles et résistantes inoculées par le TuMV ont un profil métabolique similaire. La distinction métabolique entre les accessions en fonction de leur sensibilité ne devient claire qu'à 13 dpi même si à 9 dpi une séparation entre les deux groupes d'accessions semble commencer à se mettre en place. Cette distinction clairement visible à 13 dpi est en accord avec ce qui a déjà été décrit lors de nos précédentes études en conditions extérieures. De plus, les différences entre ces accessions sont en relation avec des concentrations en métabolites primaires significativement plus importantes chez les accessions sensibles, comme cela a déjà été observé dans des études précédentes. De plus, les corrélations négatives entre biomasse fraîche et charge virale montrent qu'elles n'arrivent plus, non plus, à assurer leur croissance. Cette distinction visible à 13 dpi correspond au moment à partir duquel il est possible de détecter le virus dans les échantillons infectés par ELISA semi-quantitatif. Les perturbations métaboliques notamment chez les accessions sensibles, avec l'accumulation de métabolites primaires sont donc quantifiables à partir du moment où l'infection est détectée. Cela est en accord avec ce qui a déjà pu être observé dans d'autres études de la réponse métabolique des plantes à une infection virale. En effet, dans leur travail sur *A. thaliana* (accession Col-0) en réponse au *Tobacco rattle virus* (TRV), Fernandez-Calvino et ses collaborateurs (2014), montrent que des changements métaboliques importants sont reliés au moment où l'infection virale est maximale comme par exemple l'accumulation de sucres solubles (une quantification du virus par qRT-PCR ayant été effectuée à chaque temps de prélèvement) (Fernandez-Calvino *et al.*, 2014).

Dans le cadre de notre étude, la quantification du virus dans les échantillons infectés n'a été possible qu'à 13 dpi par ELISA. La charge virale dépasse donc les seuils de détection entre 9 et 13 dpi. Cette évolution de la maladie diffère légèrement de ce qui avait été pour le moment observé en conditions contrôlées. En effet, en chambre de culture ou même en serre, la détection du virus par ELISA est possible pour les accessions les plus sensibles dès 7 dpi. L'évolution de l'infection est donc plus lente en conditions extérieures. Il aurait donc été intéressant d'avoir un point de prélèvement intermédiaire entre 9 et 13 dpi afin de pouvoir analyser les modifications du métabolisme au moment où la charge virale devient détectable.

Cette étude a donc permis **i)** d'analyser l'empreinte métabolique des accessions d'*A. thaliana* avant inoculation au temps T0 permettant de montrer que la sensibilité au TuMV ne peut pas

être prédite sur la base du contenu en métabolites primaires des accessions ; **ii**) de confirmer qu'aucune distinction ne peut être faite entre accessions résistantes et sensibles sur la base des concentrations en métabolites primaires en absence de virus (échantillons mock-inoculés). Ce résultat a donc été vérifié sur trois essais différents en conditions extérieures en utilisant des données du métabolisme primaire et secondaire ; **iii**) de mettre en évidence que des perturbations métaboliques existent dès les phases précoces de l'infection. En revanche la distinction entre accessions résistantes et sensibles ne semble être clairement définie que lorsque l'infection devient détectable (ici en ELISA). Comme cela avait été montré sur les deux précédents essais extérieurs, les accessions sensibles accumulent significativement des métabolites primaires en comparaison aux accessions résistantes.

B-3. Architecture génétique de la réponse métabolique d'*A. thaliana* au TuMV en conditions extérieures

Introduction

La caractérisation de la réponse d'*A. thaliana* au TuMV en conditions extérieures a dans une première partie de cette thèse été effectuée à partir des variables liées à la maladie. Ce sont ces variables qui ont été utilisées dans des analyses de GWA visant à décrire les bases génétiques de la réponse d'*A. thaliana* au TuMV en conditions extérieures. Plusieurs régions génomiques ont ainsi été identifiées contenant de potentiels gènes candidats de l'interaction *A. thaliana*/TuMV. Des analyses d'enrichissement menées sur les top SNP de l'analyse GWA ont permis de regrouper les gènes selon leur appartenance à des fonctions biologiques ou moléculaires particulières. Il apparaît ainsi que les processus métaboliques en lien avec les carbohydrates sont significativement enrichis (Rubio *et al.*, 2017 – *submitted*). De plus, parmi les différents gènes annotés dans les différents loci identifiés, plusieurs sont en lien avec des fonctions métaboliques. Par exemple le gène codant pour la glutamate dehydrogenase 1 (AT5G18170) est retrouvé dans une région génomique identifiée sur le chromosome 5. Des gènes impliqués dans la réponse aux sucres solubles (glucose, fructose, saccharose) sont également retrouvés tel que le gène AT5G41670 qui code pour une phosphogluconate dehydrogenase. Même si l'ensemble de ces résultats pourraient être le reflet des variations métaboliques importantes en conditions de stress comme cela a déjà été démontré (Fernandez-Calvino *et al.*, 2014), il faut garder à l'esprit que chez *A. thaliana* les gènes impliqués dans le métabolisme cellulaire représentent une des catégories les plus abondantes du génome (22% du génome) (Kaul *et al.*, 2000).

Une des caractéristiques de l'étude du pathosystème naturel *A. thaliana*/TuMV dans le cadre de cette thèse est le phénotypage large effectué. En effet, un phénotypage en lien avec des variables métaboliques a été effectué permettant de décrire la réponse métabolique d'*A. thaliana* lors d'une infection virale en conditions extérieures (Rubio *et al.*, 2017 – *in preparation*). Ainsi, un des objectifs était d'utiliser ces données métaboliques dans des analyses GWA afin **d'analyser l'architecture génétique de la réponse métabolique d'*A. thaliana* au TuMV en conditions extérieures et d'identifier de potentielles régions génomiques co-localisant avec des loci identifiés avec les données maladie.**

Pour cela, des analyses GWA ont été menées sur un set de 130 accessions d'*A. thaliana* choisies parmi l'ensemble des accessions testées lors de l'essai extérieur de 2014. L'analyse GWA a été réalisée avec le modèle linéaire mixte GEMMA utilisé dans les diverses analyses d'association réalisées dans le cadre de cette thèse. Les 100 meilleurs SNPs de chaque analyse ont été sélectionnés. A partir de ces données, des zones comprenant des SNPs dont la distance qui les sépare est inférieure ou égale à 10kb ont été mises en évidence. Le seuil de p-value 10^{-4} ($-\log(p\text{-value}) > 4$) a été retenu afin de considérer une région comme potentiellement associée à la réponse au TuMV. Les données relatives aux mesures des 10 métabolites primaires (acides aminés, protéines, glutamate, malate, fumarate, glucose, fructose, saccharose, amidon et chlorophylle a) ont été utilisées. L'ensemble des analyses ont été effectuées sur les échantillons inoculés par le virus mais également sur les échantillons mock-

inoculés. Les échantillons inoculés par le TuMV ont été normalisés par les teneurs en métabolite des échantillons mock-inoculés.

1 GWA et métabolisme primaire chez *A. thaliana* inoculée par le TuMV

Pour chacun des dix métabolites primaires analysés, la distribution des accessions en fonction de leur concentration métabolique permet de montrer la variabilité qui existe au sein de ces 130 accessions testées (Fig. B3.1)

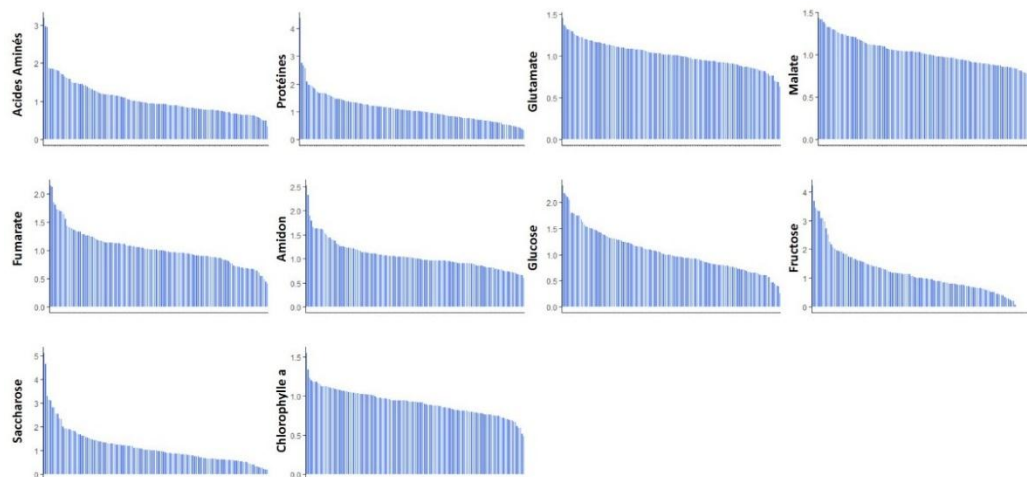


Figure B3.1 | Distribution des concentrations en 10 métabolites primaires dosées dans 130 accessions d'*A. thaliana* en conditions extérieures

La distribution de chaque métabolite est représentée par un graphique où les 130 accessions d'*A. thaliana* sont représentées par ordre décroissant des concentrations. L'ordonnée correspond aux concentrations métaboliques exprimées en $\mu\text{mol/g}$ de matière fraîche pour les acides aminés, glutamate, malate, fumarate, amidon, glucose, fructose, sucrose, chlorophylle a et en $\mu\text{g/g}$ de matière fraîche pour les protéines. Ces distributions ont été obtenues à partir des données métaboliques des échantillons inoculés avec le TuMV et normalisées par les données des échantillons mock-inoculés.

Avant de mettre en place les analyses de GWA à partir des données des concentrations en métabolites primaires, l'héritabilité (H^2) a été estimée pour chacun des traits métaboliques à partir de la même méthode utilisée dans le cadre de l'étude des traits phénotypiques liés à la maladie chez la population mondiale d'*A. thaliana* (Rubio et al., 2017 – submitted). Les valeurs sont comprises entre 0.2 et 0.74 avec une valeur moyenne de 0.51 (Tab. B3.1). Le coefficient de variation est estimé à 31.5% reflétant la variabilité de ces H^2 qui existe en fonction des métabolites. Les H^2 estimées à partir des données relatives aux concentrations en protéines ainsi qu'aux concentrations de fumarate sont très faibles. De plus, les analyses statistiques révèlent qu'il n'y a aucun effet significatif du génotype, en conditions d'infection, sur les concentrations en protéines et fumarate (données non montrées). Ainsi, les données relatives à ces deux métabolites n'ont pas été exploitées dans les analyses GWA.

Tableau B3.1 | Héritabilités estimées pour chacun des traits métaboliques

Variable Métabolique	H^2
Acides Aminés	0.71
Glutamate	0.39
Malate	0.5
Amidon	0.74
Glucose	0.54
Fructose	0.54
Saccharose	0.2
Chlorophylle a	0.47

Les premières analyses GWA ont donc été effectuées sur les données de concentrations de huit métabolites primaires (acides aminés, glutamate, malate, amidon, glucose, fructose, sucrose, chlorophylle a). La première remarque qui peut être faite sur ces résultats est en lien avec les faibles p-value observées pour l'ensemble des régions identifiées. De façon globale, les valeurs des $-\log(p\text{-value})$ sont en dessous de 5 avec une moyenne des valeurs autour de 4, à la limite du seuil de significativité fixé. Ainsi, 58 régions génomiques sont identifiées sur l'ensemble des cinq chromosomes d'*A. thaliana* dont au moins un des SNP présente une p-value au-dessus du seuil de significativité (Fig. B3.A1). De plus, 23 de ces régions génomiques ont été retrouvées avec au moins deux variables métaboliques différentes appartenant ou non à des voies métaboliques proches. De ce fait, près de 60% des loci identifiés l'ont été à partir d'une seule des huit variables analysées en GWA.

Le nombre de loci mis en évidence sur chacun des chromosomes par chacune des variables testées a été calculé. Les variables faisant ressortir le plus grand nombre de loci sont le glucose, le fructose, le saccharose, l'amidon et le malate (Tab. B3.2).

Tableau B3.2 | Nombre de zones du génome identifiées sur chaque chromosome et pour chaque trait métabolique

Variable Métabolique	Nbre d'associations				
	Chr1	Chr2	Chr3	Chr4	Chr5
Acides Aminés	0	2	3	0	0
Glutamate	1	1	5	1	1
Malate	1	1	5	3	3
Amidon	1	0	4	5	5
Glucose	4	0	1	5	0
Fructose	6	0	4	1	1
Saccharose	3	1	5	5	2
Chlorophylle a	4	0	1	1	1

Les données de concentrations en métabolites primaires ont été obtenues suite à des dosages sur des échantillons inoculés par le TuMV et récoltés 13 jours après inoculation. Seules les zones génomiques pour lesquelles au moins un des SNP présente une valeur de p-value supérieure au seuil de $-\log(p\text{-value}) = 4$ ont été comptabilisées.

Les diverses analyses effectuées au préalable sur ces données métaboliques avaient permis de mettre en évidence chez les individus inoculés par le TuMV, des différences dans le contenu métabolique entre les individus en fonction de leur sensibilité au TuMV. Ainsi, il a été montré que les accessions sensibles accumulaient de façon globale plus de métabolites primaires en condition d'infection en comparaison aux accessions résistantes. Ces différences étaient statistiquement significatives pour 6 des 10 métabolites primaires dosés (acides aminés, protéines, glutamate, sucrose, fructose et chlorophylle a). Ainsi, les 100 meilleurs SNPs de chacune des analyses GWA menées sur ces métabolites primaires s'accumulant différemment en fonction de la sensibilité des accessions ont été sélectionnés. Ces données GWA ont été analysées en commun avec les 100 meilleurs SNPs des analyses GWA menées sur les données de charge virale et de biomasse sèche. L'objectif étant de potentiellement mettre en évidence des régions génomiques retrouvées de façon commune avec les données en lien avec la maladie (charge virale et biomasse sèche) et les données métaboliques.

Ainsi, deux régions génomiques ont été identifiées à la fois avec les données en lien avec la biomasse sèche et les acides aminés (chromosome 1 – 25.921.758..25.925.441, $-\log(p\text{-value})$ top SNP = 4.83) mais aussi avec les données de biomasse sèche et le saccharose (chromosome 3 – 8.415.034..8.419.984, $-\log(p\text{-value})$ top SNP = 4.45) (Tab. B3.3a). Trois régions ressortent avec les données en lien avec la charge virale/glutamate, charge virale/acides aminés et

charge virale/fructose (Tab. B3.3b). Parmi ces trois zones, la région identifiée sur le chromosome 3 avec les données charge virale/acides aminés (21.581.868..21.592.579, $-\log(p\text{-value})$ top SNP = 4.83) co-localise avec une région caractérisée dans le cadre des analyses GWA sur les données maladie (Rubio et al. 2017, *submitted*). Il s'agit plus particulièrement de la région où se trouve le gène *RTM3* déjà caractérisé dans la réponse d'*A. thaliana* aux potyvirus (Whitham *et al.*, 1999; Decroocq *et al.*, 2006; Cosson *et al.*, 2012). Cette région fait effectivement partie des 58 régions mises en évidence sur l'ensemble des analyses GWA menées avec les données du métabolisme primaire et en particulier avec les données des concentrations en acides aminés.

Tableau B3.3| Régions du génome identifiées à partir des analyses GWA sur les données métaboliques, de charge virale et de biomasse sèche

Variables	Chr	Zone bp (Début..Fin)	Top SNP	$-\log(p\text{value})$ top SNP	Gènes	Description
(a)						
Biomasse Sèche	1	25.921.758..25.925.441	25.925.441	4.83	AT1G68935	Unknown protein
Acides Aminés					AT1G68940	Armadillo/beta-catenin-like repeat family protein - Protein ubiquitination
Biomasse Sèche	3	8.415.034..8.419.984	8.419.984	4.45	AT3G23470	Cyclopropane-fatty-acyl-phospholipid synthase - Lipid Biosynthetic process
Saccharose					AT3G23480	Cyclopropane-fatty-acyl-phospholipid synthase - Lipid Biosynthetic process
(b)						
Charge Virale	3	2.334.729..2.336.134	2.334.729	4.38	AT3G07320	O-Glycosyl hydrolases family 17 protein - Carbohydrate metabolic process
Glutamate						
Charge Virale	3	21.581.868..21.592.579	21.591.977	4.83	AT3G58290	TRAF-like superfamily protein
Acides Aminés					AT3G58300	Arabidopsis phospholipase like protein
					AT3G58310	Unknown protein
					AT3G58320	Ubiquitin specific protease family C19-related protein
					AT3G58330	Arabidopsis phospholipase like protein
					AT3G58340	TRAF-like superfamily protein
					AT3G58350	RTM3
Charge Virale	5	20.151.972..20.154.614	20.151.972	4.12	AT5G49640	Unknown protein
Fructose					AT5G49650	Xylulose kinase 2 - Xylulose metabolic process

(a) Régions génomiques identifiées en commun à partir des résultats GWA obtenus sur les données de biomasse sèche ainsi que de métabolites primaires. En l'occurrence, il s'agissait des données de concentrations en acides aminés et en saccharose **(b)** Régions génomiques identifiées en commun à partir des résultats GWA obtenus sur les données de charge virale ainsi que du contenu en glutamate, acides aminés et fructose

Chr = N° de Chromosome ; Zone bp (Début..Fin) = Délimitation des zones génomiques identifiées dont les SNPs sont distants de maximum 10kb ; Top SNP = Position du meilleur SNP de la zone génomique identifiée ; $-\log(p\text{value})$ top SNP = Valeur de la pvalue du meilleur SNP

Afin d'éventuellement améliorer la détection – même si cela limite le nombre de variables analysées – des analyses GWA ont également été conduites avec les données issues des premiers axes (axes 1 et 2) d'analyses en composantes principales comme cela a été effectué par exemple dans l'étude du métabolisme du carbone et de l'azote chez le maïs (Zhang *et al.*, 2015). Ces données ont donc été obtenues suite aux analyses en composantes principales réalisées à partir des données métaboliques ayant été regroupées par voies métaboliques. Les corrélations entre ces variables ont été vérifiées par les calculs de coefficient de Spearman (Fig. B3.2). Ces données peuvent être considérées comme des indicateurs clés du métabolisme de l'azote et du carbone.

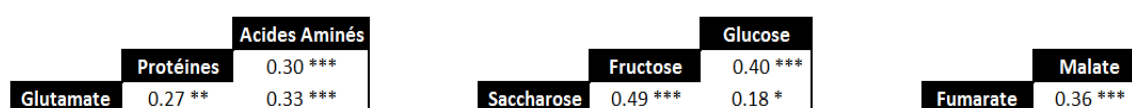


Figure B3.2| Coefficient de corrélation de Spearman entre les données métaboliques pouvant être considérées comme des indicateurs clés du métabolisme de l'azote et du carbone

Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ainsi, trois analyses en composantes principales (ACP) ont été réalisées avec **i)** les données de protéines, acides aminés et glutamate, **ii)** les données de glucose, fructose et saccharose et **iii)** les données de malate et fumarate. Les coordonnées des individus sur les deux premiers axes principaux ont été récupérées pour chacune de ces analyses ACP. Elles sont notées **i)** « Coord.Indiv.ACP1 » pour l'ACP réalisée sur les données du contenu en protéines, acides aminés et glutamate, **ii)** « Coord.Indiv.ACP2 » pour l'ACP réalisée sur les données du contenu en glucose, fructose et saccharose et **iii)** « Coord.Indiv.ACP3 » pour l'ACP réalisée sur les données du contenu en malate et fumarate. Ce sont ces données qui ont été utilisées dans des analyses de GWA. Ainsi, 19 régions du génome avec au moins un SNP avec une valeur de $-\log(p\text{-value}) > 4$ ont été identifiées sur les cinq chromosome d'*A. thaliana* (Tab. B3.4).

Tableau B3.4 | Nombre d'associations mises en évidence sur chaque chromosome à partir des coordonnées des individus obtenues sur 3 analyses ACP différentes

Variables Utilisées pour le GWA	Nombre d'associations				
	Chr1	Chr2	Chr3	Chr4	Chr5
Coord. Indiv. ACP 1	4	1	2	1	0
Coord. Indiv. ACP 2	0	1	0	1	1
Coord. Indiv. ACP3	2	2	2	1	1

Les coordonnées des individus ont été obtenues à partir des analyses ACP menées avec **i)** les concentrations en protéines, acides aminés et glutamate : Coord.Indiv.ACP1, **ii)** les concentrations en glucose, fructose et sucrose : Coord.Indiv.ACP2 et **iii)** les concentrations en malate et fumarate : Coord.Indiv.ACP3. Seules les zones génomiques pour lesquelles au moins un des SNP présente une valeur de p-value supérieure au seuil de $-\log(p\text{-value}) = 4$ ont été comptabilisées.

Parmi ces régions, aucune ne co-localise avec les 58 régions génomiques identifiées avec les analyses GWA menées avec chaque métabolite séparément. En revanche, il est intéressant de noter que deux régions sont retrouvées en commun avec les régions mises en évidence à partir des analyses GWA menées avec les données maladie et les métabolites (Tab. B3.3). Ainsi, il y a d'une part, une région au niveau du chromosome 1 entre 25.920.194 et 25.924.643 bp, mise en évidence par GWA avec les données obtenues avec l'ACP sur les données acides aminés, protéines et glutamate (Tab. B3.5a). Cette même région a été identifiée sur les analyses précédentes avec les données du contenu en acides aminés et les données de charge virale. D'autre part, il y a une région au niveau du chromosome 5 entre 20.142.679 et 20.154.614 bp mise en évidence par le GWA avec les données obtenues avec l'ACP sur les données glucose, fructose et saccharose (Tab. B3.5a). Cette région a aussi été identifiée dans l'analyse GWA sur les données de charge virale et du contenu en fructose. D'autres régions détectées avec des p-value relativement fortes ($-\log(p\text{-value}) = 5$) sont également mises en évidence sur ces différentes analyses GWA, les gènes retrouvés dans ces zones ont donc été notés (Tab. B3.5b).

Tableau B3.5 | Régions du génome identifiées par les analyses GWA à partir des coordonnées des individus sur les deux axes principaux des ACP regroupant les variables par voies métaboliques

Variables	Chr	Zone bp (Début..Fin)	Top SNP	-log(p-value) top SNP	Gènes	Description
(a)						
Coord. Indv. ACP 1	1	25.920.194..25.924.643	25.920.425	4.57	AT1G68930 Pentapeptide repeat containig protein AT1G68940 Armadillo/beta-catenin like repeat family protein - protein ubiquitination	
Coord.Indv.ACP2	5	20.142.679..20.154.614	20.154.146	6.19	AT5G49630 Amino acid permease 6 - response to water deprivation, to salt stress, aspartate transport , tryptophan transport AT5G49640 Unknown protein AT5G49650 Xylulose kinase 2 - xylulose metabolic process	
(b)						
Coord. Indv. ACP 1	2	8.448.903..8.460.847	8.448.924	6.06	AT2G19500 Cytokinin oxidase 2 AT2G19510 LOB domain containing protein 8 AT2G19520 FVE - metal ion binding - flower developement, unidimentional cell growth	
Coord.Indv.ACP2	2	1.934.087..1.947.139	1.947.139	5.71	AT2G05310 Unknown protein AT2G05320 beta 1.2 N acetylglucosaminyltransferase II - oligosaccharide biosynthetic process AT2G05330 BTB/POZ domain containing protein AT2G05335 SCR like 15 - signal transduction	
Coord.Indv.ACP3	1	2.002.978..2.020.444	2.002.978	5.49	AT1G06540 Unknown protein AT1G06550 ATP dependent caseinolytic protease/crotonase family protein AT1G06560 NOL1/NOP2/sun family protein AT1G06570 Phytoene desaturation 1 (PDS1) AT1G06580 Pentatricopeptide repeat superfamily protein AT1G06590 Unknown protein	
Coord.Indv.ACP3	2	8.338.678..8.339.277	8.339.277	5.19	AT2G19210 Leucine rich repeat transmembrane protein kinase	

(a) Régions génomiques identifiées en commun avec celles mises en évidence à partir des analyses GWA menées sur les données de charge virale et du contenu en métabolites primaires. **(b)** Régions génomiques identifiées avec de fortes p-value. Les analyses GWA ont été réalisées avec les coordonnées des individus obtenues à partir des analyses ACP menées avec **i)** les concentrations en protéines, acides aminés et glutamate : Coord.Indiv.ACP1, **ii)** les concentrations en glucose, fructose et saccharose : Coord.Indiv.ACP2 et **iii)** les concentrations en malate et fumarate : Coord.Indiv.ACP3.

Chr = N° de Chromosome ; Zone bp (Début..Fin) = Délimitation des zones génomiques identifiées dont les SNPs sont distants de maximum 10kb ; Top SNP = Position du meilleur SNP de la zone génomique identifiée ; -log(pvalue) top SNP = Valeur de la pvalue du meilleur SNP

2 Qu'en est-il des bases génétiques de la réponse métabolique d'*A. thaliana* en absence d'infection ?

Dans un premier temps, ce sont les données relatives aux échantillons inoculés par le TuMV qui ont été analysées puisque de fortes perturbations du métabolisme primaire avaient été mises en évidence chez *A. thaliana* en conditions d'infection. Les essais ayant été conduits de façon originale en conditions extérieures et ayant à disposition les données de concentrations en métabolites primaires pour les échantillons mock-inoculés, une analyse GWA a été effectuée afin d'essayer de caractériser l'architecture génétique des métabolites primaires analysés dans ces conditions extérieures.

Dans un premier temps, une estimation de l'héritabilité (H^2) a été effectuée pour chacun des dix traits métaboliques analysés (Tab. B3.6). La valeur moyenne d' H^2 obtenue sur ces échantillons mock-inoculés est de 0.44 avec des valeurs allant de 0.21 à 0.71. Le coefficient de variation de 32.4% calculé entre ces H^2 est similaire à celui évalué sur l'ensemble des H^2 calculées chez les échantillons inoculés par le TuMV.

Tableau B3.6 | Héritabilités (H^2) pour chacun des traits métaboliques mesurés en absence d'infection

Variable Métabolique	H^2
Acides Aminés	0.27
Protéines	0.21
Glutamate	0.31
Malate	0.54
Fumarate	0.41
Amidon	0.52
Glucose	0.71
Fructose	0.52
Saccharose	0.39
Chlorophylle a	0.48

Les analyses GWA ont donc été menées à partir des données sur les concentrations de ces dix métabolites primaires mesurées chez les échantillons mock-inoculés. Comme cela a déjà été observé sur les analyses GWA des échantillons inoculés par le TuMV, les valeurs des p-value obtenues restent globalement faibles, en moyenne autour d'une valeur de $-\log(p\text{-value}) = 4$ (Fig. B3.A2). Ainsi, 117 zones ont été identifiées dont au moins un SNP présente une valeur de $-\log(p\text{-value})$ au-dessus du seuil de significativité. Près de 70% de ces loci ont été identifiés avec une seule variable métabolique puisque seules 35 de ces zones sont retrouvées avec au moins deux traits différents. Les analyses GWA menées avec les teneurs en amidon, fructose et saccharose sont celles qui permettent d'identifier le plus grand nombre de régions (Tab. B3.7).

Tableau B3.7 | Zones du génome identifiées sur chaque chromosome et pour chaque trait métabolique chez échantillons mock-inoculés

Variable Métabolique	Nbre d'associations				
	Chr1	Chr2	Chr3	Chr4	Chr5
Acides Aminés	5	2	0	3	5
Protéines	2	2	2	4	3
Glutamate	2	1	3	3	2
Malate	3	1	2	2	4
Fumarate	2	2	1	4	0
Amidon	3	0	2	6	11
Glucose	2	3	0	4	2
Fructose	3	0	3	14	17
Saccharose	11	2	1	4	4
Chlorophylle a	2	6	2	3	1

Les données de concentrations en métabolites primaires ont été obtenues suite à des dosages sur des échantillons mock-inoculés et récoltés 13 jours après inoculation. Seules les zones génomiques pour lesquelles au moins un des SNP présente une valeur de p-value supérieure au seuil de $-\log(p\text{-value}) = 4$ ont été comptabilisées.

Des analyses ont également été effectuées en suivant la même démarche que précédemment avec les données relatives aux coordonnées des individus sur les deux axes principaux d'analyses en composantes principales. Les loci identifiés par ces analyses ne co-localisent avec aucune des 117 régions identifiées par les analyses GWA sur chaque métabolite. De plus, par rapport aux résultats obtenus avec les données relatives aux échantillons inoculés avec le TuMV, aucun locus commun n'a été identifié (données non montrées). Cela permet de montrer et de confirmer qu'avec les données métaboliques des échantillons inoculés avec le TuMV, ce sont bien les bases génétiques des variations métabolique en lien avec l'infection qui sont analysées. Avec les échantillons mock-inoculés, c'est donc la structure génétique des variations métaboliques dans les conditions extérieures qui est analysée. Cependant, l'annotation des gènes dans les régions génomiques identifiées par les différentes analyses GWA n'a pas permis de faire le lien de façon claire entre les données métaboliques testées et des gènes impliqués dans les voies métaboliques en lien avec ces données.

Discussion

Le phénotypage large de la réponse d'*A. thaliana* au TuMV en conditions extérieures a permis de s'intéresser de façon assez nouvelle dans le domaine de la virologie végétale aux perturbations du métabolisme primaire et secondaire ayant lieu lors d'une infection. Pour aller plus loin dans la compréhension de cette réponse métabolique, nous avons cherché à caractériser ses bases génétiques via des analyses GWA. Dans un premier temps, ces analyses ont été menées sur les échantillons inoculés par le TuMV. En effet, c'est en condition d'infection que des perturbations métaboliques fortes ont été mises en évidence (Rubio *et al.*, 2017 – *in preparation*).

Dans ces conditions d'infection, les données de concentration en protéines et fumarate n'ont pas été exploitées dans l'analyse GWA puisque les héritabilités étaient trop faibles. Ce qui n'était d'ailleurs pas le cas avec les données obtenues sur les échantillons mock-inoculés. Une explication pourrait être en lien avec le phénotypage de ces variables qui aurait pu créer une variabilité entre échantillons supérieure à la variabilité liée au génotype. Cependant, les dosages métaboliques ont été effectués sur l'ensemble des échantillons (inoculés par le virus et mock-inoculés) en même temps. Si un problème de phénotypage existait, il aurait potentiellement concerné aussi les échantillons mock-inoculés. En ce qui concerne les concentrations en protéines, cette faible héritabilité en conditions d'infection pourrait s'expliquer par le fait que dans ces conditions, les protéines de la capsidie donc les protéines virales pourraient être prises en compte dans les mesures du dosage des protéines. La variation observée ne serait donc pas en lien avec la réponse des génotypes au virus mais plutôt à la capacité multiplicative du virus. Une autre hypothèse serait de dire que ces deux métabolites en condition de stress biotiques varient de façon indépendante du génotype. Chaque échantillon accumule plus au moins de ces métabolites et ce indépendamment de son fond génétique. En ce sens, pour Fusari et ses collaborateurs (2017), les faibles héritabilités peuvent être dues à la complexité de la variable, à la variation environnementale et/ou à une interaction génotype X environnement forte (Fusari *et al.*, 2017).

Parmi les différentes zones du génome obtenues, certaines ont été identifiées à partir de plusieurs traits métaboliques en même temps. Cette co-localisation a déjà été observée sur d'autres analyses GWA avec des données métaboliques chez la tomate, *A. thaliana* et le riz (Chan *et al.*, 2010; Zhao *et al.*, 2011; Sauvage *et al.*, 2014). Cette co-localisation pourrait s'expliquer soit par **i)** le fait que les métabolites appartiennent à une même voie métabolique soit **ii)** les métabolites appartiennent à des voies métaboliques différentes qui seraient contrôlées par des mécanismes communs soit **iii)** il peut s'agir de zones composées de plusieurs gènes liés avec des fonctions dans différentes voies métaboliques (Lisec *et al.*, 2007).

Par rapport aux analyses GWA menées au préalable sur les traits phénotypiques en lien directement avec la maladie (charge virale, symptômes, fréquences d'infection), une seule région identifiée à partir des données de concentration en acides aminés est retrouvée en commun. Il s'agit d'un locus au niveau du chromosome 3 où se trouve le gène *RTM3* déjà largement caractérisé comme impliqué dans la réponse d'*A. thaliana* aux potyvirus (Whitham *et al.*, 1999; Decroocq *et al.*, 2006; Cosson *et al.*, 2012). Le fait de retrouver cette même région avec les données d'acides aminés permettrait de dire que la sensibilité des échantillons au

TuMV pourrait être caractérisée sur la base des concentrations en acides aminés. Le contenu en acides aminés serait donc caractéristique de l'infection, les acides aminés pourraient être considérés comme des marqueurs diagnostic de l'infection par le TuMV en conditions extérieures.

Une autre façon de traiter ces données a été de mettre en place des analyses GWA à partir des coordonnées des individus obtenues suite à des analyses en composantes principales réalisées à partir des variables métaboliques significativement corrélées entre elles et qui pouvaient donc être considérées comme des indicateurs clés du métabolisme de l'azote et du carbone (Zhang *et al.*, 2015). Cette analyse a permis de mettre en évidence deux régions qui co-localisent avec les locus mis en évidence sur les analyses GWA regroupant les données métaboliques et les données maladie. Les recherches bibliographiques sur les gènes annotés dans ces deux régions n'ont cependant pas permis de montrer leur éventuelle implication dans des mécanismes de réponse des plantes à des virus ou de façon plus large à des agents pathogènes.

Des analyses GWA ont également été conduites à partir des données obtenues sur les échantillons mock-inoculés. Les régions détectées sont différentes de celles observées sur les échantillons inoculés par le TuMV. Cette différence permet de penser que les zones détectées dans les conditions d'infection virale sont bien en relation avec le virus. En revanche, comme pour les analyses sur les échantillons inoculés par le TuMV, les p-value obtenues sont globalement faibles. De plus, l'annotation des gènes n'a pas permis d'identifier de façon claire des gènes décrits comme appartenant à une voie métabolique en particulier.

De façon globale, il semble donc difficile de trouver des arguments permettant de justifier le choix de tel ou tel gène candidat dans la caractérisation de la réponse métabolique d'*A. thaliana* aux conditions extérieures et/ou à l'infection par le TuMV.

Les difficultés à caractériser les bases génétiques de la réponse métabolique dans cette étude peuvent finalement ne pas être si surprenantes. En effet, d'autres études ont déjà montré que le lien entre génotype et traits métaboliques n'est pas si simple à décrire car ces variables sont sous l'influence de loci quantitatifs complexes (Kliebenstein, 2009). Les niveaux de concentration en métabolites étant sous le contrôle de différents mécanismes (voies enzymatiques, régulations transcriptionnelles et post-transcriptionnelles) (Matsuda *et al.*, 2012).

Des variations dans les héritabilités calculées existent entre les métabolites étudiés. Ces variations ont déjà été décrites dans d'autres études (Chan *et al.*, 2010; Sauvage *et al.*, 2014; Matros *et al.*, 2016). Ces variations pourraient s'expliquer par le niveau auquel le phénotypage est analysé (transcriptionnel, métabolique, physiologique). En effet, des mécanismes de régulation plus complexes concernent d'avantage le niveau métabolique que le niveau transcriptionnel. Ainsi, cette complexité va avoir tendance à diminuer l'héritabilité estimée (Kliebenstein, 2009). Le fait que ces variables métaboliques présentent des valeurs d'héritabilité parfois faibles, peut expliquer la difficulté à identifier des loci de la réponse métabolique. Il a d'ailleurs été mis en évidence que la détection de mQTL (QTL métabolique) est fortement dépendante de l'héritabilité des traits analysés. Cela a déjà été décrit chez *A.*

thaliana mais également chez le riz et la tomate (Lisec *et al.*, 2007; Schauer *et al.*, 2008; Matsuda *et al.*, 2012). De plus, il semble admis que de façon générale les métabolites secondaires sont contrôlés par un nombre réduit de loci à large effet. Au contraire, les variations du contenu en métabolites primaires seraient contrôlées par de nombreux loci à faible effet (Kliebenstein, 2009; Chan *et al.*, 2010; Luo, 2015).

La complexité du phénotypage métabolique est également soulignée dans une étude d'association chez la tomate, menée sur deux années différentes, et pour laquelle seuls 47.3% des traits métaboliques testés sont stables sur les deux années (Sauvage *et al.*, 2014). Néanmoins, Sauvage et ses collaborateurs (2014) identifient 44 loci en lien avec des traits métaboliques mesurés dans des fruits de tomate. Parmi ces loci, certains présentent des gènes candidats dont le lien avec les variables métaboliques étudiées a déjà été décrit. De plus, de nouveaux potentiels gènes candidats ont également été identifiés. La force de cette étude réside dans **i)** la diversité génétique des 163 accessions de *Solanum lycopersium* choisies comprenant des accessions sauvages et cultivées, **ii)** les deux essais mis en place sur deux années différentes et **iii)** le panel de 76 variables métaboliques analysé sachant que seulement la moitié de ces traits sont stables sur les deux années et que les 44 loci identifiés sont en lien avec la variation de 19 de ces traits métaboliques (Sauvage *et al.*, 2014). Une autre étude récente a été menée chez *A. thaliana* dans le but de caractériser les bases génétiques de métabolisme primaire central et de la croissance (Fusari *et al.*, 2017). Cette étude a été conduite sur deux années consécutives. Des analyses GWA pour 9 variables métaboliques et 24 variables enzymatiques de chaque essai ont été effectuées de façon indépendante sur 349 accessions. La sélection des pics d'association retenus a notamment été basée sur ceux retrouvés en commun sur les deux analyses. Ainsi, à partir des 70 loci identifiés, une liste de 116 gènes candidats a été constituée.

Le contrôle génétique de ce contenu métabolique peut être affecté par les conditions extérieures (Chan *et al.*, 2010). De ce fait, dans le cadre de cette étude, il aurait été pertinent d'étudier cette réponse génétique sur une deuxième année d'expérimentation. Cela aurait sans doute permis de voir certains effets dus à l'environnement. L'identification potentielle de loci communs entre les deux essais aurait pu être considérée comme particulièrement intéressante pour décrire l'architecture génétique de la réponse d'*A. thaliana* lors d'une infection par le TuMV. Cela aurait été d'autant plus intéressant que les p-value obtenues suite aux analyses GWA sont relativement faibles, ce qui ne permet pas de donner plus d'importance à un locus plus qu'à un autre. C'est d'ailleurs le fait de disposer de deux années d'études qui a permis de mettre l'accent sur certains des résultats obtenus dans le cadre des analyses GWA sur la population mondiale avec les traits liés à la maladie mais également lors de l'étude de la réponse métabolique d'*A. thaliana* au TuMV. La mise en place de plusieurs expérimentations pour étudier les bases génétiques du métabolome se retrouve sur d'autres études menées chez *A. thaliana* mais également chez le riz ou encore chez la tomate (Chan *et al.*, 2010; Sauvage *et al.*, 2014; Zhang *et al.*, 2015).

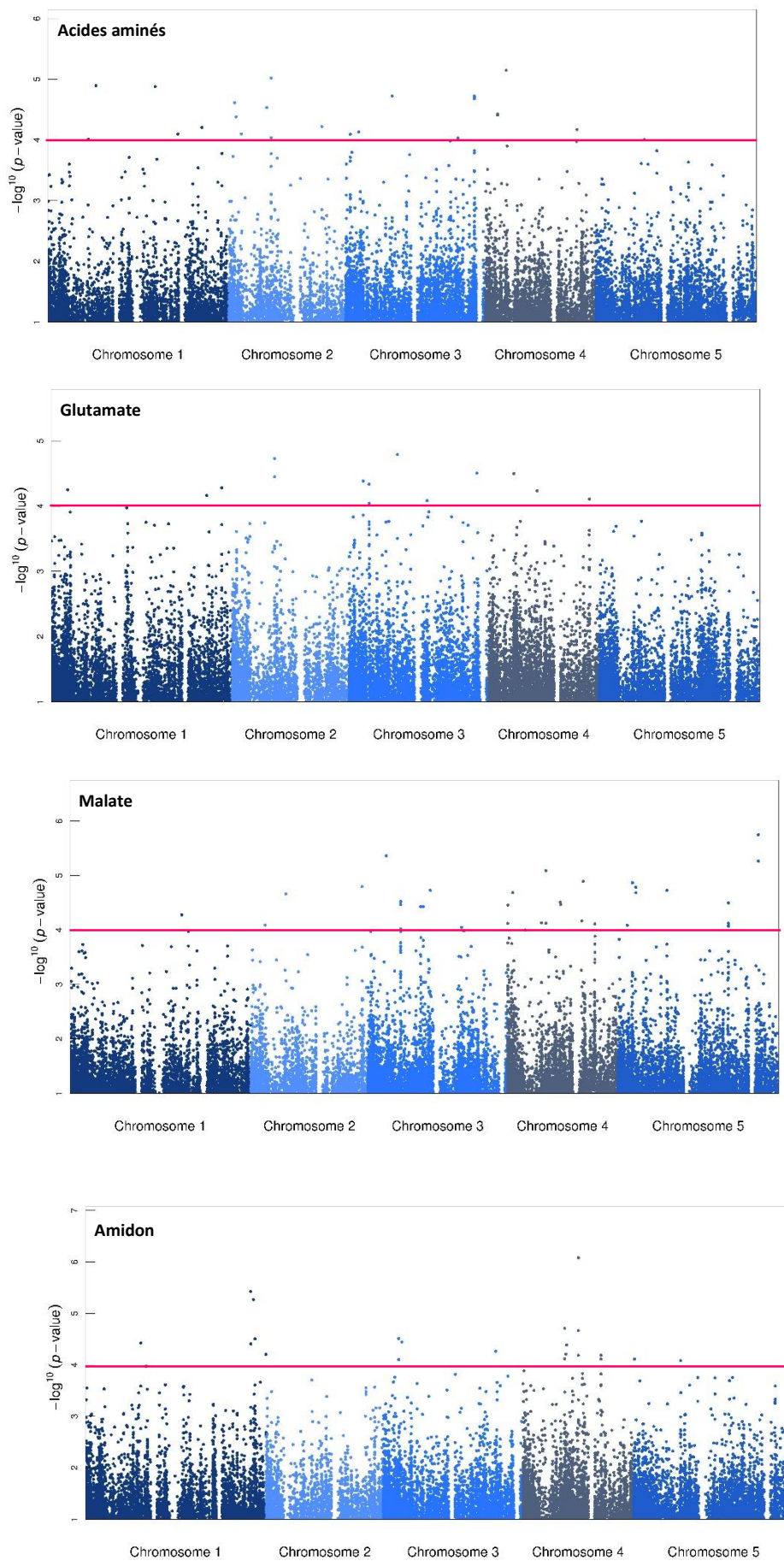
De façon originale, dans leur étude de régulateurs du métabolisme primaire chez *A. thaliana*, Wu et ses collaborateurs (2016) ont combiné les résultats d'analyses GWA avec **i)** des données issues d'analyses de réseaux de corrélation entre métabolites et transcrits ainsi que **ii)** des

données concernant des mQTL détectés à partir d'analyses menées sur des populations de RIL (Recombinant Inbred Lines) et IL (Introgression Lines) d'*A. thaliana*. C'est la combinaison de ces stratégies qui a facilité l'identification de régulateurs du métabolisme primaire puisqu'ils ont réussi à identifier 92 associations significatives entre des gènes et le contenu en certains des métabolites primaires analysés (Wu *et al.*, 2016).

Au vu de ces différentes études et de leurs caractéristiques, un certain nombre de limites peuvent être identifiées dans le cadre de notre analyse en lien notamment avec **i)** la variabilité génétique : 130 accessions ont été analysées soit quasiment trois fois moins par rapport à l'étude de Fusari et ses collaborateurs, **ii)** le nombre de variables métaboliques mesurées qui s'est limité à 10 métabolites primaires or généralement les études comprennent des dizaines de variables, **iii)** le fait de ne disposer que des données obtenues sur une seule année. Si des dosages métaboliques ont été réalisés sur une deuxième année, seules 29 accessions ont été étudiées. Tout cela n'empêche pas de mettre en avant la démarche de cette étude qui n'en reste pas moins intéressante et originale. Il faut également tenir compte du fait qu'au moment où ces essais ont été mis en place, l'idée d'utiliser les données des concentrations métaboliques dans des analyses GWA n'était pas l'objectif principal. D'autres choix en termes de nombre d'accessions, de répétitions biologiques, de nombre et types de variables métaboliques auraient pu être effectués dans le but de maximiser et d'optimiser les analyses GWA (Soltis & Kliebenstein, 2015).

Annexes

Figure B3.A1 Manhatten plots correspondants aux différentes analyses GWA à partir des concentrations en métabolites primaires des échantillons inoculés avec le TuMV



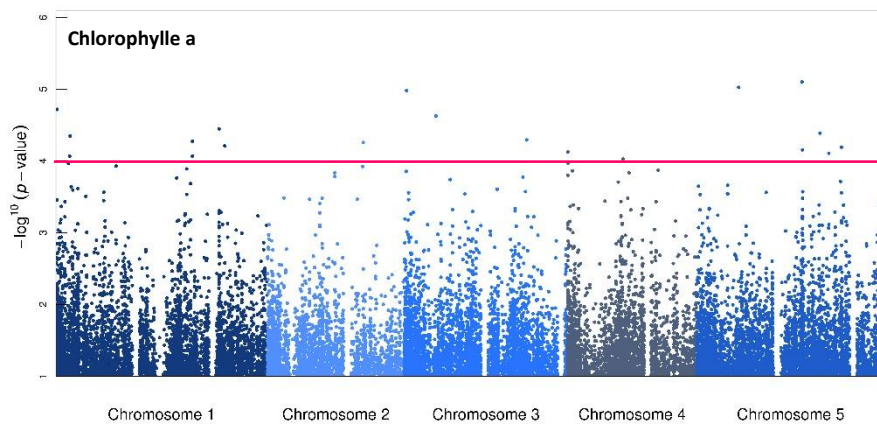
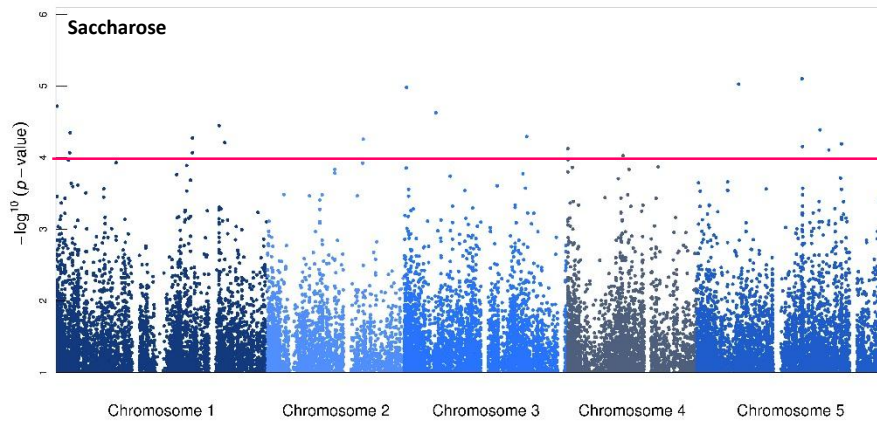
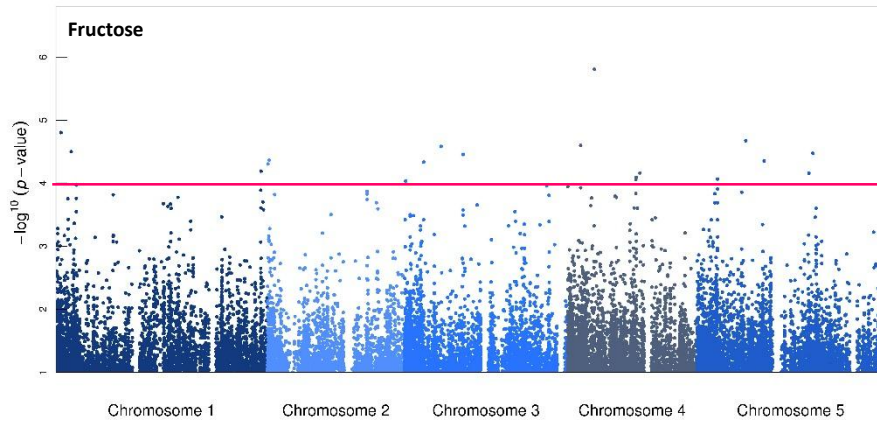
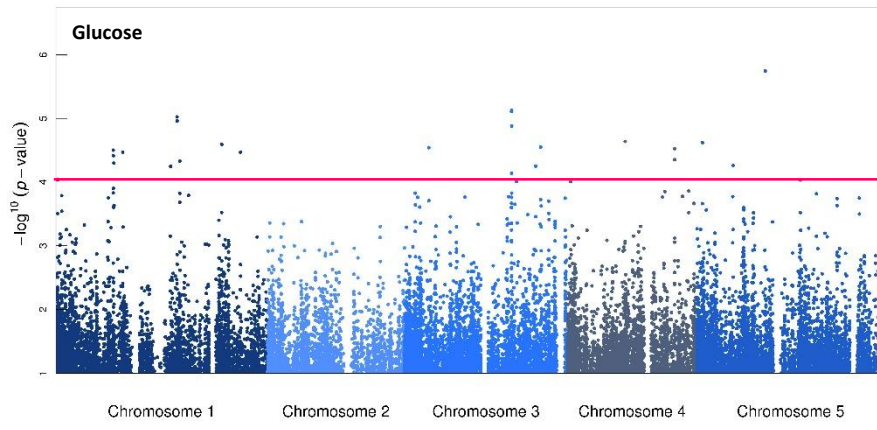
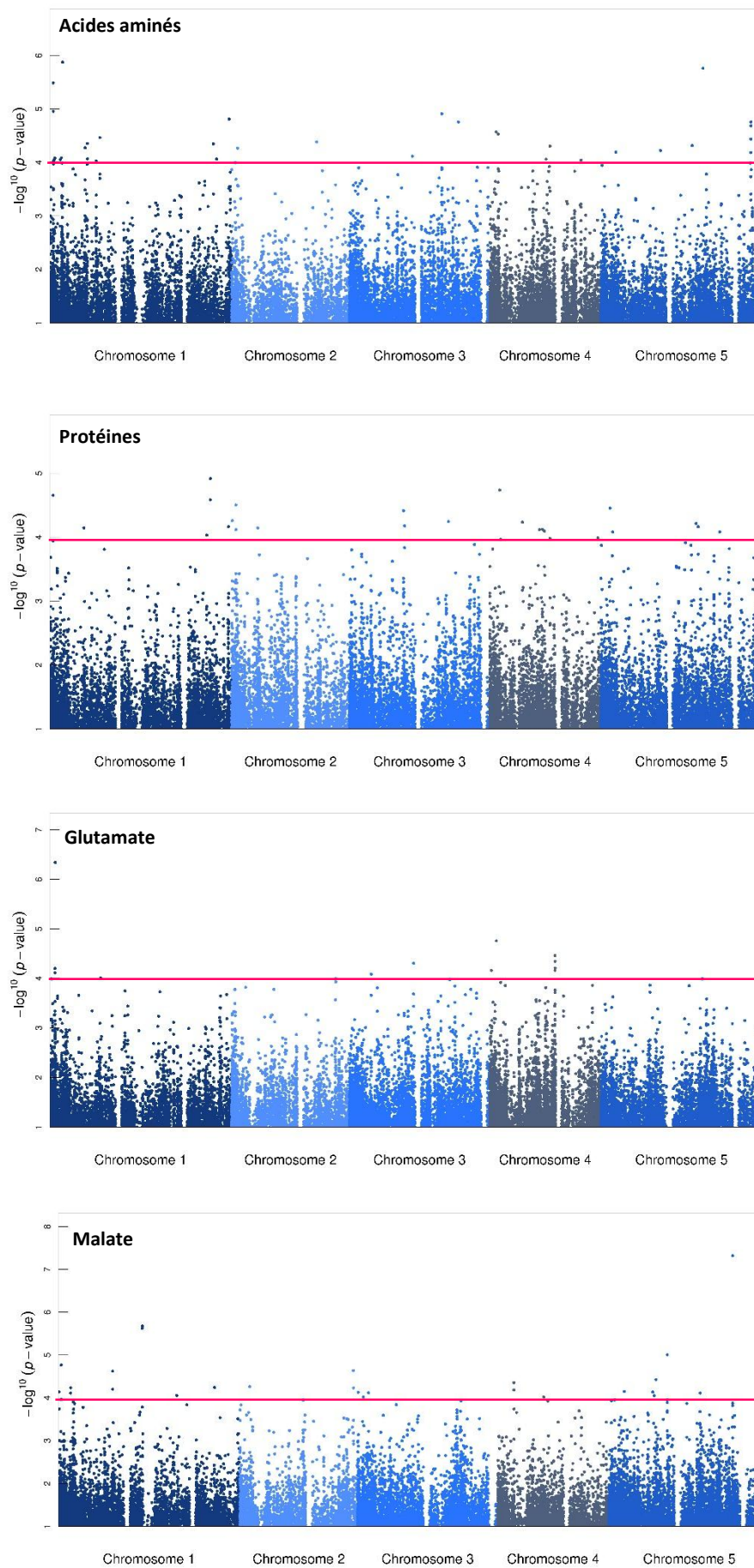
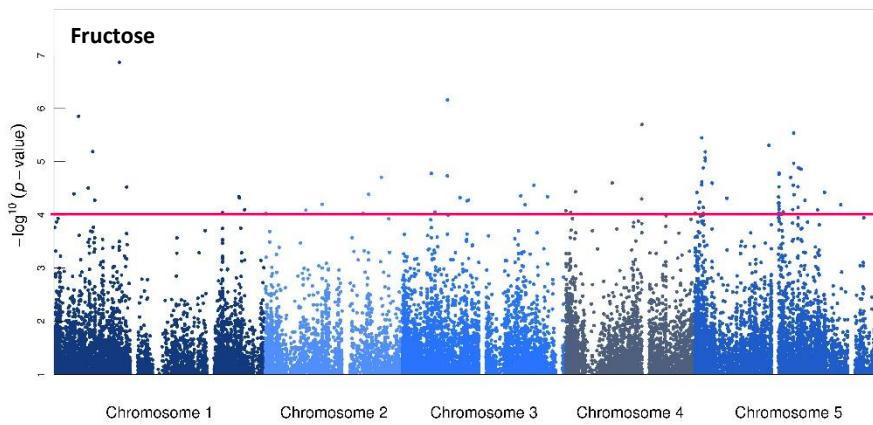
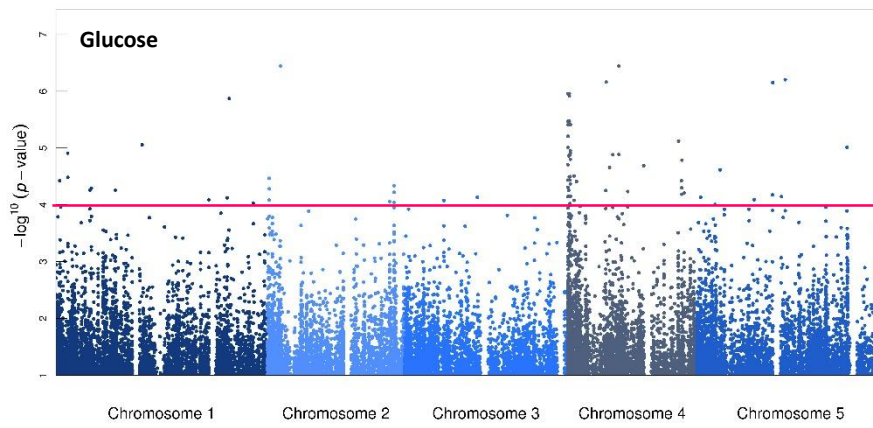
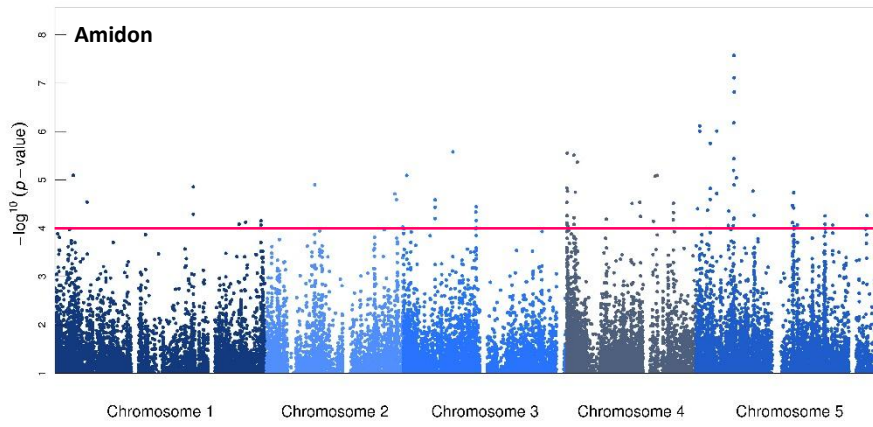
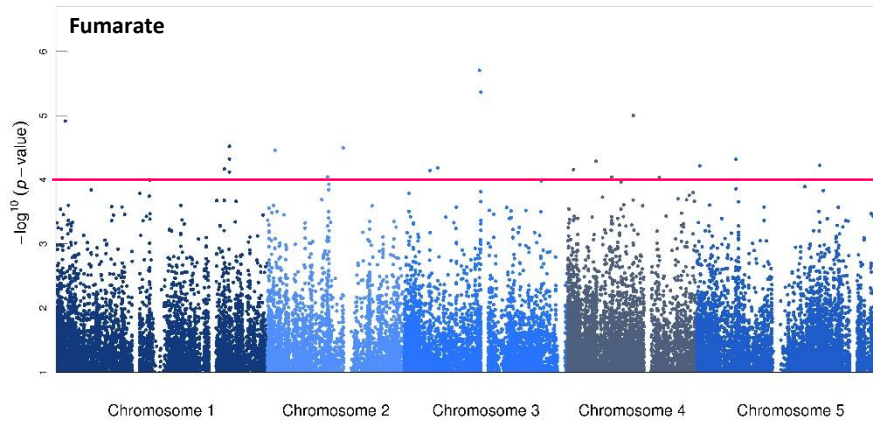
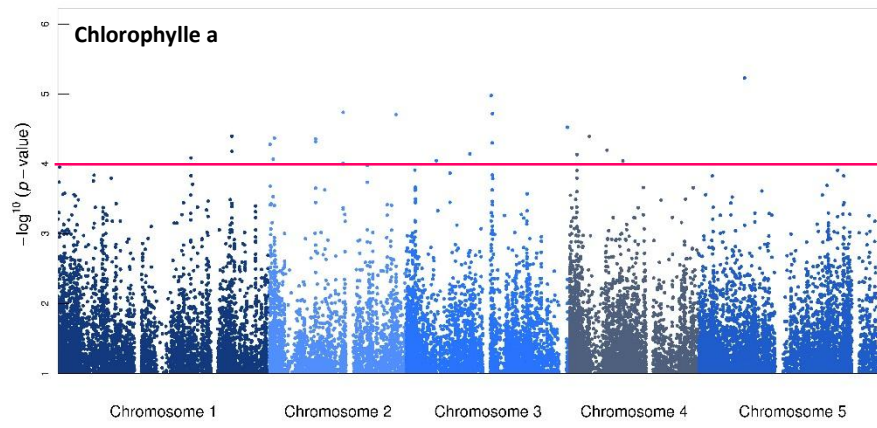
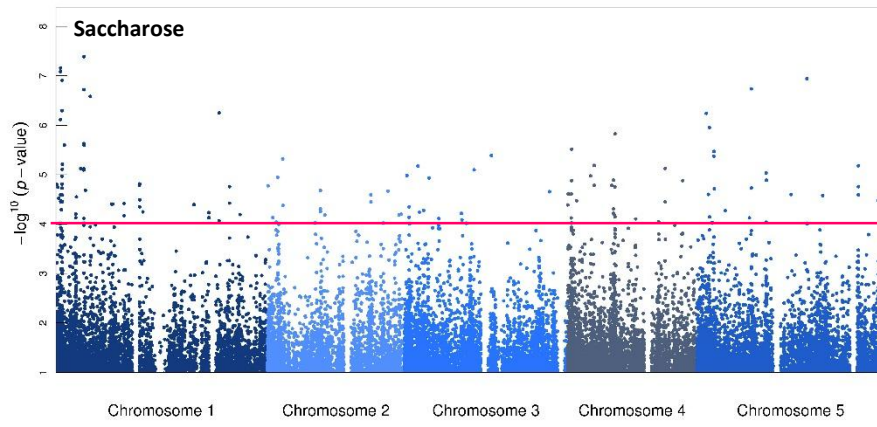


Figure B3.A2 Manhattan plots correspondants aux différentes analyses GWA à partir des concentrations en métabolites primaires des échantillons mock-inoculés







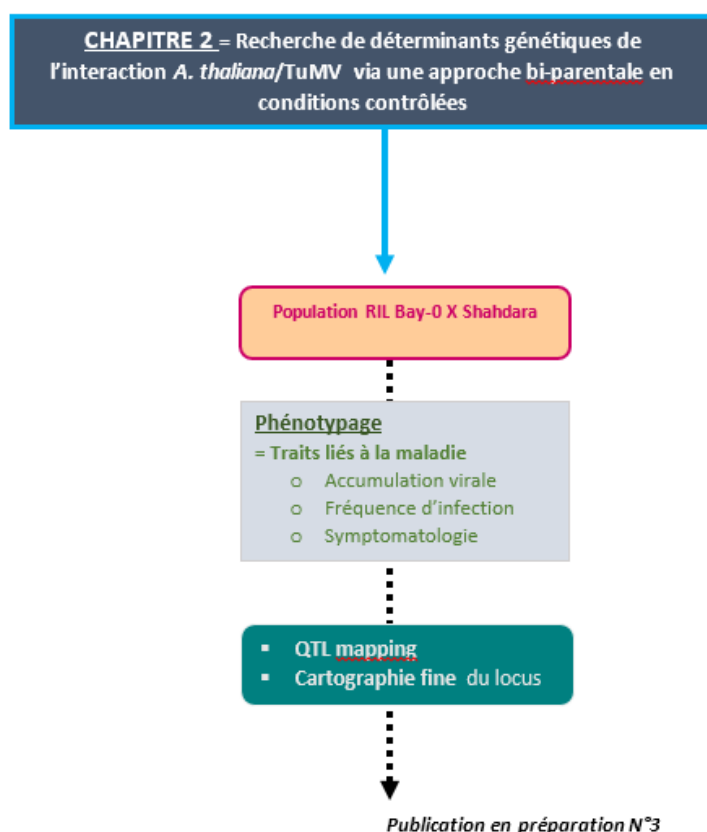
- Chapitre 2 -

**Recherche de déterminants génétiques de
l'interaction *A. thaliana*/TuMV via une approche
bi-parentale en conditions contrôlées**

Présentation Générale du Chapitre 2

L'identification de nouvelles sources de résistances est toujours un challenge majeur pour lutter contre les maladies virales. Les résultats présentés au cours du deuxième chapitre de cette thèse ont permis notamment de caractériser les bases génétiques de la réponse d'*A. thaliana* à une infection par le TuMV en conditions multistress et en utilisant des stratégies de génétique d'association (GWA). Au cours des analyses en « common garden » et d'analyses en conditions contrôlées (données non montrées) et comme cela avait déjà été montré par Martin Martin et ses collaborateurs (1999), l'accession Bay-0 présente une résistance stable et relativement peu dépendante des conditions environnementales. *A. thaliana* étant une plante modèle, de nombreuses ressources génétiques sont disponibles. Ainsi, une population de RIL (Recombinant Inbred Line) issue du croisement entre Bay-0 (accession résistance) et Shahdara (accession fortement sensible dans nos conditions) était disponible (Loudet *et al.*, 2002). Afin de mettre en évidence de nouveaux déterminants génétiques de la réponse des plantes au virus, une étude de « QTL mapping » a été conduite à partir de cette population RIL en conditions contrôlées.

L'ensemble de ce travail est présenté sous la forme d'un article scientifique.



Identification of new candidate genes involved in the natural pathosystem interaction *Arabidopsis thaliana* and Turnip mosaic virus (TuMV).

Bernadette Rubio¹, Patrick Cosson¹, Mélodie Caballero¹, Thierry Michon¹, Valérie Schurdi-Levraud¹ *

¹INRA – Université de Bordeaux, UMR 1332 Biologie du Fruit et Pathologie, Villenave D'Ornon, France

* Corresponding author

Mailing address : INRA, UMR 1332 Biologie du Fruit et Pathologie

Phone : +33 (0)5 57 12 23 77

E-mail : valerie.schurdi-levraud@inra.fr

Keywords: *Arabidopsis thaliana*, *Turnip mosaic virus*, Recombinant Inbred Line (RIL), QTL analysis, Heterozygous Inbred Family (HIF), fine mapping

Introduction

Compared to other plant pathogens, viruses are particularly unpredictable and difficult to treat. The genus Potyvirus is considered as one of the largest among plant viruses, causing considerable economic damage in crop worldwide. In view of this major economic problem, the development and use of cultivars that are genetically resistant to viruses still remains a critical factor of competitiveness for both breeders and producers and one of the key for sustainable agriculture. Thus, considering the high adaptive potential of plant viruses and resulting resistance breakdown, an extensive screen of host-virus interactors remains needed.

Arabidopsis thaliana (*A. thaliana*) is not a crop but it is an ideal host for functional strategies involved in plant responses to pathogens. Studies on disease resistances has shown that there are significant conservation in the factors that mediate defense between *A. thaliana* and cultivated crops (Carr & Whitham, 2007). Approximately 40 viral species belonging to 18 genera have been shown to infect at least one *A. thaliana* accession under controlled conditions (Ouibrahim & Caranta, 2013). Working on this model plant, gives access to many resources for understanding the genetic, molecular and the biochemical foundations of plant/virus interactions. This has contributed to great advances in understanding the molecular dialog between viral pathogens and their plant hosts. For example, the *TOBAMO-VIRUS MULTIPLICATION 1* (*TOM1*) and *TOM3* genes, which play an essential role in the replication of *Tobacco mosaic virus* (TMV), were isolated in a screen for *A. thaliana* mutants defective for infection by TMV and two homologous genes, *NtTOM1* and *NtTOM3*, have been identified in *Nicotiana tabacum* (Yamanaka *et al.*, 2002; Asano *et al.*, 2005).

Different strategies have been described in *A. thaliana* to counteract viral infections. RNA interference (RNAi) mediated-resistance has been recognized as an evolutionary conserved process in eukaryotes triggered by double-stranded RNA. Key components of RNAi have been shown to have an important protective role against viral pathogens (Pumplin & Voinnet, 2013). Besides RNAi, viruses may also have to deal with R-mediated resistances. Two dominant resistant genes, *HRT* (*HYPERSENSITIVE RESPONSE TO TCV*) and *RCY1* (*RESISTANT TO CMV*), which belong to the NB-LRR family have been identified in *A. thaliana*/*Turnip crinkle virus* (TCV) and *A. thaliana*/*Cucumber mosaic virus* (CMV) interactions. *HRT* confers systemic resistance to TCV in a SA-dependent manner and *RCY1* systemic resistance to CMV is partially dependent on SA but also involves ethylene signaling (Cooley *et al.*, 2000; Takahashi *et al.*, 2002). A third NB-LRR gene, *TTR1* (*TOLERANCE TO TRSV 1*) was identified in *A. thaliana*/*Tobacco ringspot virus* (TRSV) (Nam *et al.*, 2011). Other dominant resistance to viruses have also been discovered in the last years. This is the case, for example, of the three dominant genes *RTM1*, *RTM2* and *RTM3* identified on Col-0 accession (Mahajan *et al.*, 1998; Whitham *et al.*, 1999). *RTM1* encodes a lectin belonging to a large family of sugar-binding proteins (Chisholm *et al.*, 2000). *RTM2* encodes a protein with an N-terminal region similar to small plant heat shock proteins (Whitham *et al.*, 2000). *RTM3* encodes a MATH (mepirin and TRAF homology) protein with a coiled-coil domain on C-terminal (Cosson *et al.*, 2010). A mutation in any of these genes results in completely abolish the restriction of the long-distance movement of potyviruses. Among all these responses to viruses, monogenic resistances seem to be overrepresented in plant virus/interactions. Recessive resistances are

the consequence of a mutation or loss of a host factor recruited normally by the virus to carry out its life cycle. A large number of natural recessive resistance genes correspond to eukaryotic translation initiation factor eIF4E and eIF4G and their isoforms. Several recessive resistance loci against viruses have been identified in *A. thaliana* like *rlm1* (*resistance to lettuce mosaic virus 1*) or *rpv1* (*resistance to plum pox virus 1*) for resistance to *Lettuce mosaic virus* (LMV) and *Plum pox virus* (PPV) respectively (Revers *et al.*, 2003; Decroocq *et al.*, 2006). However, most of these recessive genes have not yet been cloned. Recently, new candidates involved in recessive resistance have been discovered in *A. thaliana*/*Watermelon mosaic virus* (WMV) interaction. A chloroplastic phosphoglycetate kinase (cPGK2) has been proved to be the *rwm1*-mediated resistance (Ouibrahim *et al.*, 2014). This also could be a potential candidate gene in the response of *A. thaliana* to PPV (Poque *et al.*, 2015).

Despite the fact that *A. thaliana* can be the host of numerous plant viruses, there is a lack of studies that have identified viruses naturally infecting *A. thaliana*. In their study, Pagan and collaborators have characterized five viruses, *Cucumber mosaic virus* (CMV), *Turnip mosaic virus* (TuMV), *Turnip yellow mosaic virus* (TYMV), *Turnip crinkle virus* (TCV) and *Cauliflower mosaic virus* (CaMV) that have been shown to infect six different *A. thaliana* natural populations in Spain, potentially leading to plant/virus coevolution (Pagan *et al.*, 2010). A lots of *A. thaliana*/virus studies have been conducted on model pathosystems, challenging *A. thaliana* with either modified lab strain viruses or with viruses for which information on possible coevolution is not known. Thus, in a recent review concerning the resistance to viruses in *A. thaliana*, only three out of twenty-one studies reviewed, *A. thaliana* was challenged with a virus that had been shown to naturally infect it (Ouibrahim & Caranta, 2013).

Turnip mosaic virus (TuMV), a member of the genus *Potyvirus*, is one of the most important viruses in the world that infect field-grown vegetables. The host range of TuMV covers at least 318 species found within over 43 dicotyledonous families, including *Brassicaceae*, *Asteraceae*, *Chenopodiaceae*, *Fabaceae* and *Caryophyllaceae* (Walsh & Jenner, 2002). Among, *Brassicaceae*, *A. thaliana* is a natural host of TuMV (Pagan *et al.*, 2010). A first exploration of the response of *A. thaliana* to TuMV has been realized in 1999 by Martin Martin and collaborators. They had studied the response of 160 accessions. Among these accessions, four of them- Bay-0, Di-0, Er-0 and Or-0- were described as resistant through OD assessment. Our experiments conducted in controlled conditions (data not shown) and under field conditions during several years (Rubio *et al.*, 2017, *submitted*) on more than 300 accessions confirmed the stability of Bay-0 resistance in different environments.

In order to explore the genetic basis of this resistance, we developed a QTL mapping strategy using a Recombinant Inbred Line (RIL) population available from the cross between the resistant Bay-0 accession and the susceptible Shahdara accession (Loudet *et al.*, 2002). Phenotype analysis and fine linkage mapping reveal the existence of a major recessive locus on chromosome 1. None of the putative eighteen candidate genes belonging to this locus have never been described before as involved in *A. thaliana* and TuMV/Potyvirus interaction. They constitute potentially very interesting new candidate genes in *A. thaliana*/*Potyvirus* interaction.

Material and Methods

Plant Material and growth conditions

An *Arabidopsis* F2 population Bay-0 X Shahdara and a population of Recombinant Inbred Lines (RILs) also derived from a cross between Bay-0 (6899) and Shahdara (6962) was used (population 33RV on the Versailles *Arabidopsis* Stock Center - <http://publiclines.versailles.inra.fr/page/33>). The RIL population (Loudet *et al.*, 2002) is composed by 165 lines (F8 seeds) that have been genotyped with 69 markers. To confirm the locus, one HIF (68) was used. HIFs are nearly-isogenic lines selected in the progeny of RILs that show a single residual heterozygous region (Supplemental Table S1). RILs were grown in a climatic chamber under temperatures and humidity controlled conditions (9h day, 20°C and relative humidity of 60%). F2 and HIF populations were tested in a greenhouse.

Viral materials and inoculations

A. thaliana plants were inoculated with *Turnip mosaic virus* isolate UK1 (Jenner & Walsh, 1996) using two different methods: **i)** mechanically using an inoculum prepared by homogenizing young symptomatic leaves of 5 weeks old turnip with three volumes of sodium dihydrogen phosphate dodecahydrate (Na₂HPO₄.12H₂O) 30mM and 0.2% of dietyldithiocarbamic acid (DIECA). After each inoculation, plants were rinsed with water and **ii)** by agroinfiltration with a pCambiaTunosTuMVGFP clone introduced in *Agrobacterium tumefaciens* (Beauchemin *et al.*, 2005). The agroinfiltration procedure was carried out starting from *A. tumefaciens* C58C1 cultures (OD₆₀₀ = 0.8) induced with acetosyringone. *Agrobacterium* cells transformed with the clone were applied with the tip of a syringe pressed against the underside of four leaves per rosette. Both methods of inoculation are applied on *A. thaliana* 4 weeks-old plants at 8-10 leaves corresponding to 1.09 - 1.10 Boyes stages (Boyes *et al.*, 2001).

Phenotyping of the parental accessions

A scoring of the symptoms visually observed on the parental accessions was realized from 7 to 15 days post inoculation (dpi). The area under the disease progress curve (AUDPC) scores were calculated (Simko & Piepho, 2012).

The response to TuMV infection of the two parental *A. thaliana* Bay-0 and Shahdara accessions was also characterized by quantitative real-time PCR. Sampling was realized at 1-3-5-7-9 dpi and at 1-3-5-7-9-15 dpi on inoculated and systemic leaves respectively. Three biological replicates measurements were made for each accession. Each replicate correspond to leaves collected from six distinct plants of the same accession. After total RNA extraction realized with the NucleoSpin RNA kit (Macherey-Nagel), all the samples were adjusted to 50 ng/μL. Reverse transcription was done using Revertaid/Ribolock reverse transcriptase kit (ThermoFisher Scientific). Quantitative real-time PCR was performed on a Light Cycler 480 II machine (Roche Diagnostic) by using Light Cycler 480 SYBR Green I master mix and one tenth of the synthesized cDNAs. TuMV was amplified and detected by forward (5'-TCAAACCATGTGTCAAACCTGC^{3'}) and reverse (5'-CGAGAAAAGAGTGGCCCTAA^{3'}) primers in the capsid gene. Relative viral accumulations are expressed in ng/μL. They were estimated from a standard range from 0.001 ng/μL to 1ng/μL consistent of different dilutions of TuMV UK1 plasmid.

Agroinfiltrations were performed with the pCambiaTunosTuMVGFP clone (Beauchemin *et al.*, 2005). The fluorescence detection of the GFP-tagged TuMV was realized at 6-8-10-12-14-16 and 23 dpi. Fluorescence was observed using the on-axis zoom microscope Axio Zoom.V16 (ZEISS) equipped with a filter with an excitation window at 460/488 nm and 496 nm of emission.

Phenotyping of the F2, RIL populations and HIFs

The presence or the absence of symptom were scored visually for each individual of F2 and RIL populations at 15 dpi. TuMV symptoms include plant stunting and developmental arrest, leaf mosaic, sawed and curly leaves and aborted flowers (Martín Martín *et al.*, 1999). This trait was expressed as a frequency related to the number of samples showing symptoms on the total number of the four replicates for each RIL.

Viral accumulation on RIL and HIF plants was estimated for each individual plant 15 days post mechanical inoculation using a semi-quantitative double antibody sandwich (DAS)-ELISA assays with a commercial anti-potyvirus monoclonal antibody kit (Adgia Biofords, France). Optical densities (OD) obtained at 405 nm were normalized using *A. thaliana* Col-0 samples infected by TuMV UK1. Plants were considered infected when the OD of the sample is above the mean value of the healthy controls. The final viral accumulation values correspond to the average of normalized OD from replicates of each RIL or HIF line. A rate of infection was estimated and corresponds to the number of samples with a positive ELISA OD on the total number of replicates for each RIL line. This frequency allows to take into account the variability of response of each line.

Genetic mapping

Quantitative trait analysis was performed using the R/qtl package (Broman *et al.*, 2003) and QTLNetwork-2.1 (Yang *et al.*, 2008). RILs genetic map was used (Loudet *et al.*, 2002). A genome scan with a single model was realized using a non-parametric analysis and the Expectation-Maximization (EM) algorithm method (Dempster *et al.*, 1977). Logarithm of odd score (LOD) significance threshold was estimated with 100 permutations and for a significant level α of 0.05. An interval estimate of the location of the QTL was calculated using the 1.5-LOD support intervals method. The 1.5-LOD support interval is the interval in which the LOD score is within 1.5 units of its maximum (Broman & Sen, 2009). The percentage of the phenotypic variation explained by the QTL corresponds to the regression value R^2 taken at the peak LOD score of the QTL.

Fine mapping of the locus detected

Fine mapping was performed on RILs. SNP (Single nuclear polymorphism) molecular markers were developed. The primers were designed using BatchPrimer3 (You *et al.*, 2008) and checked for specificity on TAIR BLAST 2.2.8 (www.arabidopsis.org/Blast). All these markers were used to score RIL samples recombinant at the QTL of interest. All the characteristics of the markers are listed in Table 1. Genotyping was performed by standard PCR amplifications on extracted DNA. Products were sequenced by Sanger sequencing (GENEWIZ Genomics).

Table 1 Characteristics of the markers developed for the fine mapping of the locus detected on chromosome 1

Name	Position in bp	Forward Primers (5' - 3')	Tm_Forward	Reverse Primers (5' - 3')	Tm_Reverse	Fragment size (bp)
SNP61800	22.825.870	AACTACTACGCTTGCTTATCG	54.35	CCAATCCACTCATAGTCATTT	54.16	215
SNP61890	22.870.540	AGACGAATACTGAAGCCAATG	56.99	TGTTTTCTTTCTCTGGTGATG	57.01	236
SNP62030	22.925.416	ACCCGTGTTCTCATCAAGTC	57.54	ACCATTCTTTGGTCAAGGA	57.08	272
SNP62710	23.225.983	ACCTTTAACAGCCTTCTGG	54.52	AAGAGATGGTGACTTTAGGC	53.56	264
SNP62750	23.234.594	TGGTTCATCTCTTCATCAGC	57.14	AAGTGGCTTTGACAGGTGAC	57.77	241
SNP63105	23.404.277	AAGCCACTAACACAGATTGG	56.90	TTCCTTTGCATCAACAACC	57.07	248
SNP63210	23.446.541	ACTGTTAAGAAAGTGACAAGTGG	55.33	CGTGGTAATAATCATCGACA	54.49	253

Identification of candidate genes

In order to reduce the number of candidate genes, an analytical protocol was developed (Supplemental Fig. S1). This protocol proposes to identify polymorphic genes affected by non-synonymous mutations in Bay-0 and Shahdara accessions compared to the *A. thaliana* reference accession, Col-0. Data concerning the polymorphism of these two accessions have been recovered from the 1001genome data center (JGIHeazlewood2008 Project - <http://1001genomes.org/data-center>). The rest of the *in silico* analysis were focused on polymorphic genes only in Bay-0 accession. To predict if a substitution on these genes could have an impact on the biological function of the protein, Provean algorithm was used (http://provean.icvi.org/seq_submit.php). A score below the threshold of -2.5 indicates that the polymorphic protein has great chance to show altered activity (Choi *et al.*, 2012). The Phyre² web portal, produces a set of potential 3D models of protein based on alignment to known protein structure (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>). It was used to compare 3D structure of reference and variant protein using in particular the ProQ² quality assessment (Ray *et al.*, 2012; Kelley *et al.*, 2015). The $\Delta\Delta G$ score calculated is based on equilibrium unfolding. It is used to determine the conformational stability of a protein. Equilibrium unfolding assumes that the protein may belong to only two thermodynamic states: the folded and unfolded state. Thus, the ΔG corresponds to the conformational stability between these two states according to a simple kinetic model. $\Delta\Delta G$ score is calculated through the difference between Bay-0 variant ΔG and Col-0 reference ΔG . It gives an indication on the predicted effect of the mutation(s) on protein stability. A negative score indicates that the mutation is stabilizing whereas a positive score indicates that the mutation is destabilizing. The ΔG were calculated using the interface for FoldX protein design program developed as a plugin for the YASARA molecular graphics suite (Van Durme *et al.*, 2011).

Statistical analysis

Data analysis using parametric and non-parametric statistical tests were realized using the statistical software R version 3.4.0 (R Core Team 2017 - www.R-project.org).

Broad sense heritability were estimated for each trait by fitting a model including a single random intercept effect for the identity of the accessions (R package lme4) $Y_{i,j} \sim \mu + \epsilon_{i,j}$ where μ is the random effect of the identity of the accessions, ($\epsilon_{i,j} \sim N(0, \sigma^2_e)$). The heritability was estimated using the formula $H^2 = \sigma^2_g / [\sigma^2_g + (\sigma^2_e/n)]$, where σ^2_g is the genetic variance, σ^2_e is the environmental variance and n is the number of replicates.

Results

Bay-0 and Shahdara present contrasted responses to TuMV

The response to TuMV infection of the two *A. thaliana* accessions, Bay-0 and Shahdara was measured using an AUDPC score calculated on the basis of symptom scores obtained during different times of the infection (7-8-9-15 dpi) (Fig. 1). Shahdara presents the same score than the Col-0 accession used as positive control of the infection. In the last time point (15 dpi), 92.6% and 87% of Col-0 and Shahdara plants present symptoms respectively. The negative control, *eIF(iso)4E* knockout mutant (E6) showed no symptoms. In the last time points of the infection, some Bay-0 plants, 21 out of 108, could present slight symptoms. But there is no significant differences with the AUDPC scores of the negative control E6.

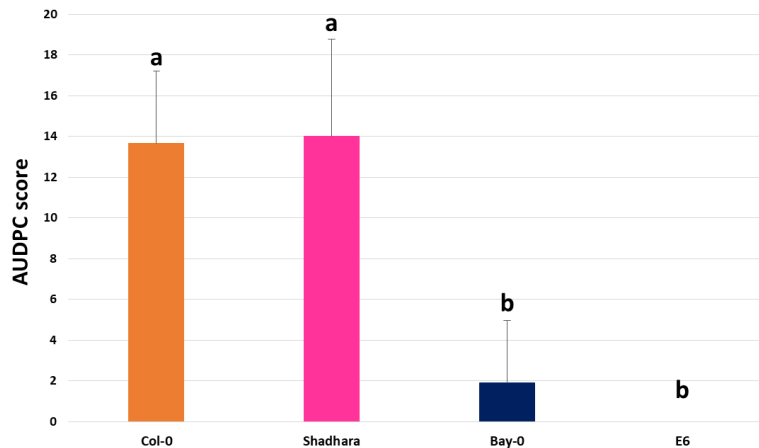


Fig. 1. AUDPC scores of Bay-0 and Shahdara accessions and controls Col-0 and E6

The AUDPC scores were calculated from symptom scores at different times of the infection (7-8-9-15 dpi). Two controls of infection were used: **i**) a positive control, Col-0 accession and **ii**) a negative control, *eIF(iso)4E* knockout mutant line named E6. Statistical analyzes were performed (Kruskal-Wallis test) to show significant differences in AUDPC scores. The results of these analyzes correspond to the letters above the score bars.

Symptom analysis was supplemented by the viral charge quantification by Q-PCR in inoculated and systemic leaves. No TuMV infection was detected locally or systemically in the negative control. On Shahdara and Col-0 inoculated leaves virus was detected as early as 5 dpi and viral charge increase constantly. On Bay-0 inoculated leaves, viral concentration could slightly increase at 9 dpi compared to the negative control. (Fig. 2a). Systemic viral accumulation seems to start earlier on Shahdara (3 dpi) comparing to Col-0 (5 dpi), followed by a strong viral accumulation over the infection. No viral accumulation could be quantified on Bay-0 systemic leaves such as the negative control (Fig. 2b). This tends to show that Bay-0 resistance to TuMV prevents early viral replication.

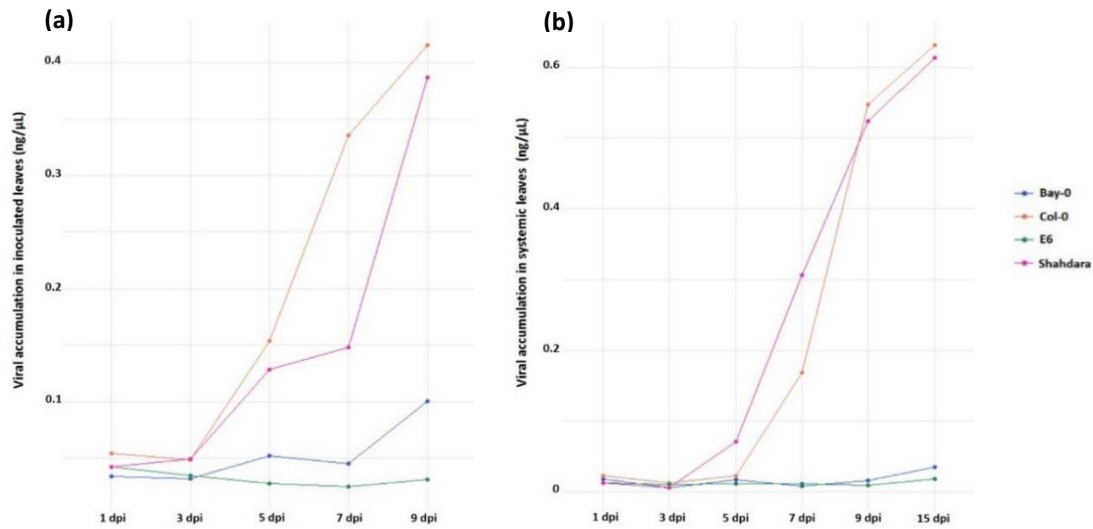


Fig. 2. Evolution of viral accumulation during TuMV infection in inoculated and systemic leaves
 Viral accumulations were expressed as concentrations of virus (in ng/μL). Evolution of viral accumulations measured in inoculated (a) and systemic (b) leaves.

These accessions have also been agroinfiltrated with the pCambiaTunosTuMV-GFP clone in order to track the spread of TuMV in the plants during the infection. GFP-fluorescence observations were realized at different times of the infection (6-8-10-12-14-16 and 23 dpi) (Fig. 3). Shahdara accession presents some visible foci at 6 dpi. The long-distance spread was visible at 10 dpi. While TuMV was able to replicate and accumulate both locally and systemically in Shahdara plants, no infection was detected locally or systemically in 95 % of Bay-0 plants tested. It is ultimately the same behavior as the negative control E6 for which no signal was observed at all. Only, 4 over the 70 Bay-0 plants tested showed a GFP signal that was detected locally and systemically. The phenotype of the plants during infection also varies between these two main accessions. Indeed, Shahdara plants in which a GFP signal was observed were smaller and show yellowing as well as stunting of the leaves. All these symptoms are characteristics of a TuMV infection (Supplemental Fig. S2).

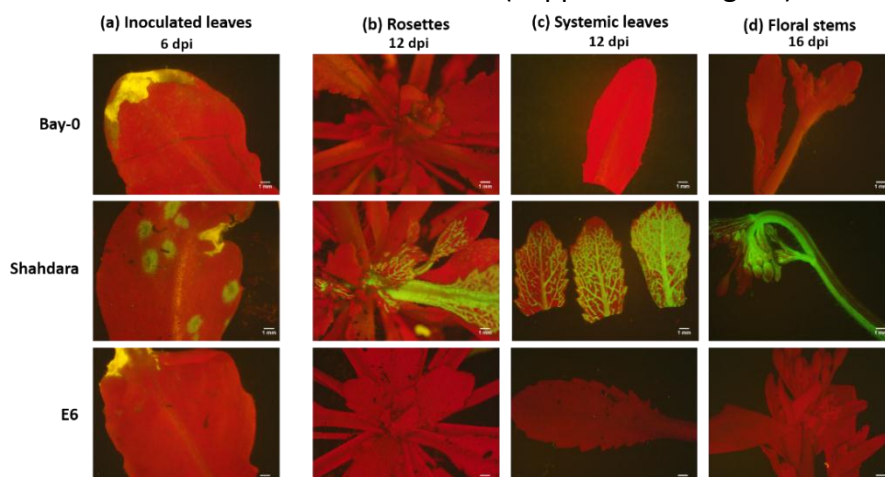


Fig. 3. Observation of green fluorescence protein (GFP)-tagged signal to different parts of *A. thaliana* plants

Photographs realized at different parts of the plants: inoculated leaves (a), rosettes (b), systemic leaves (c) and floral stems (d).

Bay-0 resistance is recessive and controlled by a major locus

In order to test the genetic determinism involved in the response to the TuMV, 200 individuals from the F2 Bay-0 X Shahdara population were tested. A symptom notation was performed at 15 dpi. The segregation data obtained was consistent with a single recessive gene model (69R:131S, p -value = 0.07, for a segregation ratio 1R:3S) (Supplemental Table S2).

RIL population resulting from the crossing between the two contrasted parents to TuMV infection, were mechanically inoculated with TuMV UK1. Their response to the viral infection were analyzed at 15 dpi by recording the presence or the absence of symptoms and by measuring the viral accumulation by DAS ELISA which allows to calculate a frequency of infection per line. The characteristics of the RIL population for each trait is summarized on Table S3. Among the 165 RIL lines tested, 28 did not show detectable virus accumulation by DAS ELISA and no symptoms were observed over all of their repeats. Heritabilities were high, estimated to 0.69 for viral charge, 0.60 for symptoms and 0.59 for the rate of infection.

A major QTL is revealed on *A. thaliana* chromosome 1 through linkage mapping of the resistance to TuMV infection in Bay-0 X Shahdara RIL population. Whatever the phenotypic data used, only one QTL was detected on chromosome 1 with R/qtl package (Fig. 4). The maximum LOD score (LOD=22.32) was obtained with the viral accumulation values (Fig. 4a). In all the analysis, the maximum LOD score is associated with the dCAPsAPR2 marker. The genetic interval calculated by the statistical methods is range from 21.888.056 (marker IND2188) to 24.374.008 (marker F514) with a peak at 22.979.476 (marker dCAPsPR2). The same genomic region was identified whatever the phenotypic trait used (Fig. 4d). The percentage of phenotypic variance explained by this main QTL ranged from 35.37% to 56.88% depending on the phenotypic trait used to perform the analysis. The QTL analysis was also carried out with QTLNetwork software and revealed the same major QTL in chromosome 1 (Supplemental Fig. S3).

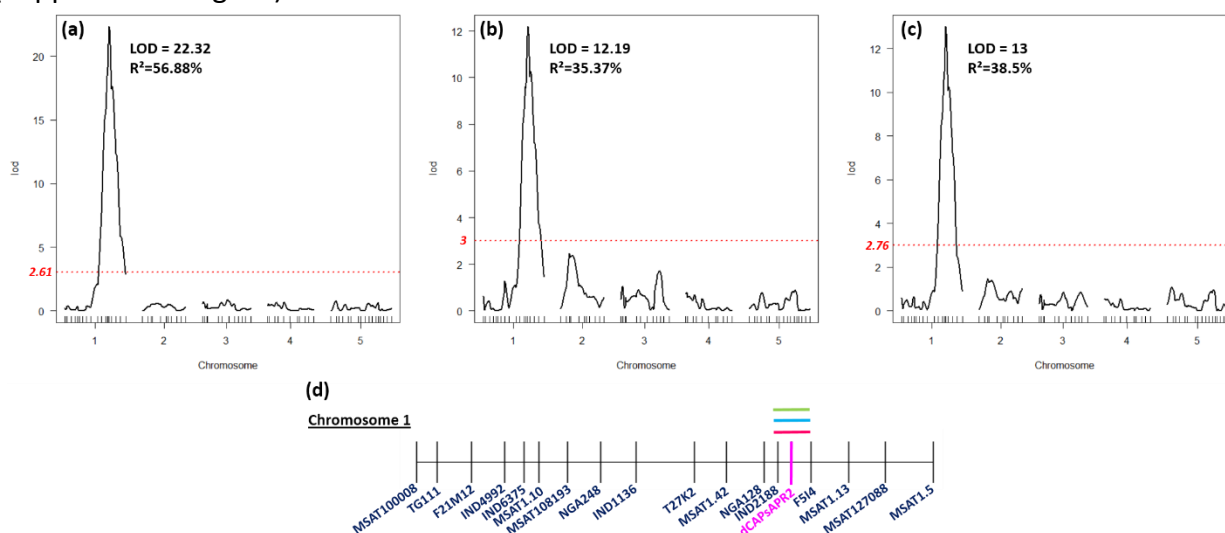


Fig. 4. Linkage mapping realized on the RIL population with three different quantitative traits

Linkage analysis performed with R/qtl package. The y axis represents the LOD score obtained by interval mapping (IM) on the 165 RIL lines. QTL analysis performed with viral accumulation (a), the rate of infection (b) and the symptoms values (c). The characteristics concerning the maximum LOD score and the proportion of the phenotypic variation explained by the peak of the LOD plot (R^2) are represented for each phenotypic trait. The threshold used for detecting QTL significance is represented for each analysis by a dot red line. The interval of the location of the QTL estimated by the 1.5-LOD support interval method are represented in each plot (lines pink (a), light-blue(b) and light-green (c)) but they are also more precisely identified on the scheme of chromosome 1 with the set of these respective markers (d). The dCAPsAPR2 marker corresponds to the maximum LOD score of each QTL analysis.

The phenotyping of HIF68 that spans from marker NGA128 to F514 confirmed the involvement of this locus in responses to TuMV traits. HIF68 was selected in the progeny of RILs because it shows a single residual heterozygous region corresponding to the genome region in which the QTL was identified. Each family is composed by lines with the female parent genotype, Bay-0 lines and Shahdara lines with the male parent genotype. HIF 68 was challenged by mechanical inoculation and the response to TuMV infection was measured by DAS ELISA. The differences in viral accumulation between Bay-0 and Shahdara lines were tested. The HIF68 shows significant differences between Bay-0 and Shahdara lines (Fig. 5).

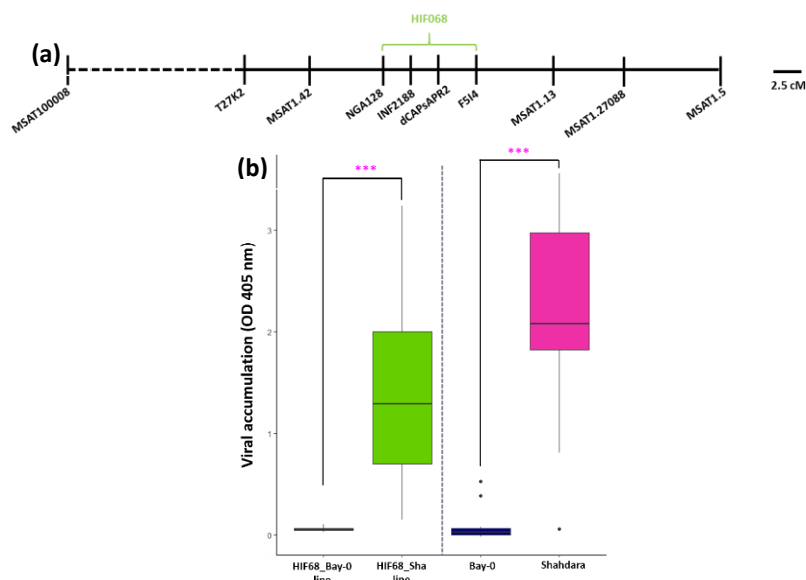


Fig. 5. Phenotypic characterization of heterogeneous inbred HIF 68

HIF 68 was challenged with TuMV. Viral accumulation was measured by DAS ELISA at 15 dpi. Each family is composed by lines with the female parent genotype, Bay-0 lines and also lines with the male parent genotype, Sha lines. **(a)** Localization of the heterozygous regions on chromosome 1. **(b)** Statistical comparisons (Student test) were realized between the female and male lines and parents as a control. Significant differences were represented by asterisks (Signif.codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1) whereas the absence of significant differences are noted "NS".

Fine mapping of the main QTL identified

The fine mapping consisted in testing new markers on three RIL lines, recombinant at the genomic region of interest. This allowed narrowing down the QTL interval to 364 kb, between positions 22.870.540 and 23.234.594 on the long arm of chromosome 1 (Fig. 6).

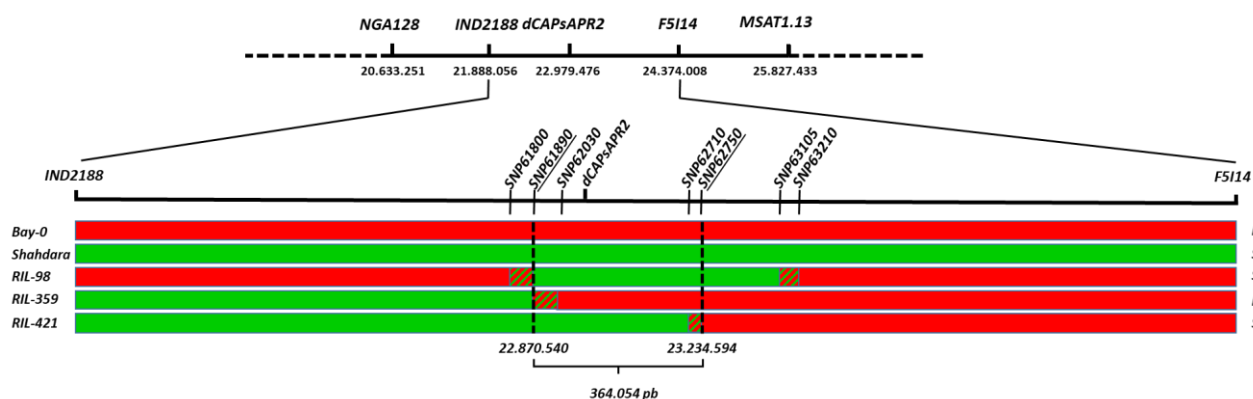


Fig. 6. Schematic representation of RILs used to fine map the locus identified

Markers developed to perform the fine mapping are identified by their name. Red regions correspond to genome regions with the genotype of Bay-0 and green regions to genome regions with the genotype of Shahdara. The phenotype – according to viral accumulation data – are noted as R for the resistant lines and S for the susceptible lines. Red/green boxes correspond to the recombination area.

The quantitative trait locus spans 18 putative candidates involved in Bay-0 resistance to TuMV

The QTL interval retained following the fine-mapping analyzes includes 91 genes. In order to reduce to number of potential candidate genes, an analytical protocol was established (Supplemental Fig. S1). This protocol proposes to select in a first time the polymorphic genes affected by non-synonymous mutations in Bay-0 and/or Shahdara accessions with respect to the *A. thaliana* reference accession, Col-0. Thus, 61 genes were retained among all the genes included in the QTL interval (Supplemental Table S4). Among these 61 candidate genes, 34 genes show polymorphism in both parental accessions whereas 18 of these genes are polymorphic only in the resistant accession, Bay-0 (Table 2). In order to go further in the selection of candidate genes, *in silico* analysis of these 18 genes were carried out (Table 3). The main goal was to look at the effect of mutations on the structure and function of proteins. Thus the Provean software showed four genes AT1G62020 (Coatomer, alpha subunit protein), AT1G62120 (Mitochondrial transcription terminator factor), AT1G62420 (Unknown protein) and AT1G62600 (Flavin-binding monooxygenase family protein) for which at least one mutation had a deleterious effect on protein function. For two of these proteins, AT1G62120 (Mitochondrial transcription terminator factor) and AT1G62420 (Unknown protein), highly negative $\Delta\Delta G$ scores strongly suggest a high impact of mutation of protein conformation and function (Bershtein *et al.*, 2006; Tokuriki & Tawfik, 2009). These two proteins can be considered more stabilized in the Bay-0 variant compared to the Col-0 reference. However, this effect is not confirmed by Phyre² analysis. Only the protein encoded by the AT1G62510 (Lipid transfert protein) showed a decrease in the stability of the protein structure. Indeed, the ProQ² quality assessment used to predict local as well as global quality of protein models is worse in Bay-0 protein model than in Col-0 (Supplemental Fig. S4). For all the other proteins appearing in Table 2, no clear effect of Bay-0 mutations on structure, stability and function of the protein could be established.

Table 2 List of the 18 non-synonymous polymorphic genes in Bay-0 identified in the QTL interval

Gene Name	Protein Description	Protein Function
AT1G62020	Coatomer, alpha subunit protein	Protein transport
AT1G62040	AUTOPHAGY 8C - Ubiquitin-like superfamily protein	Autophagy - Protein transport
AT1G62060	TBA2 - Member of a mucilage protein family	Mucilage biosynthetic process
AT1G62110	Mitochondrial transcription termination factor family protein	Regulation of transcription
AT1G62120	Mitochondrial transcription termination factor family protein	Regulation of transcription
AT1G62250	Orotidine 5-phosphate decarboxylase	Pyrimidine biosynthesis
AT1G62262	SLAC1 HOMOLOGUE 4	Ion homeostasis in guard cells
AT1G62320	Early-responsive to dehydration stress (ERD) family protein	Response to drought stress
AT1G62410	MIF4G domain-containing protein	Translation
AT1G62420	Unknown protein	-
AT1G62480	Vacuolar calcium-binding protein-like protein	Response to cadmium ion
AT1G62500	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	Lipid transport
AT1G62510	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	Lipid transport
AT1G62600	Flavin-binding monooxygenase family protein	Oxidation-reduction process
AT1G62640	3-ketoacyl-acyl carrier protein synthase III (KAS III)	Fatty acid biosynthetic process
AT1G62700	NAC domain transcription factor	Regulation of transcription
AT1G62730	Terpenoid synthases superfamily protein	Biosynthetic process
AT1G62750	ATSCO1 - Nuclear encoded protein	Translation elongation factor activity

Table 3 Effect of non-synonymous mutations on protein sequence and stability through *in silico* modeling analysis

Gene Name	Mutations	Provean	Phyre ² Analysis			ΔΔG
			Conf.	% ID	3D modeling structure	
AT1G62020	S129F	deleterious	100	58	ND	
	V138I	neutral				
AT1G62040	N3E	neutral	100	64	ND	-0.46 kcal/mol
	L8E	neutral				
AT1G62060	G58A	neutral	X	X	X	X
AT1G62110	G47A	neutral	100	23	ND	-0.98 kcal/mol
	S134N	neutral				
	K206R	neutral				
	E308K	neutral				
AT1G62120	R411G	neutral				
AT1G62120	D313E	deleterious	100	17	ND	-13.3 kcal/mol
AT1G62250	T162S	neutral	X	X	X	X
AT1G62262	S129A	neutral	100	20	ND	4.67 kcal/mol
	I336L	neutral				
	S337G	neutral				
AT1G62320	Q318P	neutral	99.39	12	ND	X
AT1G62410	H40D	neutral	100	26	ND	-4 kcal/mol
	A49E	neutral				
	R71G	neutral				
	D74E	neutral				
	E88K	neutral				
AT1G62420	K143N	neutral				
AT1G62420	N66D	neutral	90.66	10	ND	-28.67 kcal/mol
	G138C	deleterious				
	Y152H	neutral				
AT1G62480	G32A	neutral	X	X	X	X
	T39V	neutral				
AT1G62500	T297P	neutral	99.96	39		X
AT1G62510	C64M	neutral	99.96	44	ProQ ² : Mutant < WT	0.04 kcal/mol
AT1G62600	A122P	deleterious	100	28	ND	-6.07 kcal/mol
	E268G	deleterious				
	T452A	neutral				
AT1G62640	M401V	neutral	100	29	ND	7.01 kcal/mol
AT1G62700	D348E	neutral	100	52	ND	X
AT1G62730	D214G	neutral	100	17	ND	-3.58 kcal/mol
AT1G62750	T56H	neutral	100	57	ND	X
	Q76K	neutral				

The polymorphic position in the protein was identified in the “Mutation” column. The reference amino acid (Col-0) is on left side of the polymorphic position, Bay-0 variant amino acid is on the right side. The Provean software result was noted either deleterious or neutral depending on a calculated score. A score below the threshold of -2.5 indicates that the polymorphic protein has great chance to show altered activity. Phyre² software results includes i) Confidence (Conf.) = the probability that the match between the sequence impute and the template is a true homology ii) the likely accuracy of the model (% ID) = percentage of identity between the sequence impute and the template iii) comparisons of the ProQ² quality assessment between the reference (Col-0) and the variant (Bay-0) sequence. The ProQ² is the model quality assessment algorithm that uses support vector machines to predict local as well as global quality of protein models. ND = if no differences were observed between models obtained with reference and variant protein sequence. ΔΔG calculated corresponds to the stability score that indicated the predicted effect of the mutation on protein stability. It shows the energy difference between variant and reference. This ΔΔG was calculated only if the mutation positions are in the matching sequence of the template.

Discussion

This paper reports the fine mapping of a natural recessive resistance locus involved in Bay-0 response to its natural pathogen, TuMV. Bay-0 presents a nearly complete resistance to TuMV with a resistant phenotype similar to the control *eIF(iso)4E* mutant with no virus detection and early arrest of the infection. Nevertheless, the virus could be somehow detected in a small fraction of the inoculated plants. The QTL fine mapping analysis performed with three different phenotypic traits linked to the viral accumulation and the presence of symptoms allowed us to identify a new genome region on the long arm of the chromosome 1 involved in *A. thaliana*/TuMV interaction. A final locus of 364 kb between 22,870,540 and 23,234,594 bp positions was detected. Sixty-one genes showing non synonymous mutations in their coding sequence related to Col-0 susceptible genotype could be considered as putative candidates. Among them, we focused our analysis on the 18 genes whose non-synonymous mutations were carried exclusively by Bay-0.

In order to estimate the effect of non-synonymous mutations between Col-0 and Bay-0 on protein conformation, an *in silico* analysis was performed.

For five proteins - a coatomer alpha subunit protein (AT1G62020), a mitochondrial transcription termination factor family protein (AT1G62120), an unknown protein (AT1G62420), a bifunctional inhibitor/lipid transfer protein/seed storage 2S albumin superfamily protein (AT1G62510) and a flavin-binding monooxygenase family protein (AT1G62600) – a deleterious effect of the mutations on the predicted protein was identified. For two of them, a mitochondrial transcription terminator factor and an unknown protein stability was affected. As mitochondrial transcription terminator factors are a big family in *Arabidopsis* with more than 30 genes coding for these proteins in *Arabidopsis thaliana* (Kleine, 2012), it is hard to say if this protein could be a relevant candidate for Bay-0/TuMV interaction. Even though a role in abiotic stress tolerance has been shown for some of them (Robles *et al.*, 2015).

Beside these five genes, two others for which non synonymous mutation in Bay-0 has been observed but without any predicted effect on protein, could be relevant candidates. The first gene – AT1G62040 – encodes an ubiquitin-like superfamily protein with a role in autophagy. Autophagy is now generally recognize to contribute to anti-microbial defenses, disease resistance (Hofius *et al.*, 2017) including resistance to viruses (Clavel *et al.*, 2017). One recent report have provided novel molecular insights on the antiviral role of autophagy in plant viral responses. Haxim and collaborators (2017), showed that autophagy helps to protect plants against three different geminiviruses (with circular single-stranded DNA genomes). Using the satellite-encoded protein bC1 avirulence factor of *Cotton leaf curl Multan virus* (ClCuMuV) to perform yeast two-hybrid screening of a tomato cDNA library, they identified an ATG8-related protein as an interactor. These studies were carried out in response to DNA viruses belonging to *Geminiviridae* family (Haxim *et al.*, 2017). Recently, Choi *et al.*, (2016) demonstrated autophagy induction in PVY⁰, a potyvirus infecting tobacco leaves. They performed an ultrastructural analysis and observed double-membranous autophagosome-like vacuoles, including lamellar structures or electron-dense materials and multivesicular bodies in PVY⁰ infected leaves. Moreover, a critical autophagy marker in mammals, LC3B/Atg8, was found in extracts of PVY⁰ infected tobacco leaves using Western blotting (Choi *et al.*, 2016).

The second gene –AT1G62410 – encodes a 233 aa MIF-4G domain-containing protein. A protein BLAST analysis revealed a high level of homology (55% of identity – Evalue = 5e-57) with the central protein domain of eIF(iso)4G1, a protein involved in the translation initiation process eIF4F complex. Duprat in 2002 had shown that the lack of eIF(iso)4E inhibits the infection cycle of TuMV. The eIF(iso)4E protein was found to interact with the VPg of TuMV (Wittmann *et al.*, 1997) and this interaction was further found to be necessary for virus infectivity (Léonard *et al.*, 2000). Nicaise and collaborators had shown that TuMV infection, in addition to eIF(iso)4E, requires both eIF(iso)4G1 and eIF(iso)4G2 demonstrating the crucial role of eIF4G factors and their isoforms during *A. thaliana* infection by potyviruses (Nicaise *et al.*, 2007). The analysis of the polymorphisms of this eIF factors on Bay-0, Shahdara and Col-0 accessions did not reveal any differences that might explain the response of these accessions to TuMV (data not shown), that allows to retain the AT1G62410 gene – encoding the MIF-4G domain containing protein – as a potential candidate gene.

As no particular mutation can fully explain the observed phenotype, it is therefore difficult to exclude or retain one or more genes from this list of candidates. It has also to be said that all the strategies used in the *in silico* analysis are predictions based on protein sequence homology. They are highly dependent of the databases quality. Second, the fact that a mutation has no effect on the structure and/or functions of the protein, does not exclude the genes encoding these proteins from the list of candidates as plants has to retain an active form of the protein. Third, an altered structure and/or function highlighted by *in silico* analysis does not mean that genes encoding these proteins are necessarily involved in response to the viral infection. Homologous proteins could perform the same functions.

It appears that ten proteins out of 18 did not appear to be affected by the identified variations. This was the case for the two genes AT1G62040 and AT1G62410 described above. It is also important to underline that all these analyses were performed on genes with polymorphisms in the coding regions of the genes. However, there are numerous, well-characterized examples of mutations in promoter regions. For example, mutations in the pathogenesis related gene PR-1a promoter reduce the level of inducible gene expression for a hypersensitive plant response and systemic acquired resistance (Buchel *et al.*, 1999; Shu *et al.*, 2012). The analysis of the expression of our candidate genes in the available databases has not yielded major results. Exploration of our own results on *A. thaliana* gene expression in response to various potyviruses (data not shown) as the expression dataset provided by the work of Yang *et al.*, in 2007 did not show any variation in response to potyviruses. However, transcriptomic analysis should be performed on our own samples and in our experimental conditions. Finally, our analyses focused initially on polymorphic genes only in Bay-0 accession. But, the QTL interval contains 43 other genes that cannot be excluded from future analyses.

This is the first time that this genomic region of chromosome 1 was mapped in relation to *A. thaliana* response to potyvirus. This locus does not co-localize with any genome region mapped by GWA analyses performed on a world *A. thaliana* population infected by TuMV in a study carried out by our team in common garden conditions (Rubio *et al.*, 2017 – *submitted*).

Two hypotheses could explain this result **i)** the resistance shown is only carried by Bay-0 in the accession set used in common garden experiments. It is therefore a rare allele which cannot be detected by GWA analysis (Roux & Bergelson, 2016) **ii)** the experimental conditions are not the same. Indeed, the Bay-0 X Shahdara RIL population was tested under controlled conditions whereas GWA study was conducted in common garden conditions on a world *A. thaliana* population. This shows the interest of developing studies under different experiment conditions and the complementary support of diverse genetic resources.

Although the involvement of a particular single gene in *A. thaliana* response to TuMV has not been identified yet, this study is original because of the exploration of a natural pathosystem. It revealed a new locus in chromosome 1 associated with a recessive response and highlighted two main candidate genes.

Supplemental Data

Table S1 Heterozygous regions characteristics of the HIF068 from RIL BAY-0 X Shahdara progeny on chromosome 1

Locus	Pos_cM	Pos_Mb	HIF-068
MSAT100008	0	0.008639	S
T1G11	3.5	1.24325	B
F21M12	9.7	3.212191	B
IND4992	15.4	4.992444	B
IND6375	19	6.375557	B
MSAT1.10	21.6	7.296649	B
MSAT108193	26.6	8.192951	B
NGA248	32.4	9.887359	B
IND1136	38.7	11.362151	B
T27K12	49.1	15.926702	B
MSAT1.42	54.7	18.153083	B
NGA128	61.3	20.633251	H
IND2188	63.8	21.888056	H
dCAPsAPR2	66.1	22.979476	H
F514	69.6	24.374008	H
MSAT1.13	76.3	25.827433	S
MSAT1.27088	82.7	27.088847	S
MSAT1.5	91.3	29.016886	S

The heterozygous region is noted by a red “H” and highlighted in yellow. The “B” and “S” correspond to alleles of Bay-0 and Shahdara accessions respectively. Each family is composed of 3 lines with the male parent genotype (corresponding to the Shahdara accession) and 3 lines with the female parent genotype (corresponding to the Bay-0 accession) at the segregating region.

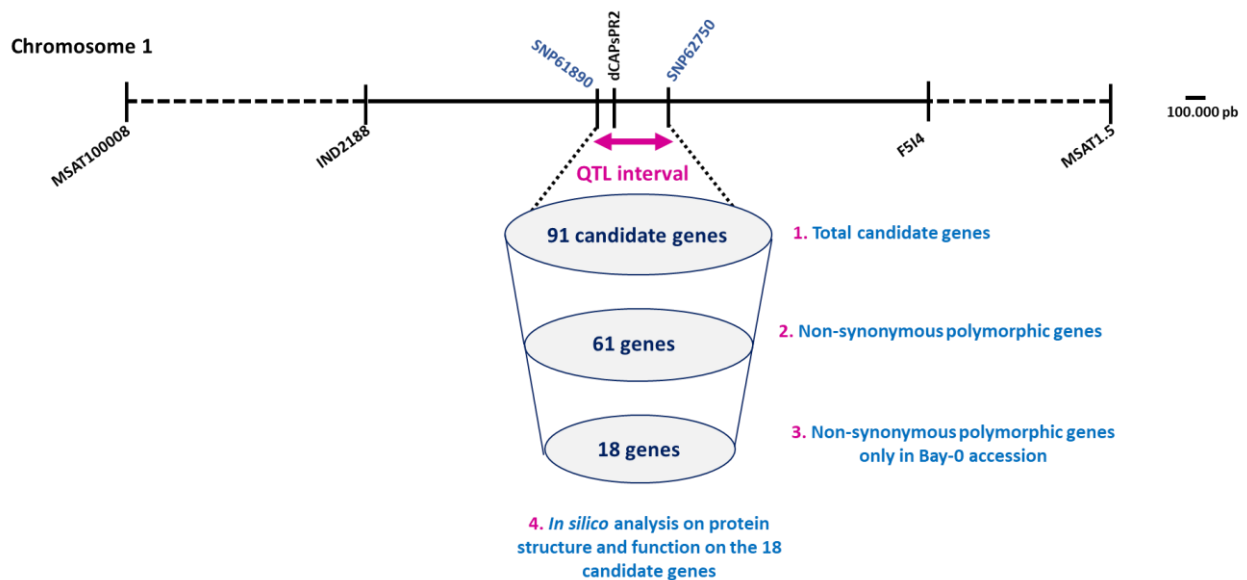


Figure S1 Analytical protocol of the *in silico* analysis of the QTL interval of interest

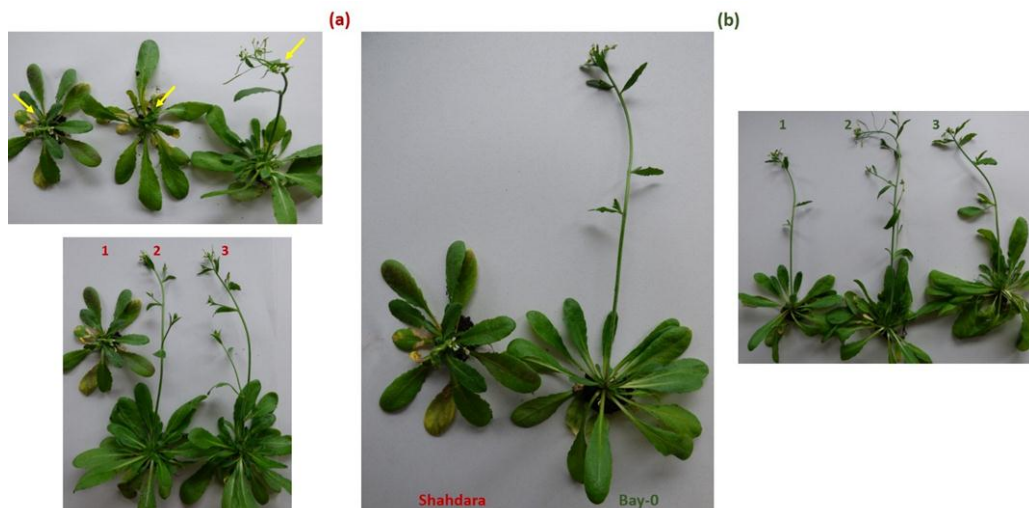


Figure S2 Phenotypic characteristics of Bay-0 and Shahdara accessions at 23 dpi

The photograph of the middle corresponds to Shahdara (left side) and Bay-0 (right side) accessions. Shahdara accessions is smaller, without stem and with yellowing and stunting of the leaves at the core of the rosette. **(a)** Phenotyping of Shahdara accessions: i) the top photograph shows the variety of symptoms (yellow arrows) and ii) the bottom photograph corresponds to an agro-infiltrated Shahdara plant with TuMV (1) and two controls of the assay : Shahdara plant agro-infiltrated with the empty plasmid (2) and Shahdara plant agro-infiltrated with the inoculation solution. **(b)** Phenotyping of Bay-0 accession: the photograph corresponds to an agro-infiltrated Bay-0 plant with TuMV (1) and two controls of the experiment: Bay-0 plant agro-infiltrated with the empty plasmid (2) and Bay-0 plant agro-infiltrated with the inoculation solution (3).

Table S2 Statistical inheritance analysis in the F2 Bay-0 X Shahdara population at 15 dpi

Bay-0 X Shahdara F2 population	
Number tested	198
Segregation	69R:131S
Genetic determinism tested	1R:3S
χ^2 (p-value)	0.0683

P-values calculated for a χ^2 of 1 degree of freedom resulted from a chi-square test of fit of the data to a single recessive gene model (1R:3S). When p-value > 5%, the hypothesis that the observed ratio of segregation is consistent with the expected ration cannot be rejected R for resistance which correspond to individuals that did not show symptoms at 15 dpi opposite to S for sensible individuals.

Table S3 Phenotypic variation for all the traits measured on the Bay-0 X Shahdara RIL population

Phenotypic Trait	Bay-0	Shahdara	RIL mean	RIL range
Viral charge	0.039	0.562	0.297	0 to 1.499
Rate of infection	0.152	0.788	0.464	0 to 1
Symptoms	0.091	0.788	0.566	0 to 1

Three phenotypic traits were measured, **i)** the viral accumulation (viral charge) which corresponds to the average of normalized OD from replicated of each accession (Bay-0 & Shahdara) or each RIL line, **ii)** the rate of infection which corresponds to the number of samples with a positive ELISA OD on the total number of replicates for each accession or RIL line and **iii)** symptoms expressed as a frequency related to the number of samples showing symptoms on the total number of replicates for each accession or RIL line. The "RIL mean" column was obtained by averaging the different traits recorded on all the 165 RIL lines. The "RIL range" corresponds to the variation interval for each of the traits over the 165 RIL lines.

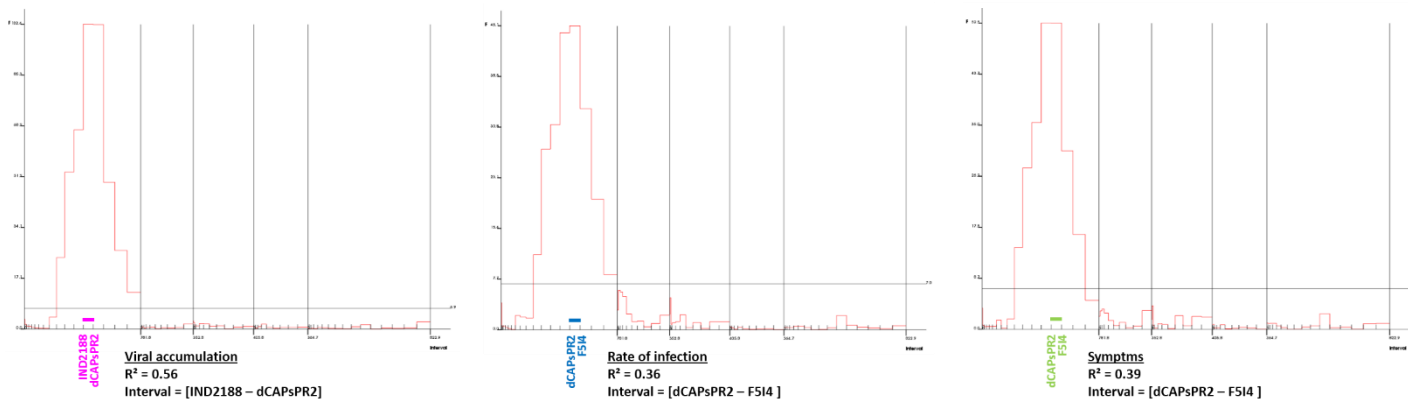


Figure S3 QTL analysis performed with QTLNetwork software

QTL mapping analysis were performed using the three phenotypic traits. The putative interval of QTL was represented by a line of color and by the flanking markers. The percentage of the phenotypic variation explained by the QTL (R^2) was also calculated.

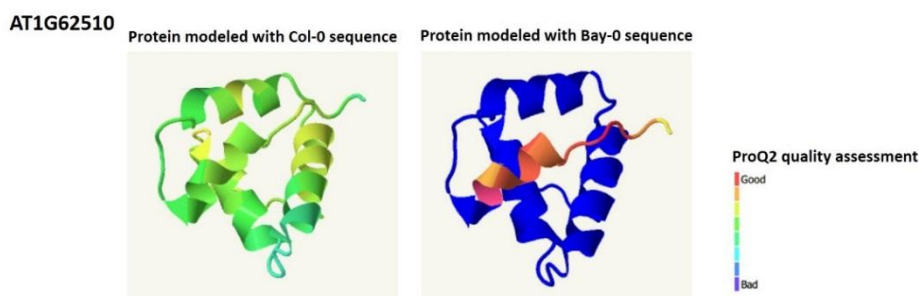


Figure S4 Comparison of the proteins models coded by the Col-0 and Bay-0 sequence of the AT1G62510 gene

ProQ2 is a model quality assessment algorithm uses to predict local as well as global quality of protein models (Ray et al., 2012). The protein coded by Bay-0 sequence has a poorer ProQ2 quality assessment than the protein coded by Col-0 sequence. The polymorphism in AT1G62510 Bay-0 gene decreases the quality of the protein model.

Table S4 List of the 61 non-synonymous polymorphic genes of the QTL interval

Polymorphic genes with non-synonymous mutations in Bay-0 accession	
AT1G62020	Coatomer, alpha subunit
AT1G62040	AUTOPHAGY 8C - Ubiquitin-like superfamily protein
AT1G62060	TESTA ABUNDANT 2 - Encodes a member of a mucilage protein family. Predicted in silico to be glycosylated
AT1G62110	Mitochondrial transcription termination factor family protein
AT1G62120	Mitochondrial transcription termination factor family protein
AT1G62250	orotidine 5-phosphate decarboxylase
AT1G62262	SLAC1 HOMOLOGUE 4 - Predicted to encode a protein with similarity to the SLAC1 protein involved in ion homeostasis in guard cells.
AT1G62320	ERD (early-responsive to dehydration stress) family protein
AT1G62410	MIF4G domain-containing protein
AT1G62420	Unknown protein
AT1G62480	Vacuolar calcium-binding protein-like protein
AT1G62500	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G62510	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G62600	Flavin-binding monooxygenase family protein
AT1G62640	3-ketoacyl-acyl carrier protein synthase III (KAS III)
AT1G62700	ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 26,
AT1G62730	erpenoid synthases superfamily protein
AT1G62750	SNOWY COTYLEDON 1

Polymorphic genes with non-synonymous mutations in Bay-0 AND Shahdara accessions	
AT1G62030	Cysteine/Histidine-rich C1 domain family protein
AT1G61990	Mitochondrial transcription termination factor family protein
AT1G62000	Member of a mucilage protein family. Predicted in silico to be glycosylated
AT1G62070	Hypothetical protein
AT1G62085	Mitochondrial transcription termination factor family protein
AT1G62130	AAA-type ATPase family protein
AT1G62150	Mitochondrial transcription termination factor family protein
AT1G62170	Serine protease inhibitor (SERPIN) family protein
AT1G62180	Adenosine 5'-phosphosulfate reductase, involved in sulfate assimilation
AT1G62190	Kua-ubiquitin conjugating enzyme hybrid localization domain-containing protein
AT1G62200	Major facilitator superfamily protein
AT1G62220	Transmembrane protein
AT1G62260	MITOCHONDRIAL EDITING FACTOR 9 (MEF9), an E subclass PPR protein required for RNA editing.
AT1G62270	F-box and associated interaction domains-containing protein
AT1G62280	SLAC1 HOMOLOGUE 1 - has similarity to the SLAC1 protein involved in ion homeostasis in guard cells
AT1G62305	Core-2/-branching beta-1,6-N-acetylglucosaminyltransferase family protein
AT1G62310	Transcription factor jumonji (jmjC) domain-containing protein
AT1G62330	O-fucosyltransferase family protein
AT1G62340	Subtilisin-like serine protease required for epidermal surface formation in embryos and juvenile plants
AT1G62350	Pentatricopeptide repeat (PPR) superfamily protein
AT1G62430	CDP-diacylglycerol synthase, involved in phospholipid biosynthesis
AT1G62440	Paralog of LRX1 (LEUCINE-RICH REPEAT/EXTENSIN 1) which acts synergistically with LRX1 in root hair cell morphogenesis.
AT1G62490	Mitochondrial transcription termination factor family protein
AT1G62530	Hypothetical protein
AT1G62540	Flavin-monoxygenase (FMO) family, encodes a glucosinolate S-oxygenase that catalyzes the conversion of methylthioalkyl glucosinolates to methylsulfanylalkyl glucosinolates
AT1G62560	Flavin-monoxygenase (FMO) family, encodes a glucosinolate S-oxygenase that catalyzes the conversion of methylthioalkyl glucosinolates to methylsulfanylalkyl glucosinolates
AT1G62580	Flavin monooxygenase that binds NO, has a higher affinity for NO than for O(2) and can generate cGMP from GTP in vitro in an NO-dependent manner.
AT1G62630	Disease resistance protein (CC-NBS-LRR class) family
AT1G62660	Glycosyl hydrolases family 32 protein
AT1G62670	Pentatricopeptide repeat protein required for 5' end processing of nad9 and cox3 mRNAs in mitochondria.
AT1G62680	Pentatricopeptide repeat (PPR) superfamily protein
AT1G62690	Hypothetical protein
AT1G62720	PPR protein gene that localizes to the mitochondrion and is required for seed germination.
AT1G62050	Ankyrin repeat family protein;

Polymorphic genes with non-synonymous mutations in Shahdara accession	
AT1G61890	MATE efflux family protein
AT1G61940	TUBBY LIKE PROTEIN 4 - Member of TLP family
AT1G61970	Mitochondrial transcription termination factor family protein
AT1G61980	Mitochondrial transcription termination factor family protein
AT1G62010	Mitochondrial transcription termination factor family protein
AT1G62290	PUTATIVE ASPARTIC PROTEINASE A2 - Saposin-like aspartyl protease family protein
AT1G62570	Flavin-monoxygenase (FMO) family, encodes a glucosinolate S-oxygenase that catalyzes the conversion of methylthioalkyl glucosinolates to methylsulfanylalkyl glucosinolates
AT1G62710	ASPARAGINYL ENDOPEPTIDASE 2 - Encodes a vacuolar processing enzyme belonging to a novel group of cysteine proteases essential for the proper processing of storage proteins.
AT1G62740	HOP2- Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins with potential to interact with Hsp90/Hsp70 as co-chaperones

References

- Asano M, Satoh R, Mochizuki A, Tsuda S, Yamanaka T, Nishiguchi M, Hirai K, Meshi T, Naito S, Ishikawa M. 2005.** Tobamovirus-resistant tobacco generated by RNA interference directed against host genes. *FEBS Letters* **579**: 4479–4484.
- Beauchemin C, Bougie V, Laliberté J-F. 2005.** Simultaneous production of two foreign proteins from a potyvirus-based vector. *Virus Research* **112**: 1–8.
- Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. 2006.** Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**: 929–932.
- Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Görlach J. 2001.** Growth stage–based phenotypic analysis of *Arabidopsis* a model for high throughput functional genomics in plants. *The Plant Cell* **13**: 1499–1510.
- Broman KW, Sen S. 2009.** *A Guide to QTL Mapping with R/qtl*. New York, NY: Springer New York.
- Broman KW, Wu H, Sen S, Churchill GA. 2003.** R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889–890.
- Buchel AS, Brederode FT, Bol JF, Linthorst HJ. 1999.** Mutation of GT-1 binding sites in the Pr-1A promoter influences the level of inducible gene expression in vivo. *Plant molecular biology* **40**: 387–396.
- Carr T, Whitham SA. 2007.** An Emerging Model System : *Arabidopsis* as a Viral Host Plant. *Plant Cell Monographs* **7**.
- Chisholm ST, Mahajan SK, Whitham SA, Yamamoto ML, Carrington JC. 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.
- Choi D, Park J, Oh S, Cheong H. 2016.** Autophagy induction in tobacco leaves infected by *Potato virus Y O* and its putative roles. *Biochemical and Biophysical Research Communications* **474**: 606–611.
- Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012.** Predicting the functional effect of amino acid substitutions and indels (AG de Brevern, Ed.). *PLoS One* **7**: e46688.
- Clavel M, Michaeli S, Genschik P. 2017.** Autophagy: A double-edged sword to fight plant viruses. *Trends in Plant Science* **22**: 646–648.
- Cooley MB, Pathirana S, Wu H-J, Kachroo P, Klessig DF. 2000.** Members of the *Arabidopsis* HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens. *The Plant Cell* **12**: 663–676.
- Cosson P, Sofer L, Schurdi-Levraud V, Revers F. 2010.** 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**: 1321–1323.

- Decroocq V, Sicard O, Alamillo JM, Lansac M, Eyquard JP, Garcia JA, Candresse T, Le Gall O, Revers F. 2006.** Multiple resistance traits control *Plum pox virus* infection in *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions* **19**: 541–549.
- Dempster AP, Laird NM, Rubin DB. 1977.** Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (methodological)*: 1–38.
- Haxim Y, Ismayil A, Jia Q, Wang Y, Zheng X, Chen T, Qian L, Liu N, Wang Y, Han S, et al. 2017.** Autophagy functions as an antiviral mechanism against geminiviruses in plants. *eLife* **6**: e23897.
- Hofius D, Li L, Hafrén A, Coll NS. 2017.** Autophagy as an emerging arena for plant–pathogen interactions. *Current Opinion in Plant Biology* **38**: 117–123.
- Jenner CE, Walsh JA. 1996.** Pathotypic variation in *Turnip mosaic virus* with special reference to European isolates. *Plant Pathology* **45**: 848–856.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015.** The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**: 845–858.
- Kleine T. 2012.** *Arabidopsis thaliana* mTERF proteins: evolution and functional classification. *Frontiers in Plant Science* **3**: 233.
- Léonard S, Plante D, Wittmann S, Daigneault N, Fortin MG, Laliberté J-F. 2000.** Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. *Journal of Virology* **74**: 7730–7737.
- Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F. 2002.** Bay-0 × Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *TAG Theoretical and Applied Genetics* **104**: 1173–1184.
- Mahajan SK, Chisholm ST, Whitham SA, Carrington JC. 1998.** Identification and characterization of a locus (RTM1) that restricts long-distance movement of *Tobacco etch virus* in *Arabidopsis thaliana*. *The Plant Journal* **14**: 177–186.
- Martín Martín Á, y Poch HLC, Herrera DM, Ponz F. 1999.** Resistances to *Turnip mosaic potyvirus* in *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions* **12**: 1016–1021.
- Nam M, Koh S, Kim SU, Domier LL, Jeon JH, Kim HG, Lee S-H, Bent AF, Moon JS. 2011.** *Arabidopsis* TTR1 causes LRR-dependent lethal systemic necrosis, rather than systemic acquired resistance, to *Tobacco ringspot virus*. *Molecules and Cells* **32**: 421–429.
- Nicaise V, Gallois J-L, Chafiai F, Allen LM, Schurdi-Levraud V, Browning KS, Candresse T, Caranta C, Le Gall O, German-Retana S. 2007.** Coordinated and selective recruitment of eIF4E and eIF4G factors for potyvirus infection in *Arabidopsis thaliana*. *FEBS Letters* **581**: 1041–1046.
- Ouibrahim L, Caranta C. 2013.** Exploitation of natural genetic diversity to study plant-virus interactions: what can we learn from *Arabidopsis thaliana*?: Natural virus resistance in *Arabidopsis thaliana*. *Molecular Plant Pathology* **14**: 844–854.
- Ouibrahim L, Mazier M, Estevan J, Pagny G, Decroocq V, Desbiez C, Moretti A, Gallois J-L, 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* **79**: 705–716.
- Pagan I, Fraile A, Fernandez-Fueyo E, Montes N, Alonso-Blanco C, Garcia-Arenal F. 2010.** *Arabidopsis thaliana* as a model for the study of plant-virus co-evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1983–1995.
- Poque S, Pagny G, Ouibrahim L, Chague A, Eyquard J-P, Caballero M, Candresse T, Caranta C, Mariette S, Decroocq V. 2015.** Allelic variation at the *rpv1* locus controls partial resistance to *Plum pox virus* infection in *Arabidopsis thaliana*. *BMC Plant Biology* **15**: 159–172.
- Pumplin N, Voinnet O. 2013.** RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nature Reviews Microbiology* **11**: 745–760.
- Ray A, Lindahl E, Wallner B. 2012.** Improved model quality assessment using ProQ2. *BMC Bioinformatics* **13**: 224–235.
- Revers F, Guiraud T, Houvenaghel M-C, Mauduit T, Le Gall O, Candresse T. 2003.** Multiple resistance phenotypes to *Lettuce mosaic virus* among *Arabidopsis thaliana* accessions. *Molecular Plant-Microbe Interactions* **16**: 608–616.
- Robles P, Micol JL, Quesada V. 2015.** Mutations in the plant-conserved MTERF9 alter chloroplast gene expression, development and tolerance to abiotic stress in *Arabidopsis thaliana*. *Physiologia Plantarum* **154**: 297–313.
- Roux F, Bergelson J. 2016.** The genetics underlying natural variation in the biotic interactions of *Arabidopsis thaliana*. *Current Topics in Developmental Biology* **119**: 111–156.
- Shu QY, Forster BP, Nakagawa H (Eds.). 2012.** *Plant mutation breeding and biotechnology*. Cambridge, Mass: CABI International.
- Simko I, Piepho H-P. 2012.** The area under the disease progress stairs: calculation, advantage, and application. *Phytopathology* **102**: 381–389.
- Takahashi H, Miller J, Nozaki Y, Takeda M, Shah J, Hase S, Ikegami M, Ehara Y, Dinesh-Kumar SP. 2002.** RCY1, an *Arabidopsis thaliana* RPP8/HRT family resistance gene, conferring resistance to *Cucumber mosaic virus* requires salicylic acid, ethylene and a novel signal transduction mechanism. *The Plant Journal* **32**: 655–667.
- Tokuriki N, Tawfik DS. 2009.** Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology* **19**: 596–604.
- Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F. 2011.** A graphical interface for the FoldX forcefield. *Bioinformatics* **27**: 1711–1712.
- Walsh JA, Jenner CE. 2002.** *Turnip mosaic virus* and the quest for durable resistance. *Molecular Plant Pathology* **3**: 289–300.
- Whitham SA, Anderberg RJ, Chisholm ST, Carrington JC. 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**: 569–582.

- Whitham SA, Yamamoto ML, Carrington JC. 1999.** Selectable viruses and altered susceptibility mutants in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **96**: 772–777.
- Wittmann S, Chatel H, Fortin MG, Laliberté J-F. 1997.** Interaction of the viral protein genome linked of *Turnip mosaic potyvirus* with the translational eukaryotic initiation factor (iso) 4E of *Arabidopsis thaliana* using the yeast two-hybrid system. *Virology* **234**: 84–92.
- Yamanaka T, Imai T, Satoh R, Kawashima A, Takahashi M, Tomita K, Kubota K, Meshi T, Naito S, Ishikawa M. 2002.** Complete inhibition of tobamovirus multiplication by simultaneous mutations in two homologous host genes. *Journal of Virology* **76**: 2491–2497.
- Yang J, Hu C, Hu H, Yu R, Xia Z, Ye X, Zhu J. 2008.** QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* **24**: 721–723.
- You FM, Huo N, Gu Y, Luo M, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD. 2008.** BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9**: 253.

- Discussion Générale -

Ce travail de thèse s'inscrit dans les objectifs généraux de recherche de l'équipe Virologie Végétale de l'UMR Biologie du Fruit et Pathologie qui visent à mieux comprendre les interactions entre les potyvirus et leurs plantes hôtes. Ainsi, cette thèse a porté plus particulièrement sur l'étude du **pathosystème naturel *Arabidopsis thaliana*/TuMV**. Différentes méthodes et approches ont été utilisées afin de caractériser de façon fine la réponse d'*A. thaliana* au TuMV dans le but de pouvoir identifier de potentiels facteurs de l'hôte impliqués dans cette interaction et de façon plus large dans les interactions plante/virus. Ce travail a ainsi permis d'explorer la diversité naturelle d'*A. thaliana* via des **approches génétiques et métaboliques** avec des essais menés majoritairement en conditions extérieures. L'ensemble des résultats obtenus ont permis **i)** de détecter via des analyses de génétique d'association et de QTL mapping en conditions extérieures et en conditions contrôlées, plusieurs loci. Parmi eux, certains contiennent des gènes candidats déjà décrits comme impliqués dans la réponse des plantes à des virus. De façon plus originale, de nouveaux gènes candidats potentiels de l'interaction *A. thaliana*/TuMV ont également été décrits ; **ii)** de caractériser, de façon novatrice, la réponse métabolique d'*A. thaliana* en condition d'infection. Cela a notamment permis de montrer qu'il était possible de discriminer les accessions en fonction de leur sensibilité au TuMV sur la base de leurs profils en métabolites primaires et secondaires. Les résultats majeurs de chacune des études et parties de ce travail de thèse sont repris dans le schéma ci-dessous.

Réponse d'*Arabidopsis thaliana* au Turnip mosaic virus (TuMV) en conditions extérieures et en conditions contrôlées : phénotypage fin de traits de maladie et métaboliques et architecture génétique associée

- OBJECTIFS**
- Caractérisation des bases génétiques de la réponse d'*A. thaliana* au TuMV en conditions extérieures et en conditions contrôlées
 - Identification de nouveaux gènes candidats de l'interaction *A. thaliana*/TuMV
 - Phénotypage métabolique fin afin de mieux caractériser l'interaction *A. thaliana*/TuMV

CHAPITRE 1 = Architecture génétique de la réponse d'*A. thaliana* au TuMV en conditions extérieures et phénotypage fin de la réponse métabolique

PARTIE A = Bases génétiques de la réponse d'*A. thaliana* au TuMV en conditions extérieures

PARTIE B = Phénotypage métabolique fin de la réponse au TuMV en conditions extérieures

Population mondiale

Population française

Population mondiale

Les résultats majeurs :

- Le locus **RTM3** apparaît être un **locus majeur de réponse au TuMV**, et à d'autres potyvirus chez *A. thaliana*
- L'exploration de la **diversité mondiale** d'*A. thaliana* en conditions extérieures permet d'identifier de **nouveaux loci de réponse au virus**
- L'exploration d'une **diversité locale** d'*A. thaliana* permet d'identifier des **loci différents de la population mondiale**
- Les **profils métaboliques** permettent de **différencier** clairement les **accessions sensibles et résistantes**. Certaines molécules apparaissent fortement présentes dans les génotypes résistants
- Les **perturbations métaboliques** apparaissent dès les **temps précoces** de l'infection mais, la **discrimination** des accessions **sensibles et résistantes** est observée lorsque la charge virale atteint un certain niveau en ELISA
- Les bases génétiques de la réponse métabolique sont complexes à décrire

CHAPITRE 2 = Recherche de déterminants génétiques de l'interaction *A. thaliana*/TuMV via une approche bi-parentale en conditions contrôlées

Population RIL Bay-0 X Shahdara

Les résultats majeurs :

- **Bay-0** est une **accession résistante** au TuMV présentant un blocage du virus dès les premières étapes de l'infection. La résistance est **récessive**
- Un **locus majeur** de 364 kb est identifié sur **chromosome 1**
- **61 gènes** présentent des **mutations non-synonymes**
- **18/61 gènes** avec des **mutations non-synonymes uniquement chez Bay-0** révèlent **deux nouveaux candidats potentiellement intéressants**, **AT1G62040** codant pour une protéine impliquée dans les mécanismes d'autophagie et le gène **AT1G62410** codant pour une protéine avec un domaine MIF-4G

L'interaction *Arabidopsis thaliana*/TuMV : bénéficiaire de la richesse de l'espèce en l'explorant dans le cadre d'un pathosystème naturel

Les recherches menées sur *A. thaliana* ont permis de mettre en évidence et de décrire divers mécanismes chez les plantes contribuant de ce fait à faire avancer les connaissances dans différents domaines du végétal. L'ensemble des ressources génétiques et génomiques ainsi que d'importantes données de transcriptomique, protéomique et métabolomique recensées dans diverses conditions (environnements, stades de développement) ont fait d'*A. thaliana* une excellente source de référence pour les études concernant les plantes cultivées (Sivasubramanian *et al.*, 2015; Provar *et al.*, 2016). Cette plante modèle est également considérée comme un hôte compatible avec plusieurs des principaux genres et espèces virales. En effet, *A. thaliana* constitue un hôte favorable pour le développement de 40 espèces virales appartenant à 18 genres différents (Carr & Whitham, 2007; Ouibrahim & Caranta, 2013). Les différents outils, ressources génétiques et moléculaires disponibles rendent l'utilisation de cette plante modèle intéressante dans la recherche de nouvelles sources de résistances aux virus. De plus, il a été montré que certaines bases moléculaires de la sensibilité des plantes aux virus sont relativement bien conservées dans le règne végétal (Carr & Whitham, 2007; Ouibrahim & Caranta, 2013). Le transfert des connaissances à des espèces cultivées étant en plus facilité par le développement des nouvelles technologies (séquençage haut débit, outils biotechnologiques : mutagenèse dirigée, transgénèse, CRISPR-Cas9). Ainsi, un certain nombre de gènes et de loci impliqués dans la réponse à des infections virales ont été identifiés et décrits chez *A. thaliana* (Ouibrahim & Caranta, 2013). En revanche, peu de connaissances existent sur la capacité des virus à infecter naturellement *A. thaliana*. Seule une étude semble s'être intéressée à cette question en recensant, pendant quatre ans, la présence de cinq espèces virales, le *Turnip mosaic virus* (TuMV), le *Turnip yellow mosaic virus* (TYMV), le *Turnip crinkle virus* (TCV), le *Cucumber mosaic virus* (CMV) et le *Cauliflower mosaic virus* (CaMV) dans six populations naturelles d'*A. thaliana*, originaires du centre de l'Espagne (Pagan *et al.*, 2010). Ainsi, dans la plupart des études, les informations concernant une possible coévolution entre *A. thaliana* et les virus étudiés sont inexistantes. *A. thaliana* a été montrée comme naturellement infectée par le TuMV. Ce travail de thèse porte sur l'étude d'un **pathosystème naturel**. Des facteurs de l'hôte impliqués dans la réponse à l'infection virale peuvent être le résultat de cette possible coévolution entre *A. thaliana* et le TuMV. En étudiant ce pathosystème naturel, on essaie d'être plus proche des mécanismes qui pourraient être naturellement retrouvés chez les plantes en réponse aux virus. Ces mécanismes pouvant certainement être différents de ceux mis en place dans le cadre d'interactions « naïves » entre les plantes et les virus (Pagan *et al.*, 2010).

Afin d'étudier ce pathosystème naturel, différentes populations d'*A. thaliana* ont été utilisées telles que **i)** des accessions faisant partie de la collection mondiale, **ii)** une population locale française (population TOU-A) ainsi qu'une **iii)** population de RIL issue du croisement entre deux parents au phénotype contrasté. L'exploration de la variabilité génétique de ces différentes populations a permis de mettre en évidence de nouveaux loci avec de potentiels gènes candidats de l'interaction *A. thaliana*/TuMV.

Cette variabilité génétique naturelle a conduit à considérer *A. thaliana* comme un modèle dans l'étude des interactions avec des facteurs biotiques. Des analyses de génétique d'association ont de ce fait été développées permettant l'identification de facteurs impliqués dans ces interactions (Roux & Bergelson, 2016). Cependant, Bartoli et Roux dans leur revue décrivant les études de GWA menées dans le cadre d'interaction plantes/agents pathogènes, ont montré qu'aucune analyse de GWA n'a été menée en réponse à des virus en conditions naturelles. Cela permet de mettre l'accent sur l'originalité des analyses GWA utilisées pour caractériser les bases génétiques de la réponse au TuMV des différentes populations d'*A. thaliana* étudiées en conditions extérieures. De plus, l'architecture génétique de la réponse à l'infection par le TuMV a été étudiée chez des populations d'*A. thaliana* à différentes échelles géographiques puisque la variabilité génétique d'une population mondiale mais également une population locale française ont été explorées. Ainsi, ces analyses ont permis de confirmer, par exemple, l'implication potentielle, des protéines avec un domaine TRAF, dans la réponse au TuMV, voire même dans la réponse aux potyvirus. En effet, un cluster de 11 gènes codant pour des protéines avec un domaine TRAF a été identifié sur deux loci situés à proximité (~ 10kb) sur le chromosome 3 par les analyses GWA menées sur ces deux populations. Dans ce cluster de 11 gènes se trouve le gène *RTM3* déjà décrit comme impliqué dans la réponse d'*A. thaliana* aux potyvirus (Whitham *et al.*, 1999; Decroocq *et al.*, 2006; Cosson *et al.*, 2012). Ces analyses GWA, ont également mis en évidence des loci différents entre les deux populations permettant d'illustrer l'idée que la variabilité naturelle de certains traits phénotypiques dépend de l'échelle géographique considérée (Brachi *et al.*, 2013).

L'ensemble de ces résultats renforcent l'idée que la variabilité génétique des populations d'*A. thaliana* permet d'identifier des gènes dont certains n'auraient pas pu l'être en utilisant les accessions classiquement étudiées en laboratoire. Un exemple clair illustrant cette idée est l'identification des loci impliqués dans la floraison *FRIGIDA* (FRI) et *Flowering Locus C* (FLC). En effet, une floraison tardive n'a lieu que lorsque ces deux loci possèdent des versions alléliques actives, or dans les accessions d'*A. thaliana* communément utilisées au laboratoire (Ler, Col-0) au moins un des loci est inactif. Cette relation d'épistasie fait que certains phénotypes (ici la floraison tardive) n'apparaissent que dans certains fonds génétiques (Caicedo *et al.*, 2004; Koornneef *et al.*, 2004).

Enfin, dans certaines conditions, les variations génétiques ne sont pas détectées en utilisant la simple comparaison entre accessions. Dans ce cas, l'étude sur des populations issues de croisements entre individus présentant un phénotype contrasté devient intéressante (Koornneef *et al.*, 2004). C'est ainsi que l'analyse QTL menée sur la population de RIL entre Bay-0 et Shahdara a permis d'identifier un locus majeur sur le chromosome 1 de la réponse au TuMV. Malgré les études ayant décrit le phénotype résistant de Bay-0 et malgré la disponibilité de cette population de RIL, à notre connaissance aucune étude ne s'était jusqu'à alors intéressée à étudier le déterminisme génétique de la résistance de Bay-0 au TuMV (Martín Martín *et al.*, 1999; Carr & Whitham, 2007).

Un phénotypage élargi et fin apporte de nouvelles connaissances sur l'interaction entre *Arabidopsis thaliana* et le TuMV

Le premier objectif de ce travail de thèse a été de caractériser la réponse d'*A. thaliana* au TuMV. Ce travail est une première étape indispensable, il s'agit en quelques sortes de mieux connaître son ennemi afin de pouvoir mettre en place les meilleures stratégies pour pouvoir le battre. Pour cela le **phénotypage** de la réponse d'*A. thaliana* lors d'une infection virale revêt une importance cruciale. Alors que des méthodes et approches de génotypage ont facilité la caractérisation à grande échelle des génomes végétaux, le phénotypage est devenu le principal facteur limitant. Or, la capacité à disposer de phénotypes clairs, précis et robustes est d'une importance majeure afin de pouvoir les associer au génotype et identifier les gènes qui aideraient à mieux comprendre et maîtriser les caractères complexes. La précision du phénotypage est indispensable puisqu'elle constitue la base du développement des différentes approches et analyses développées afin de comprendre la réponse biologique étudiée.

Dans le cadre de cette thèse, la description du comportement d'*A. thaliana* lors d'une infection par le TuMV a été effectuée en utilisant différents traits phénotypiques. De façon classique – comme dans la plupart des études portant sur les interactions plantes/agents pathogènes – **des traits en lien avec la maladie** ont été utilisés. Parmi ces traits phénotypiques se retrouvent **i)** l'accumulation virale mesurée par ELISA semi-quantitatif, **ii)** des notations de symptômes liées à leur présence/absence ou en utilisant une échelle de notation en fonction de leur sévérité, **iii)** des fréquences d'infection qui rendent compte de la variabilité de la réponse au virus en fonction de l'accession mais également **iv)** des notations en lien avec la croissance des plantes (mesure de la hauteur des hampes, biomasse sèche/fraîche).

En parallèle, des **analyses métaboliques** ont été effectuées. Au travers de dosages métaboliques à haut-débit non-ciblés et ciblés, la réponse d'*A. thaliana* au TuMV a ainsi été décrite avec un autre angle de vue. L'étude de la réponse métabolique d'*A. thaliana* au TuMV apparaît être une des originalités principales de cette thèse.

Le développement des techniques à haut débit a permis – entre autres – de réaliser des profilages métaboliques permettant d'appréhender la complexité du vivant dans son ensemble. L'utilisation de ces données métaboliques contribue à apporter des informations sur le fonctionnement de certaines réponses biologiques. Ces métabolites sont également utilisés comme des outils pour l'évaluation de critères de performance dans le cadre de process industriels. Enfin, les métabolites peuvent aussi être utilisés afin d'évaluer les critères de performance des plantes soumises à des stress. Si un certain nombre d'études ont révélé le rôle de métabolites dans la réponse à des stress abiotiques, l'utilisation de ces données métaboliques est relativement récente dans les études des plantes soumises à des stress biotiques et cela est d'autant plus vrai dans l'étude des interactions plante/virus (Zabotina, 2013; Fernandez *et al.*, 2016).

L'objectif au départ de ce travail, était de voir s'il était possible de pouvoir prédire – avant infection – la sensibilité des accessions étudiées à partir de leur contenu en métabolites primaires. Cependant, l'ensemble des analyses effectuées, sur différentes années montrent qu'il n'y a aucune différence dans le contenu métabolique pouvant être associée à la sensibilité des accessions avant infection. Ce premier résultat a été une réelle déception au départ.

En revanche, un résultat majeur obtenu sur un grand nombre d'accessions montre des différences importantes dans le contenu en métabolites primaires entre accessions en fonction de leur sensibilité en condition d'infection. Les accessions sensibles au TuMV présentent une forte accumulation en métabolites primaires par rapport aux accessions résistantes. L'utilisation de modèles statistiques du type OPLS-DA a permis de mettre en évidence le fait qu'avec une dizaine de métabolites primaires, il était possible d'obtenir des coefficients de prédiction tout à fait intéressants avec une part non négligeable de la variabilité de la réponse qui est expliquée par le métabolisme. Ce résultat montre qu'une contribution du métabolisme à la réponse au virus est importante chez les plantes. Cette contribution n'est pas seulement une conséquence de l'infection mais accompagne la réplication virale et la systémisation du virus. L'importance de facteurs métaboliques de l'hôte a été, ces dernières années, très étudiée dans le modèle *Tomato blushy stunt virus* (TBSV)/levure. Il a été montré que certaines formes lipidiques contribuent clairement au support de la réplication virale (Xu & Nagy, 2015, 2016) tandis que la glycolyse et le recrutement d'une pyruvate kinase glycolytique sont des facteurs essentiels de fourniture d'ATP (Chuang *et al.*, 2017). La démonstration des perturbations métaboliques et du rôle de la glycolyse ont également été soulignés pour des virus animaux. C'est le cas pour le virus de l'hépatite C (Diamond *et al.*, 2010) et pour les virus de la Dengue ainsi que les Semliki Forest virus (SFV) et Sindbis virus (SINV) (Fontaine *et al.*, 2015; Findlay & Ulaeto, 2015).

Les dosages métaboliques non ciblés ont permis **i)** d'augmenter le nombre de variables métaboliques analysées afin d'augmenter la puissance statistique du modèle **ii)** d'avoir accès à un autre niveau d'information en lien avec le métabolisme secondaire. Ces analyses nous ont permis de confirmer définitivement qu'en absence d'infection par le TuMV aucune prédiction de sensibilité ne peut être effectuée. En revanche ce jeu de données conséquent a permis d'améliorer de façon notable les modèles prédictifs mis en place. Une réelle perturbation du métabolisme existe chez les accessions sensibles au TuMV avec notamment une grande accumulation de métabolites secondaires. Au contraire, les accessions résistantes semblent maintenir un équilibre/compromis entre croissance et défense. Cela se traduit notamment par l'accumulation d'un nombre beaucoup plus réduits de composés secondaires qui pourraient contribuer ou être le résultat de mécanismes d'action efficaces dans la lutte contre le TuMV (Rubio *et al.*, 2017 – *en préparation*).

Si la majorité des réponses a été analysée à un moment précis de l'infection, correspondant à l'apparition des premiers symptômes, l'analyse cinétique de cette réponse métabolique a permis de mettre en évidence des perturbations métaboliques dès les premières étapes de l'infection. En revanche, la discrimination des accessions sensibles et résistantes ne s'effectue qu'au moment où l'infection atteint un certain niveau de détection. D'ailleurs, dans les études menées sur la réponse métabolique à des infections virales, les modifications importantes du

métabolisme correspondent au moment où l'accumulation virale est considérée comme maximale (Fernandez-Calvino *et al.*, 2014).

La plupart des études qui se sont intéressées à la réponse métabolique des plantes lors d'infections virales ont été menées en conditions contrôlées, sur un faible nombre de génotypes et sur des systèmes souvent sensibles. Ainsi, l'ensemble des analyses métaboliques, menées au cours de cette thèse, a donc permis de caractériser la réponse métaboliques d'un large panel d'accessions d'*A. thaliana* en conditions extérieures. Cela a conduit à mettre en évidence des marqueurs métaboliques de la réponse au virus particulièrement intéressants notamment en ce qui concerne les accessions résistantes. Leur identification, en cours, permettrait notamment de mieux comprendre les mécanismes de réponse des plantes au virus.

Un phénotypage en conditions extérieures : se rapprocher des conditions réelles de l'infection

Un autre point original de cette thèse réside dans les conditions d'expérimentation utilisées pour étudier la réponse d'*A. thaliana* au TuMV puisque des essais ont été menés en conditions extérieures. Très peu d'études concernant les interactions plantes/virus ont été menées en conditions naturelles (Hily *et al.*, 2016). Or, dans ces conditions, les plantes sont soumises à diverses variations, plus ou moins intenses, tout au long de la journée. Les conditions naturelles sont de ce fait très différentes des conditions contrôlées que ce soit en terme de température, qualité et intensité lumineuse, humidité, entre autres, cela se traduit notamment par des phénotypes différents (Jänkänpää *et al.*, 2012; Annunziata *et al.*, 2017). Les paramètres tels que la lumière et la température vont avoir un impact important sur la physiologie des plantes. Par exemple, en conditions contrôlées les écarts de température sont faibles voire parfois inexistantes entre le jour et la nuit alors que de fortes variations (entre 7 et 12°C) sont enregistrées au champ. Ainsi, les mécanismes qui dépendent de la température telle que la division cellulaire seront stimulés en conditions contrôlées et souvent proches des optimums de croissance des plantes. Il est nécessaire d'avoir conscience de ces différences puisqu'elles pourraient avoir un impact dans le transfert des connaissances du laboratoire au champ. En effet, si le fossé est trop important entre les conditions d'études et celles sur lesquelles le transfert des résultats doit s'effectuer, il se peut que dans des situations extrêmes l'énergie et le matériel dépensés en conditions contrôlées l'aient été en vain. L'effet des conditions d'expérimentations a été particulièrement observé au cours de nos différents essais, les plantes présentant des feuilles et des hampes plus épaisses ainsi que des colorations violettes. Ces différences observées, ont un impact clair sur le phénotype et pourraient être à l'origine d'une réponse au TuMV différente de celle déjà décrite jusqu'à présent en conditions contrôlées. D'ailleurs si un locus au niveau du chromosome 3 co-localise avec le gène *RTM3* et le locus *sha3*, identifiés en conditions contrôlées en réponse à différents potyvirus, de nouveaux loci, jamais encore identifiés en réponse à des potyvirus ont été mis en évidence. L'ensemble de ces observations illustrent bien la plasticité phénotypique des plantes et leur capacité à répondre de différentes façons en fonction des conditions environnementales.

C'est souvent cette variabilité de réponse, difficile à prendre en compte, qui a conduit à réaliser la majorité des essais dans des conditions plus contrôlées (Poorter *et al.*, 2016).

La reproduction de ces fluctuations naturelles est très complexe à mettre en place en conditions contrôlées. Plusieurs questions se posent dans ces conditions concernant l'application des contraintes abiotiques. En effet, l'objectif est de pouvoir se rapprocher au maximum des conditions naturelles, les contraintes abiotiques ne doivent pas être appliquées de façon trop intense. La question concernant l'ordre d'application de ces stress se pose également : faut-il les appliquer en simultanément, en séquentiel ? Toutes ces questions sont d'autant plus importantes, qu'il a été montré que la réponse des plantes à des stress multiples ne peut être prédite sur la base des connaissances obtenues en réponse à des stress uniques. De plus, les réponses ne sont pas les mêmes si les stress sont appliqués l'un après l'autre ou en même temps (Mittler, 2006; Suzuki *et al.*, 2014; Coolen *et al.*, 2016; Pandey *et al.*, 2017). Néanmoins, la quantification des stress subis est plus difficilement réalisable à l'extérieur et certaines variables plus difficilement accessibles qu'en conditions contrôlées (exemple de la prise d'image en continu pour évaluer le développement). On peut cependant penser que les plateformes – déjà opérationnelles ou à venir de phénotypage haut-débit en champ – permettront de se rapprocher de certains phénotypes haut-débits disponibles en conditions contrôlées. Pour le moment, les deux stratégies apparaissent complémentaires.

Les études menées en conditions extérieures au cours de cette thèse ont été complétées par des études menées en conditions contrôlées (chambre de culture et serre). Cela a permis de mettre en évidence des réponses différentes notamment en termes de déterminants génétiques identifiés. Comme cela est souligné par Poorter et ses collaborateurs, c'est l'utilisation de ces différentes approches expérimentales qui semble être une des meilleures façons de maximiser l'extraction d'informations génétiques utiles tout en minimisant l'énergie investie. Toutes ces analyses ont donc contribué à une meilleure caractérisation de la réponse d'*A. thaliana* au TuMV permettant d'identifier de nouveaux loci potentiellement impliqués dans cette interactions en conditions extérieures et en conditions contrôlées.

Vers une validation des candidats et une contribution à la durabilité des résistances

Les différentes approches génétiques développées, sur différentes population d'*A. thaliana*, en conditions extérieures et en condition contrôlées, ont permis d'identifier plusieurs nouveaux loci potentiellement impliqués dans l'interaction *A. thaliana*/TuMV. Parmi les différents candidats identifiés, des loci co-localisant avec le gène *RTM3* ont été identifiés via les analyses GWA sur les populations mondiales et françaises. Ce gène a déjà été caractérisé comme impliqué dans l'interaction d'*A. thaliana* avec d'autres potyvirus en conditions contrôlées (Whitham *et al.*, 1999; Decroocq *et al.*, 2006; Cosson *et al.*, 2010). Ce gène code pour une protéine avec un domaine TRAF, or dans les régions génomiques identifiées, se trouve un cluster de 11 gènes codant pour des protéines avec des domaines TRAF. Cela voudrait donc dire, qu'un ou plusieurs des gènes de ce cluster dont *RTM3* fait partie,

pourrai(en)t être impliqué(s) dans un mécanisme généraliste de la réponse aux potyvirus et ce indépendamment des conditions environnementales. Cependant, même si RTM3 semble impliqué dans le blocage du mouvement en systémie du virus, sa fonction précise n'a à ce jour pas été décrite ni dans son rôle vis-à-vis du virus, ni dans son rôle cellulaire (Cosson *et al.*, 2010). Le criblage d'une banque de cDNA de cellules compagnes d'*A. thaliana* avec les protéines de capsid du *Lettuce mosaic virus* (LMV) et les différentes protéines RTM, ont permis d'identifier 14 protéines dont 2 interagissent avec les protéines de capsid et les protéines RTM, suggérant un complexe multiprotéique dont la localisation cellulaire serait à la périphérie des chloroplastes (Sofer *et al.*, 2017). Les travaux sur *RTM3* seraient intéressants à poursuivre. Une recherche d'orthologues chez des espèces d'intérêt agronomique permettrait également d'évaluer leur intérêt.

L'ensemble des autres loci identifiés, montre qu'il est possible d'apporter de nouvelles connaissances qui pourraient potentiellement constituer de nouvelles sources de résistances vis-à-vis des maladies virales. Le projet Plant KBBE COBRA-2013-2017, coordonné par V. Decroocq (A COmbination of systems Biology and experimental high-throughput approaches to engineer durable Resistance against plAnt viruses in crops) s'intègre parfaitement dans cette problématique puisque son objectif était d'intensifier les recherches appliquées et fondamentales concernant les interactions plantes/virus afin de diversifier les gènes cibles qui pourraient être pyramidés afin de s'adapter aux contraintes de production. Lors de la réunion de clôture du projet, une des conclusions a été de dire qu'aucun nouveau facteur de l'hôte original n'avait été identifié. Des réflexions autour de ce constat ont permis de mettre en évidence le fait que souvent ce sont les mêmes approches qui ont été développées en utilisant les mêmes traits pour caractériser les réponses. La question s'est également posée en ce qui concerne la limitation des ressources naturelles et le type de pathosystème étudié. Ce travail de thèse permet ainsi de montrer qu'en essayant de changer certaines conditions d'étude, l'accès à de nouvelles connaissances est possible.

Il faudrait à présent pouvoir aller plus loin dans l'identification et validation de certains de ces gènes candidats. L'utilisation des mutants T-DNA KO d'*A. thaliana* avec un fond sensible a montré ses limites. Par ailleurs, l'utilisation de mutants n'est pas toujours possible, puisque les gènes identifiés peuvent affecter des fonctions importantes pour le développement, fonctionnement de la plante. C'est notamment le cas du gène codant pour la phosphoglycérate kinase chloroplastique 2, cPGK2, montré comme impliqué dans l'interaction *A. thaliana*/*Watermelon mosaic virus* (WMV). Le mutant KO n'étant pas viable, la validation fonctionnelle a été effectuée chez *N. benthamiana* qui est un hôte compatible pour le WMV et pour lequel un orthologue de cPGK2 a été identifié. L'utilisation de la méthode du VIGS (« virus-induced gene silencing ») a permis de supprimer l'expression du gène codant pour la cPGK chez *N. benthamiana*. La diminution de l'expression de ce gène corrèle avec une diminution de l'accumulation du WMV. Cela a permis de démontrer l'implication de cPGK dans l'infection par le WMV (Ouibrahim *et al.*, 2014).

Un travail de bio-informatique, sur les différents loci identifiés dans nos différentes analyses, semble nécessaire. Il permettrait de comparer les séquences entre accessions sensibles et

résistantes dans le but de pouvoir éventuellement identifier des haplotypes de résistance ou de sensibilité. Cela permettrait de sélectionner de façon plus spécifique des gènes candidats de la réponse au TuMV. Des analyses de validation fonctionnelle pourraient de ce fait être développées, sur ces gènes candidats identifiés, dans le but de valider l'implication de ces gènes dans la réponse au virus. Cette validation fonctionnelle pourrait passer par la mise en place d'analyses de complémentation. Cette méthode a d'ailleurs aussi été utilisée afin de valider l'implication de *cPGK2* dans l'interaction *A. thaliana*/WMV. En effet, l'accession d'*A. thaliana* résistante au WMV (Cvi-0) a été complétement avec le gène codant pour la *cPGK2* de l'accession sensible, Ler-1. Les plantes de Cvi-0 complémentées ont une sensibilité au WMV identique à celle de l'accession sensible (Ouibrahim *et al.*, 2014). Plus récemment, l'utilisation de CRISPR-Cas9 offre une nouvelle opportunité de caractériser fonctionnellement des gènes candidats chez les plantes (Liu *et al.*, 2017). Ce système a déjà été utilisé de façon efficace chez les plantes dans le cadre d'interactions avec des agents pathogènes. Ainsi, CRISPR-Cas9 a été utilisé pour invalider le gène *eIF(iso)4E* dont la protéine est impliquée dans le complexe d'initiation à la traduction utilisé par les virus pour leur réplication. L'invalidation de ce gène chez *Cucumis sativus*, lui confère la résistance à trois virus différents (Chandrasekaran *et al.*, 2016; Andolfo *et al.*, 2016).

L'identification de ces nouveaux loci potentiellement impliqués dans l'interaction *A. thaliana*/TuMV, s'inscrit dans un contexte plus large où la compréhension des interactions plantes/virus est indispensable dans le but de garantir la durabilité des résistances. La durabilité des résistances peut être gérée à différentes échelles. En général, la gestion de cette durabilité s'effectue à travers un management spatio-temporel du déploiement des variétés portant les gènes de résistance. Cependant, cette gestion peut également s'effectuer à l'échelle du génotype lors du développement du cultivar (Quenouille *et al.*, 2014). Des études ont ainsi montré qu'une architecture génétique complexe associant des loci majeurs de résistance et un fond génétique particulier permettait d'augmenter la durabilité des résistances. Le fond génétique des variétés développées semble donc avoir un rôle non négligeable dans le mécanisme de durabilité (Quenouille *et al.*, 2014).

La durabilité des résistances a longtemps été associée à une vision très dichotomique où pyramidage/mélange des cultures ainsi que résistances qualitatives/quantitatives s'opposaient en terme de stratégies de déploiement et de mécanismes de résistances respectivement (Mundt, 2014). Cette vision très binaire de la durabilité des résistances a cependant évolué, puisque maintenant toutes les approches ayant un impact sur la durabilité sont prises en considération de manière à pouvoir les utiliser seules ou en combinaison avec d'autres stratégies en fonction des circonstances. Le champ des recherches porte désormais sur l'étude d'une plus grande diversité d'espèces de plantes pérennes et annuelles confrontées à divers agents pathogènes. De plus, les recherches sur la durabilité des résistances intègrent plus largement des informations provenant de diverses disciplines/domaines (Mundt, 2014). En ce sens, les résultats en lien avec les analyses métaboliques, pourraient être utilisés pour aller rechercher dans des espèces à intérêt

agronomique des idéotypes qui auraient par exemple la capacité d'accumuler les composés identifiés chez les accessions résistantes d'*A. thaliana* en conditions d'infection. Sur ce point, c'est plutôt un fond génétique en particulier qui serait privilégié ou plus précisément la capacité à répondre à une infection virale via la production d'une molécule conférant la résistance. Il aurait été encore plus intéressant s'il avait été possible d'identifier à partir de ces données métaboliques des gènes candidats en lien avec les variations métaboliques observées en réponse au virus. Cependant, il s'est avéré compliqué de caractériser ces bases génétiques par des analyses GWA réalisées à partir des données du contenu en dix métabolites primaires sur 130 accessions d'*A. thaliana*.

De par les différents loci identifiés et les réponses métaboliques fines décrites chez *A. thaliana* en réponse au TuMV, ce travail de thèse a permis d'apporter des connaissances qui pourraient contribuer à une meilleure compréhension des interactions plante/virus, indispensable dans le développement de variétés résistantes et durables. Il est important de garder à l'esprit que la durabilité des résistantes dépend d'un grand nombre de paramètres (spécificité des agents pathogènes, pratiques culturales, types de résistances, stratégies de déploiement des résistances). La complexité de ce mécanisme a conduit à élargir les études à divers types de pathosystèmes et à utiliser les compétences et connaissances de différentes disciplines.

Conclusion

Ce travail de thèse s'est intéressé à étudier le pathosystème naturel *A. thaliana*/TuMV en développant des stratégies complémentaires afin d'essayer d'apporter de nouvelles connaissances permettant de mieux comprendre et donc de mieux lutter contre les infections virales. Si de nombreuses études sont développées en virologie végétale, nous avons essayé de développer des approches nouvelles à travers notamment les différents essais menés en conditions extérieures ou encore à travers le phénotypage large comprenant en particulier les analyses métaboliques. Nous avons également exploité le fait de travailler avec une plante modèle, ce qui donne accès à de nombreuses ressources génétiques. Au-delà de l'ensemble des données et résultats obtenus, ce travail de thèse constitue un bon exemple illustrant le grand intérêt qu'il y a à développer des approches multidisciplinaires dans la compréhension d'une réponse biologique. Une vision plus globale de la réponse biologique peut ainsi être mise en évidence. Elle permet également de se rapprocher au mieux des conditions des plantes au champ. L'objectif – à ne pas perdre de vue – étant d'apporter des réponses aux problématiques de notre agriculture. Ce travail de thèse souligne également l'importance des collaborations entre équipes dans le travail de recherche. Dans cette idée de développement d'études plus globales, les collaborations permettent d'exploiter et unir les compétences et connaissances de chacun. La complexité des défis auxquels l'agriculture d'aujourd'hui et demain doit faire face implique des systèmes d'étude qui tiennent compte de plusieurs facteurs en même temps. Les approches transdisciplinaires doivent être dans ce contexte favorisées. Si « la recherche agronomique au sens large n'était plus prioritaire ni pour les professionnels ni pour les citoyens européens à la fin du XXe siècle. Elle est désormais interpellée pour accompagner l'agriculture dans ses mutations et contribuer à une sécurité alimentaire et nutritionnelle durable » (*Document d'orientations de l'INRA – Science pour l'impact*).

- Matériels et Méthodes -

Une **population locale française**, la **population TOU-A** a également été étudiée au cours d'un essai mené en conditions extérieures en 2016 (Frachon *et al.*, 2017). Cette population a aussi été mise à disposition par F. Roux. Il s'agit d'une population de l'Est de la France (village de Toulon-sur-Arroux, France). Cette population est composée de 190 accessions dont 76 ont été collectées en 2002 (TOU-A1) et 114 en 2010 (TOU-A6). Cette population a fait l'objet d'un génotypage dense de 1.902.203 SNPs.

Une **population de 165 RILs (Recombinant Inbreed Lines)** issue du croisement Bay-0 X Shahdara (Loudet *et al.*, 2002) a été également utilisée dans le cadre d'une stratégie biparentale de recherche de déterminants génétiques.

Trois **mutants T-DNA KO** (« Knock-Out ») ont été utilisés dans une stratégie de validation de gènes candidats. Les mutants T-DNA KO utilisés sont des lignées SALK dont le fond génétique correspond au génotype sauvage de référence d'*A. thaliana* Col-0 (Tab. M1)

Tableau M1 | Liste des mutants T-DNA d'*A.thaliana* testés

Locus	Nom_Locus	Nom_Mutant	Stock code
AT2G12900	Basic leucine zipper 33 (BZIP33)	N682366	SALK_054897C
AT3G58350	RESTRICTED TEV MOVEMENT (RTM3)	N517845	SALK_017845
AT5G18170	Glutamate dehydrogenase 1 (GDH1)	N860072	SALK_042736

B. Description et caractéristiques des expérimentations

1 Expérimentations menées en conditions extérieures

1.1 Description du design expérimental

Les différents essais menés en extérieur (2013-2014-2015-2016-2017) ont suivis les mêmes étapes du semis à la récolte (Fig. M2). Ils ont également été réalisés à la même période de l'année à partir du mois de mars jusqu'à mi-mai.

Les graines d'*A. thaliana* sont préparées deux jours avant le semis dans une solution d'agar à 0.1% ce qui permet de les maintenir en suspension facilitant de ce fait le semis. Elles sont conservées à 4°C jusqu'au moment du semis.

Le **semis** s'effectue dans une serre froide (non chauffée, non éclairée) afin de garantir une germination la plus homogène possible malgré la grande diversité génétique des accessions utilisées. Les graines sont déposées dans des plateaux de 40 alvéoles remplis avec du terreau professionnel horticole (106S cope – PletraCom) traité au VECTOBAC®. Deux **éclaircissements** sont effectués afin de ne laisser au final qu'une seule plante par alvéole.

Trois semaines après le semis, l'ensemble de l'essai est transféré sous un tunnel ouvert sur les côtés. Cette étape constitue une période d'adaptation pour les plantes avant leur mise sur le terrain. C'est à ce niveau que **l'inoculation mécanique** est réalisée sur des plantes au stade 8-10 feuilles (Boyes *et al.*, 2001).

L'essai est ensuite transféré sur le terrain. Les plateaux sont disposés à même le sol de façon aléatoire. L'essai est ensuite recouvert par un filet insect-proof protégeant les plantes de la pluie et de la grêle. Un arrosage est effectué une fois par semaine en fonction des précipitations. La **récolte** a lieu 13 jours après l'inoculation pour les essais de 2013, 2014, 2015 et 2016. Il s'agit du moment où l'on voit apparaître les premiers symptômes. Pour l'essai en 2017, plusieurs prélèvements ont été effectués : un prélèvement initial (T0) avant inoculation, puis cinq prélèvements sont effectués à 5-7-9-14-21 jours après inoculation. Lors du

prélèvement, l'ensemble de la rosette – au-dessus des feuilles inoculées – est récoltée dans des fioles de scintillation Zinger Analytic Polyvials[®]. Les échantillons prélevés sont plongés dans de l'azote liquide puis conservés à -80°C jusqu'à leur traitement

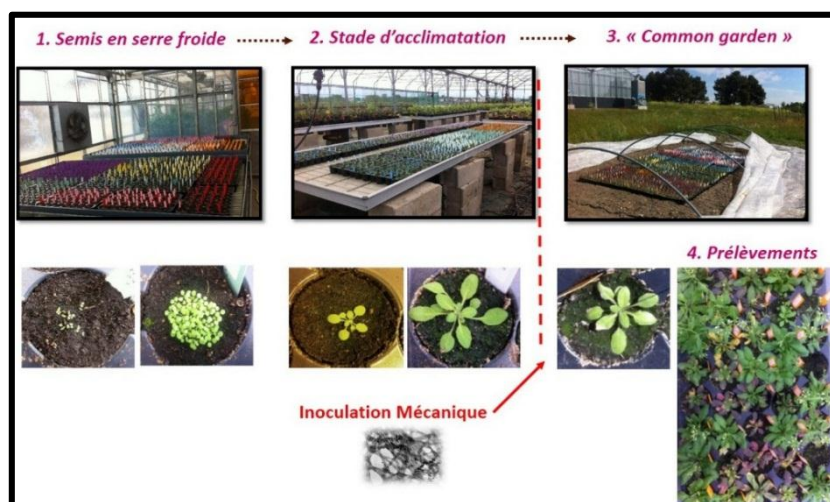


Figure M2| Design expérimental des essais mis en place en conditions extérieures (Photos B.Rubio)

Les grandes étapes de ces essais sont illustrées : **1/**semis en serre froide afin de garantir une germination homogène de l'ensemble des accessions d'*A.thaliana*. Deux éclaircissages sont réalisés afin de garder une seule plante/alvéole **2/**Transfert de l'ensemble de l'essai sous un tunnel ouvert sur les côtés = phase d'acclimatation pendant laquelle l'inoculation mécanique du TuMV est effectuée **3/**Transfert de l'ensemble de l'essai au champ jusqu'au moment du prélèvement

1.2 Spécificités de chaque essai extérieur

Au total, cinq essais ont été menés en conditions extérieures sur cinq années d'étude différentes (2013 – 2014 – 2015 – 2016 – 2017). Le premier essai (2013) a été mis en place avant mon arrivée dans l'équipe. J'ai participé au traitement et exploitation des données. L'essai de 2014 a été mis en place dans le cadre de mon stage de fin d'étude de ma formation d'Ingénieur Agronome. Enfin les trois derniers essais (2015 – 2016 – 2017) ont été menés lors du travail de thèse.

Si les grandes étapes du design expérimental sont identiques quel que soit l'essai, quelques spécificités existent en fonction chaque expérimentation. L'ensemble de ces spécificités sont regroupées dans le tableau M2.

Tableau M2| Bilan des caractéristiques spécifiques de chaque essai mené en conditions extérieures

Année de l'essai	Population d' <i>A.thaliana</i> étudiée	Nbre d'accessions	Date de Prélèvement	Design Experimental				Traits Phénotypiques			
				Nbre de Blocs	Nbre de répétitions	Charge Virale	Symptômes	H. Hampes	Biomasse	Met. 1aires	Met. 2ndaires
2013	Pop. Mondiale	301	13 dpi	3 blocs	1/accession/bloc	X	X	X	-	-	-
2014	Pop. Mondiale	326	13 dpi	3 blocs	1/accession/bloc	X	X	X	X	X	-
2015	Pop. Mondiale	29	13 dpi	4 blocs	4/accession/bloc	X	X	X	X	X	X
2016	Pop. TOU-A FR	190	13 dpi	3 blocs	3/accession/bloc	X	X	X	X	-	-
2017	Pop. Mondiale	12	5-7-9-13-21 dpi	-	9/accession/date	X	X	X	X	X	-

Les principales différences concernent le type de population d'*A. thaliana* étudiée. Ainsi, quatre essais (2013 -2014 -2015 – 2017) ont été conduits avec des accessions de la population mondiale. Un essai (2016) s'est focalisé sur une population locale française, la population TOU-A. Les prélèvements ont tous été effectués en point final à 13 dpi sauf pour l'essai de 2017 où des prélèvements cinétiques ont été menés à 5, 7, 9, 13 et 21 jours après inoculation. La gamme de traits phénotypiques étudiés est plus ou moins importante en fonction des objectifs à atteindre. Ce phénotypage concerne des traits liés à la maladie comprenant **i)** la charge virale mesurée par DAS ELISA, **ii)** les symptômes notés en présence/absence ou selon une échelle de notation, **iii)** la hauteur des hampes (en cm) et **iv)** la biomasse sèche pour les essais de 2014, 2015 et 2016 et la biomasse fraîche pour l'essai de 2017. Le phénotypage

comprend aussi des traits métaboliques via le dosage de métabolites ciblés (2014, 2015 et 2017) et non ciblés (2015).

2 Expérimentation en conditions contrôlées

Des expérimentations ont également été menées en conditions contrôlées (serre ou chambre de culture). Ainsi, une analyse phénotypique de mutants T-DNA a été menée en serre. L'étude de la population bi-parentale Bay-0 X Shahdara a été menée en chambre de culture et en serre. La description de cette étude est retrouvée dans le matériel et méthode correspondant au chapitre 2 de ce travail de thèse.

Expérimentation sur les mutants T-DNA

Trois mutants T-DNA KO sélectionnés, ont été testés dans le cadre d'une infection par le TuMV menée dans un essai mis en place en serre HC (Haut Confinement – T° moyenne journalière de 20°C – 9h jour). L'ensemble de l'essai a été inoculé mécaniquement avec le TuMV, 1 mois après la date de semis. Des prélèvements à 3-5-7 et 9 dpi sont réalisés sachant que chaque mutant-TDNA est répété 9 fois à chaque date. Le génotype Col-0 et le mutant *eIF(iso)4E* (noté E6) constituent les témoins positif et négatif respectivement de l'inoculation. A chaque date sont prélevées pour chaque mutant d'une part les feuilles inoculées et d'autre part des feuilles en systémie. Pour chaque mutant des pools de plantes sont effectués. En effet, 1 échantillon d'un mutant correspond à un pool de 3 plantes du même mutant. Au total, pour chaque mutant et à chaque date de prélèvement il y a 3 échantillons de feuilles inoculées et 3 échantillons de feuilles en systémie (Fig. M3) La charge virale des échantillons est ensuite quantifiée par Q-PCR

Des notations de symptômes ont également été réalisées en utilisant l'échelle de notation pour le TuMV.

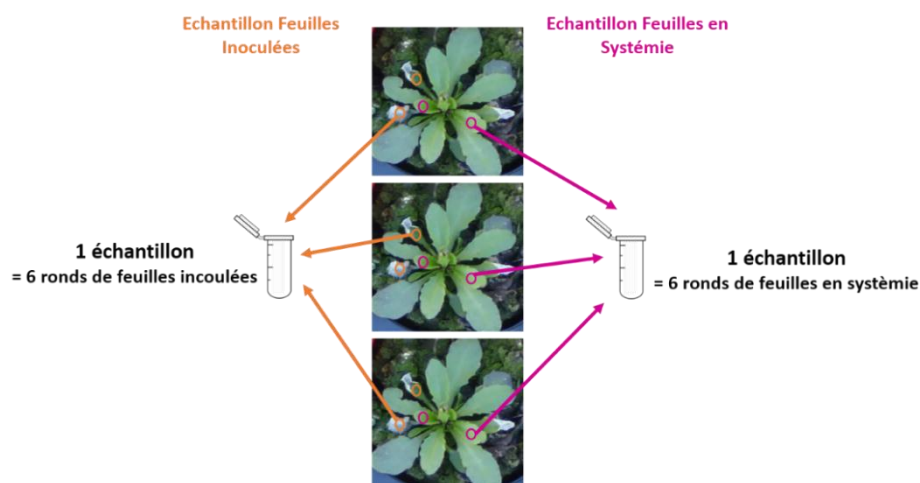


Figure M3 | Prélèvements de feuilles inoculées et en systémie pour chaque mutant analysé

Pour chaque mutant et à chaque date 3 échantillons de feuilles inoculées et 3 échantillons de feuilles en systémie sont prélevés. Chaque échantillon correspond en réalité à un pool de 3 plantes avec deux ronds (17mm de diamètre) de feuilles prélevées/plante à l'emporte-pièce

C. Méthodes Générales

1 Méthodes d'inoculation

1.1 Inoculation Mécanique

L'inoculum utilisé dans les différents essais a toujours été préparé le jour même de l'inoculation à partir de feuilles de navets présentant des symptômes de TuMV broyées dans un tampon d'inoculation (sodium phosphate dibasique dodecahydrate ($\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$) à 30mM + diéca à 0,2%) à raison d'1 g. pour 3 mL. Le broyat est récupéré dans des tubes eppendorf[®] de 2 mL qui sont centrifugés à 4°C, à 13.200 rpm pendant 10 min. L'ensemble des surnageants sont collectés constituant ainsi un inoculum standardisé.

L'inoculation mécanique est réalisée au stade 8-10 feuilles (Boyes *et al.*, 2001). Sur quatre feuilles de chaque rosette d'*A.thaliana*, 20 μL d'inoculum sont déposés (Fig. M4). Elles sont ensuite frottées à l'aide d'une poudre abrasive (carborandum[®]) puis rincées à l'eau.

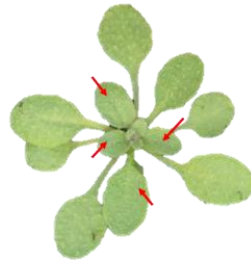


Figure M4| Feuilles inoculées sur une rosette d'*A.thaliana*

Les feuilles inoculées sont signalées par une flèche rouge.

Rosette d'*A.thaliana* 1 mois après semis

1.2 Inoculation par agro-infiltration

La mise en culture et préparation des solutions d'agro-infiltration s'effectuent sur trois jours consécutifs.

Jour 1 : Préparation des pré-cultures

Une pré-culture est préparée dans 5 mL de LB (Luria-Bertani medium, MP Biomedicals, LLC: 2% bactotryptone, 0.5% extrait de levure, 10 mM NaCl, 2.5 mM KCl) dans lequel sont rajoutés 100 μL d'un stock glycérolé de la souche TuMV-UK1-GFP conservée à -80°C avec de la Kanamycine 25 $\mu\text{g}/\text{mL}$. Cette pré-culture est mise à incuber 24h à 28°C à 200 rpm.

Jour 2 : Début des cultures d'*A. tumefaciens*

Transfert des 5 mL de pré-culture dans un Erlen contenant 100 mL de LB, 20 μL d'acétosyringone 0.1 M (3,5-Dimethoxy-4-hydroxyacetophenone) et 200 μL de Kanamycine 25 $\mu\text{g}/\text{mL}$. Incubation 24h à 28°C à 200 rpm.

Jour 3 : Préparation de la solution d'agro-infiltration et agro-infiltration des plantes

Centrifugation des cultures pendant 5 min à 3000 g. Le culot bactérien obtenu est rincé à l'eau stérile puis centrifugé à nouveau pendant 5 min à 3000 g. Le culot obtenu est re-suspendu dans une solution d'agro-infiltration (10 mM MES, 2-morphinoethanesulfonic acid, pH 6.3, 10 mM MgCl_2 et 150 μM d'acétosyringone). La densité optique (DO) à 600nm de la solution d'agro-infiltration est ajustée à 0.8 avec la solution d'agro-infiltration. Incubation de la solution préparée pendant 3h à température ambiante et à l'obscurité.

L'agro-infiltration s'effectue sur des plantes au stade 8-10 feuilles (Boyes *et al.*, 2001). A l'aide d'une seringue, 4 feuilles/rosette sont agro-infiltrées sur la face inférieure. La seringue est placée en bout de feuille, la solution d'agro-infiltration diffuse ainsi sur l'ensemble de la feuille par capillarité.

2 Notation de Symptômes

Des notations de symptômes ont été effectuées de deux manières différentes. Soit il s'agissait d'une simple notation traduisant la présence/absence de symptômes notée 1, pour la présence et 0, pour l'absence. Soit il s'agissait d'une notation en utilisant une échelle de notation allant de 0, aucun symptôme à 4, présence de symptômes sévères (Fig. M5). Cette échelle de notation tient donc compte de la sévérité des symptômes. A partir de ces notations des calculs d'AUDPC (Area Under the Disease Progress Curve) ont été effectués à partir de la formule suivante :

$$\sum_{i=1}^{N_i-1} \frac{symp_i + symp_{i+1}}{2} (t_{i+1} - t_i)$$

où N_i est le nombre de dates de mesure, t_i la date de mesure et $symp_i$ la valeur du score de symptôme notée à partir de l'échelle de notation.



Figure M5 | Echelle de notation de symptômes d'*A. thaliana* infectée par le TuMV

3 Evaluation de la charge virale

3.1 Dosage immunoenzymatique (ELISA semi-quantitatif)

L'accumulation du TuMV en systémie est évaluée par un test de dosage immunoenzymatique DAS-ELISA (Double Antibody Sandwich-Enzyme Linked ImmunoSorbent Assay) en utilisant le kit PathoScreen® développé par Agdia®. Ce test se base sur la reconnaissance spécifique des protéines de la capsid par des anticorps monoclonaux (PTY1) développés par les laboratoires de l'USDA Florist and Nursery Crops à Beltsville (Maryland, USA).

Dans un premier temps, les échantillons préalablement broyés sont dilués au 1/3 avec une solution d'extraction (« Indirect samples extraction buffer 1X » : sodium carbonate 15 mM + sodium bicarbonate 35 mM + polyvinylpyrrolidone (PVP40) 2%). Les échantillons sont déposés (en doublons – 100 µL/puits) dans une plaque ELISA de 96 puits. Une incubation de 1h à température ambiante est réalisée avant d'effectuer 8 lavages avec du tampon PBS (Phosphate buffered saline)-Tween 10X (139.9 mM NaCl, 1.47 mM KH₂PO₄, 2.68 mM KCl, 8.1 mM Na₂HPO₄ et 0.05% v/v Tween 20). L'anticorps dilué au 1:200 dans du tampon « ECI buffer » contenant de l'ovalbumine (tampon PBS-Tween 10X + albumine from chicken egg white 0.2%, PVP-40 2%). Il est déposé à raison de 100 µL/puits. Les plaques sont mises à incuber toute la nuit à 4°C. Le lendemain, les plaques sont lavées 8 fois avec le tampon PBS-Tween 1X. L'anticorps conjugué à la phosphatase alcaline est dilué au 1:200 dans du tampon « ECI buffer », déposé à raison de 100 µL/puits. Une incubation de 2h est réalisée à 37°C. Les plaques sont à nouveau lavées 8 fois avant de déposer 100 µL/puits de pNPP (p-nitrophenyl phosphate) dilué à raison de 2 pastilles dans du tampon contenant de la diéthanolamine 9.7%. Les plaques sont incubées pendant 15 min à température ambiante à l'obscurité. La coloration jaune due à la transformation du substrat en para-nitrophénol par l'activité de la phosphatase alcaline est mesurée à 405 nm à l'aide d'un spectrophotomètre multiplaque EPOCH (Biotek®). Sur l'ensemble des plaques testées, des échantillons d'*A. thaliana* Col-0 (Ecotype *Columbia-0*) sains ainsi que des échantillons d'*A. thaliana* Col-0 infectés par le TuMV ont été déposés. Un traitement des DO (densité optique) obtenues pour chaque échantillon est effectué plaque à plaque. Ainsi, dans un premier temps, la valeur moyenne de DO des échantillons d'*A. thaliana* Col-0 sains est soustraite. Une normalisation est ensuite effectuée à partir des valeurs de DO moyennes obtenues pour les échantillons d'*A. thaliana* Col-0 infectés par le TuMV.

Le résultat de ces tests ELISA comporte de ce fait trois résultats possibles où la « DO sain » correspond à la valeur de DO de l'échantillon *A. thaliana* Col-0 sain (Tab. M3)

Tableau M3| Seuil pour les résultats des analyses ELISA

Echantillon Négatif	Charge Virale Moyenne	Charge Virale Forte
DO < DO sain	DO sain < DO < 2*DO sain	DO > 2 * DO sain

3.2 PCR quantitative en temps réel (Q-PCR)

La quantification de la charge virale par Q-PCR nécessite au préalable l'extraction des ARN viraux effectuée à l'aide du kit NucleoSpin® RNA Plant de chez Macherey-Nagel. Après extraction, une étape de purification des ARN est effectuée à l'aide du kit TURBO DNA-free TM (Ambion®). La synthèse des ADNc est réalisée grâce à la RevertAid H Minus Reverse Transcriptase (200 U/µL - ThermoFisher Scientific™). C'est à partir de ces ADNc que sera

réalisée la quantification virale par Q-PCR. L'ensemble de ces méthodes sont détaillées ci-dessous.

3.2.1 Extraction des ARN totaux en utilisant le kit NucleoSpin® RNA Plant (Macherey-Nagel)

Les extractions ARN effectuées à partir de ce kit ont été réalisées à partir de matériel végétal frais (échantillons conservés à -80°C). Il s'agissait plus précisément de disques foliaires d'*A. thaliana* pouvant atteindre un poids maximal d'environ 100 mg et ayant été broyés à froid (Vibro-broyeur Retsch – 2*30 sec – 30 mouvements/sec). Ajout de 350 µL de tampon RA1 (Guanidium thiocyanate 30-60%) + 3.5 µL de β-mercaptoéthanol/tube. Les échantillons sont vortexés puis centrifugés 1 min à 11.000 rcf à 4°C. Le surnageant de chaque échantillon est transféré dans une colonne de filtration (NucleoSpin®Filter) puis centrifugé 2 min à 11.000 rcf à 4°C. Récupération du filtrat auquel sont ajoutés 350 µL d'éthanol 70%. Après homogénéisation, transfert dans une colonne de fixation (NucleoSpin®RNA Plant Column). Centrifugation 30 sec à 11.000 rcf à 4°C. Ajouter 350 µL/colonne de tampon MDB (Membrane Desalting Buffer – Guanidinium thiocyanate 1-15% + éthanol 5-20%). Centrifugation 1 min à 11.000 rcf à 4°C. Réalisation à cette étape d'un traitement à la DNase (rDNase lyophilisée + 550 µL H₂O RNA-free) pendant 15 min à température ambiante. Ajout de 200 µL de tampon RAW2 (Guanidine hydrochloride 50-66%) puis centrifugation 30 sec à 11.000 rcf à 4°C. Ajout de 600 µL de tampon de lavage RA3 (RA3+ 100 mL ethanol absolute) puis centrifugation 30 sec à 11.000 rcf à 4°C. Ajout à nouveau de 250 µL du tampon de lavage RA3 puis centrifugation 2 min à 11.000 rcf à 4°C. L'élution des ARN en ajoutant 40 µL d'H₂O RNA-free.

3.2.2 Traitement des échantillons à la DNase : utilisation du kit Turbo DNA-free™

Le traitement de la totalité des ARN extraits s'effectue en utilisant le kit TURBO DNA-free™ (Ambion®) qui permet d'éliminer d'éventuelles contaminations ADN lors de l'extraction des ARN totaux. Pour cela, ajout dans un premier temps de 1 µL/échantillon de TURBO DNase + 0.1 volume de tampon TURBO DNase Buffer 10X. Mélanger délicatement par inversion puis incubation 30 min à 37°C. Ajout de 5 µL/échantillon du tampon d'inactivation DNase Inactivation Reagent (0.1 volume) et bien mélanger. Incubation 2 min à température ambiante en agitant de temps en temps puis centrifugation 5 min à 5400 g. Réalisation de la quantification des ARN totaux extraits, en déposant 2 µL de chaque échantillon sur une plaque Take3™ puis lecture dans le lecteur de microplaques EPOCH (Biotek®). L'ensemble des échantillons sont ensuite ramenés à une concentration de 50 ng/µL avec de l'eau RNA free.

3.2.3 Synthèse des ADNc en utilisant le RevertAid H Minus Reverse Transcriptase (ThermoScientific™)

Un premier mix réactionnel est réalisé à partir de 10 µL d'ARN totaux à 50 ng/µL, 1 µL d'amorce oligodT(18) à 100 µM ajusté à 12.5 µL avec de l'H₂O RNA-free. Incubation de ce mélange 5 min à 65°C dans un thermocycleur puis les échantillons sont maintenus dans la glace. Ajout d'un deuxième mix réactionnel contenant 4 µL de tampon 5X RT (fourni dans le kit), 2 µL de dNTP à 10mM, 0.5 µL de RiboLock RNase Inhibitor (40 U/µL), 1 µL de reverse-transcriptase RevertAid H Minus (200 U/µL). Les échantillons sont placés dans un thermocycleur à 42°C pendant 1h

puis à 70°C pendant 10 min. Les ADNc obtenus sont dilués au 1/10^{ème} avant d'effectuer la Q-PCR (5 µL d'ADNc + 45 µL d'H₂O RNA-free).

3.2.4 Quantification de la charge virale par Q-PCR

La charge virale de chaque échantillon a été mesurée par une technique de PCR en temps réel (Q-PCR) en utilisant du Sybr Green I. Le Sybr Green I est une molécule intercalante de l'ADN double brin. L'émission de la fluorescence est mesurée à la fin de chaque étape d'élongation, par un système de lecture intégré à l'appareil de PCR de temps réel. Au cours des cycles, la fluorescence émise augmente proportionnellement au nombre de copies d'ADN amplifiées. Le mélange de Q-PCR se compose pour un échantillon de 5 µL d'ADNc, 10 µL de LightCycler® 480 SYBR Green I Master (Roche Applied Science : FastStart Taq DNA Polymerase, tampon de réaction, dNTPs, du SYBR Green I, MgCl₂), 0,6 µL d'amorce sens (10 µM), 0,6 µL d'amorce anti-sens (10 µM). Les amorces utilisées CP1F/CP1R sont spécifiques de la protéine de capsid de TuMV. Un gène de ménage, le gène de l'ubiquitine (AT2G36060) est utilisé comme contrôle de la Q-PCR (Tab. M4).

Tableau M4 | Caractéristique des amorces utilisées pour la quantification virale par Q-PCR

Amorces	Amorce sens	Amorce anti-sens
CP1 : Amplification de la capsid	TCAAACCATGTGTCAAACCTGC	CGAGAAAAGAGTGGCCCTAA
C34260 : Amplification du gène l'ubiquitine	CCTACCTTTGAGGGGCTTC	GAAGTCGTGAGACAGCGTT

La Q-PCR est réalisée à l'aide d'un LightCycler® 480 (Roche Applied Science). Le programme de Q-PCR comprend les étapes suivantes :

- Pré-incubation à 95°C pendant 15 min
- 40 cycles :
 - Dénaturation : 95°C pendant 30 sec
 - Hybridation : 59°C pendant 30 sec
 - Elongation : 72°C pendant 30 sec = Etape d'acquisition de la fluorescence
- Courbe de Fusion
 - 95°C pendant 5 sec
 - 59°C pendant 1 min
 - 97°C = Acquisition continue de la fluorescence

4 Vérification des mutants T-DNA KO

Dans le cadre des stratégies de validation fonctionnelle de gènes candidats des mutants d'insertion T-DNA sont largement utilisés en génétique inverse. Ainsi, trois mutants TDNA KO ont été testés. Les graines fournies sont de génération T3 et peuvent être **i)** homozygotes pour l'insertion, **ii)** hétérozygotes pour l'insertion ou **iii)** dépourvues de l'insertion. Avant de pouvoir utiliser ces mutants T-DNA, leur statut (homozygote/hétérozygote) doit être vérifié afin de ne garder que les individus homozygotes pour la mutation à partir desquels un stock de graine sera produit par autofécondation en vue des essais à mettre en place pour les tester en condition d'infection par le TuMV.

4.1 Extraction d'ADN en plaque

Avant extraction, le prélèvement de deux feuilles/mutants a été effectué puis un broyage à froid est réalisé en présence d'une bille d'acier (Vibro-broyeur Retsch – 2*30 sec – 30 mouvements/sec). L'extraction débute par l'ajout de 400 µL/échantillon de tampon d'extraction (200 mM Tris-HCl pH 7.5 + 250 mM NaCl + 25 mM EDTA + 0.5% SDS). Après agitation, centrifugation 20 min à 5500 rpm. Transfert de 200 µL de surnageant dans 200 µL d'isopropanol froid. La précipitation de l'ADN se fait à température ambiante pendant 10 min puis la solution est centrifugée 20 min à 5500 rpm. Le surnageant est éliminé et les culots sont lavés avec 500 µL d'éthanol 70%, séchés puis repris dans 200 µL d'H₂O stérile.

4.2 Amplification de l'ADN par PCR

La vérification du statut de ces mutants fait intervenir 3 amorces différentes :

- une amorce spécifique de l'extrémité gauche de l'ADN-T : « left border », **LBa1** = TGGTTCACGTAGTGGGCCATCG
- une amorce LP : « left primer »
- une amorce RP : « right primer »

Deux réactions PCR sont effectuées en parallèle. La première utilise le couple d'amorces LP/RP et amplifie de façon spécifique l'allèle sauvage. La deuxième utilise le couple d'amorces LBa1/RP afin d'amplifier spécifiquement la partie du gène où a été inséré l'ADN-T. La PCR va donc être réalisée en utilisant ces amorces sur les échantillons de mutants mais également sur des échantillons témoin d'*A. thaliana* Col-0 (Fig. M6). Le mix PCR utilisé se compose pour un échantillon, de 2 µL de tampon de réaction pour la Taq (10 mM Tris-HCl + 50 mM KCl + 1.5 mM MgCl₂ pH 8.3 - Biolabs®), 0.16 µL de dNTP 25 mM, 0.4 µL MgCl₂ 50 mM, 0.4 µL de chaque amorce (10µM), 0.1 µL de Taq polymérase (Biolabs®) et 2 µL d'ADN génomique extrait comme indiqué précédemment dans un volume final de 20 µL. Les amorces LP et RP caractéristiques de chaque mutant testé dans cette thèse sont répertoriées dans le tableau M5 ci-dessous.

Le programme de PCR réalisé comporte les étapes suivantes :

- Dénaturation initiale à 94°C pendant 2 min
- 35 Cycles
 - o Dénaturation à 94°C pendant 30 sec
 - o Hybridation à la température déterminée par le couple d'amorces (varie entre 52°C et 57°C) pendant 30 sec
 - o Elongation à 72°C pendant 30 sec
- Elongation finale à 72°C pendant 10 min

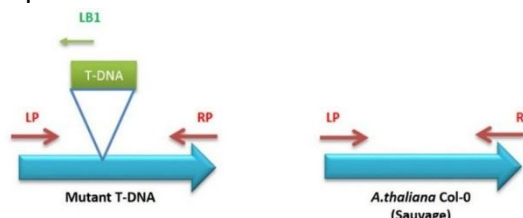


Figure M6 | Couples d'amorces (LP/RP & LBa1/RP) utilisés pour vérifier le statut des mutants T-DNA

Tableau M5 | Liste des mutants testés et des amorces LP et RP correspondantes

Nom_Mutant	Stock code	Locus	Nom_Locus	Amorce LP	Amorce RP
N682366	SALK_054897C	AT2G12900	Basic leucine zipper 33 (BZIP33)	TCTGGGTCTGTTTTGGATCAG	ACCGTGACGATATCGAGATTG
N517845	SALK_017845	AT3G58350	RESTRICTED TEV MOVEMENT (RTM3)	TGAAGGAAGGAGTTGAAAGC	TAGTTTGATTGCAGAGGAAGC
N860072	SALK_042736	AT5G18170	Glutamate dehydrogenase 1 (GDH1)	GATCGCAAGAGTGAGTTTTGC	TATGAACGAGATCTTGGTGG

4.3 Migration sur gel d'agarose

Les acides nucléiques étant chargés négativement, leur migration va s'effectuer dans un champ électrique en direction de l'anode sous une tension comprise entre 90 et 120V. Ils seront donc séparés selon leur taille lors de leur migration sur gel d'agarose à 2% (agarose en poudre, dilué dans du tampon TBE 1X (Tris 89 mM pH 8.3 + Borate 89 mM + EDTA 20 mM). L'ajout de la molécule intercalante entre les bases de l'ADN, le Bromure d'Ethidium (BET 0.5 µg/mL) permet la visualisation des acides nucléiques sous lumière UV. Ainsi, les échantillons sont déposés (5 µL) en présence de 2 µL de tampon de charge Orange (0.5% orange G, 30% glycérol). Cette migration va permettre de distinguer les plantes sauvages (WT), les mutants hétérozygotes (HZ) et homozygotes (HM) pour l'insertion du T-DNA (Fig. M7).

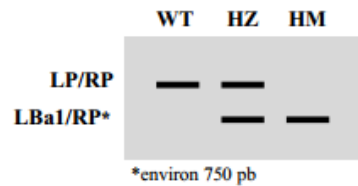


Figure M7 | Détermination du statut des mutants T-DNA testés avec les couples d'amorces LP/RP et LBa1/RP

5 Dosages ciblés des métabolites primaires

L'ensemble des dosages ciblés de métabolites primaires ont été effectués sur la Plateforme Métabolome de Bordeaux en particulier en utilisant le plateau Hit-Me (High throughput Metabolic phenotyping) qui permet de réaliser des quantifications ciblées de métabolites et d'activités enzymatiques en haut débit grâce à l'automatisation d'une grande partie de l'analyse (station de cryobroyage, station robotique de pipetage 96-canaux, lecteurs microplaques). Ainsi le dosage de **12 métabolites primaires** (acides aminés, protéines, glutamate, glutathion, ascorbate, malate, fumarate, glucose, fructose, sucrose, amidon, chlorophylles) a pu être effectué sur des échantillons issus des différents essais menés au cours de ce travail de thèse.

5.1 Préparation des échantillons avant dosage

L'ensemble des échantillons utilisés dans le cadre de ces dosages ciblés ont suivi le même traitement avant extraction. Lors du prélèvement, les échantillons récoltés (ensemble de la rosette au-dessus des feuilles inoculées) dans des fioles de scintillation Zinger Analytic Polyvials[®] ont été broyés à froid (azote liquide) grâce à un broyeur de type 2010 Geno/Grinder[®] à 1750 rpm, 2*30 sec de façon à obtenir une poudre fine et homogène. Des pesées sont ensuite réalisées toujours à froid dans des tubes MICRONIC de 1.1 mL disposés dans des Roborack-96. La quantité de poudre pesée dépend du métabolite dosé, elle peut être soit 17 mg/échantillon (acides aminés, protéines, glutamate, malate, fumarate, glucose, fructose, sucrose, amidon, chlorophylles) soit de 30 mg/échantillon (glutathion, ascorbate). En parallèle aux échantillons à tester, dans chaque Roborack-96, des standards biologiques sont également pesés. Il s'agit de façon générale d'un échantillon d'*A. thaliana* Col-0 spécifique à chaque essai (non infecté, non inoculé). Ce standard est disposé aléatoirement au moins 5 fois dans chaque Roborack-96. De même, des tubes MICRONIC vides sont utilisés pour constituer des blancs.

Les échantillons sont conservés à -80°C jusqu'au moment de l'extraction.

5.2 Extraction éthanolique des échantillons

Une extraction éthanolique est réalisée afin de pouvoir effectuer par la suite le dosage de 10/12 des métabolites primaires analysés (acides aminés, protéines, glutamate, malate, fumarate, glucose, fructose, sucrose, amidon, chlorophylles) à partir des 17 mg de poudre/échantillon pesés. Il s'agit d'une extraction par épuisement en utilisant deux solutions d'éthanol :

- Ethanol 80 % (v/v) + HEPES/KOH 10 mM à pH6
- Ethanol 50 % (v/v) + HEPES/KOH 10 mM à pH6

Remarque : utilisation d'une solution d'HEPES/KOH à 1M, pH6 (acide hydroxyéthyl pipérazine éthane sulfonique/potassium hydroxyle).

Cette extraction utilise la station robotique de pipetage 96-canaux afin d'effectuer les différentes étapes de dépôts et prélèvements successifs décrits ci-dessous :

1. Dépôt de 250 μ L de solution d'éthanol 80% - HEPES/KOH 10 mM pH6
Incubation 20 min au bain-marie à 80°C
Centrifugation 5 min à 4000 rpm à 4°C
Récupération du surnageant dans de nouveaux Roborack-96
2. Dépôt de 150 μ L de solution d'éthanol 80% - HEPES/KOH 10 mM pH6 dans les Roborack-96 contenant les culots
Incubation 20 min au bain-marie à 80°C
Centrifugation 5 min à 4000 rpm à 4°C
Récupération du surnageant dans les Roborack-96 contenant les premiers surnageants déjà prélevés.
3. Dépôt de 250 μ L de solution d'éthanol 50% - HEPES/KOH 10 mM pH6 dans les Roborack-96 contenant les culots
Incubation 20 min au bain-marie à 80°C
Centrifugation 5 min à 4000 rpm à 4°C
Récupération du surnageant dans les Roborack-96 contenant les deux premiers surnageants déjà prélevés

Après extraction conservation des Roborack-96 contenant les culots et ceux contenant les surnageants à -20°C.

5.3 Extraction des échantillons à l'acide trichloroacétique (TCA) 6%

Une extraction au TCA 6% est réalisée afin de pouvoir effectuer par la suite le dosage de 2/12 des métabolites primaires analysés (glutathion et aspartate) à partir des 30 mg de poudre/échantillon pesés. Il s'agit d'une extraction très simple débutant par le dépôt de 400 μ L/échantillon de TCA 6%. Après une agitation vigoureuse des échantillons, centrifugation 25 min à 4°C à 4000 g (speed maxi). Les dosages sont ensuite réalisés sur les surnageants.

5.4 Dosage effectués à partir des échantillons des extractions éthanoliques

Les échantillons obtenus par l'extraction éthanolique permettent d'effectuer les dosages métaboliques des chlorophylles, acides aminés, protéines, glutamate, malate, fumarate, glucose, fructose, saccharose, amidon. Le détail de chacun de ces dosages est décrit ci-dessous

en distinguant d'une part les dosages effectués à partir des surnageants et ceux effectués à partir des culots obtenus lors de l'extraction éthanolique.

5.4.1 Dosages effectués à partir des surnageants

5.4.1.1 Dosage des chlorophylles

Le dosage des chlorophylles est effectué immédiatement après l'extraction. Ce dosage est effectué à partir de 50 µL de chaque échantillon et 120 µL d'éthanol à 98% dans des plaques ELISA. Les lectures sont effectuées dans des lecteurs microplaques SAFAS MP96 à 645 et 665 nm. L'obtention des concentrations en chlorophylles a de chaque échantillon passe par l'utilisation d'une équation empirique (§5.6.1.2).

5.4.1.2 Dosage des acides aminés

Les acides aminés sont dosés à partir de 2 µL de chaque extrait dans un mix réactionnel composé de 15 µL/échantillon de tampon sodium borate 0.1 M pH 8 et de 100 µL/échantillon d'H₂O milliQ. Les mesures sont effectuées dans des plaques ELISA-96 après ajout de 90 µL/échantillon de fluorescamine 0.1% et une incubation de 5 min à température ambiante dans l'obscurité. Les lectures sont réalisées à l'aide du spectrofluorimètre pour microplaques SAFAS Xenius[®] XM. La fluorescamine est une molécule non fluorescente qui après réaction avec des amines primaires forme un produit très fluorescent après excitation par les UV. Les concentrations en acides aminés de chaque extrait sont quantifiées à partir d'une gamme de dilution glutamate 0-100-200-400-800-1000 µM (§ 5.6.1.1): 2 µL de chaque point de dilution sont déposés et répétés quatre fois par dosage.

5.4.1.3 Dosage du glutamate

Le glutamate est dosé à partir de 15 µL de chaque extrait dans un mix réactionnel composé de :

- 133 µL H₂O milliQ
- 20 µL de tricine KOH 1 M pH 8.5
- 5 µL de triton X100 10% (v/v)
- 10 µL de MTT (bromure de 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium) 10 mM
- 10 µL de NAD⁺ (forme oxydée du nicotinamide adénine dinucléotide) 30 mM
- 5 µL de PES (phenazine ethyl sulfate) 2.5 mM
- 2 µL d'ADP (adénosine diphosphate) 50 mM

Les mesures en cinétique sont effectuées dans des plaques ELISA-96 à l'aide des lecteurs de microplaques SAFAS MP96. Une mesure est donc effectuée toutes les minutes à 37°C à 570 nm. Dans un premier temps, les lectures sont réalisées uniquement avec le mix réactionnel et les extraits jusqu'à l'atteinte d'un plateau (environ 10-20 min). Ensuite, la glutamate déhydrogénase (500 U/mL) est ajoutée (2 µL/échantillon). Les lectures sont poursuivies jusqu'à l'atteinte d'un deuxième plateau. La réaction qui se produit lors de l'ajout de l'enzyme est caractéristique du fonctionnement des dehydrogenenase. C'est en réalité le passage de l'état oxydé (incolore – soluble dans l'eau) à réduct (formazan insoluble coloré) du MTT qui est dosé dû à l'apparition d'une coloration violette dont la longueur d'onde d'absorption est à 570nm (Fig. M8). Les concentrations en glutamate de chaque extrait sont quantifiées à partir d'une gamme de dilution glutamate 0-0.1-0.25-0.5-1 mM. Chaque point de dilution est déposé et répété quatre fois par dosage (§ 5.6.1.1).

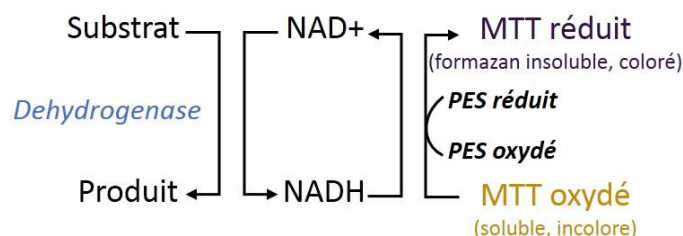


Figure M8 | Schéma réactionnel impliquant une déshydrogénase

5.4.1.4 Dosage du malate/fumarate

Le malate et le fumarate sont dosés dans le cadre d'un même protocole qui nécessite 20 μL de chaque extrait dans un mix réactionnel composé de :

- 50 μL de Tricine/KOH 0.2 M, pH 9
- 10 μL de MTT 10 mM
- 5 μL de PES 2.5 mM
- 10 μL de NAD+ 30 mM
- 5 μL de triton X100 10% (v/v).

Les mesures en cinétique sont effectuées dans des plaques ELISA-96 à l'aide des lecteurs de microplaques SAFAS MP96. Une mesure est donc effectuée toutes les minutes à 37°C à 570 nm. Dans un premier temps, les lectures sont effectuées uniquement avec le mix réactionnel et les extraits jusqu'à l'atteinte d'un plateau (environ 10-20 min). Ensuite, la malate déshydrogénase (1000 U/mL) est ajoutée (2 μL /échantillon). Les lectures sont poursuivies jusqu'à l'atteinte d'un deuxième plateau. Ensuite la fumarase (100 U/mL) est ajoutée (2 μL /échantillon). Les lectures se poursuivent jusqu'à l'atteinte d'un troisième plateau.

Le principe du dosage du malate/fumarate est identique à celui mis en place pour le dosage du glutamate (Fig. M8).

Les concentrations en malate de chaque extrait sont quantifiées à partir d'une gamme de dilution malate 0-0.1-0.25-0.5-1 mM et d'une gamme de dilution fumarate 0-0.1-0.25-0.5-1 mM : 20 μL de chaque point de dilution sont déposés et répétés quatre fois par dosage (§ 5.6.1.1).

5.4.1.5 Dosage des Glucose-Fructose-Saccharose (GFS)

Le dosage de ces trois hexoses est réalisé dans le cadre d'un même protocole qui nécessite 15 μL de chaque extrait dans un mix réactionnel composé de (pour 1 plaque ELISA-96) :

- 15,5 mL d'HEPES/KOH 0.1M, MgCl_2 3 mM, pH7
- 480 μL ATP (adénine triphosphate) 60 mg/mL
- 480 μL NADP (forme réduite du nicotinamide adénine dinucléotide) 36 mg/mL
- 80 μL G6PDH grade II (glucose-6-phosphate déshydrogénase)

Les mesures en cinétique sont effectuées dans des plaques ELISA-96 à l'aide des lecteurs de microplaques SAFAS MP96. Une mesure est donc effectuée toutes les minutes à 37°C à 340 nm. Dans un premier temps, les lectures sont effectuées uniquement avec le mix réactionnel et les extraits jusqu'à l'atteinte d'un plateau (environ 10-20 min). Ajout de 2 μL d'hexokinase (1500 U/mL) puis les lectures sont poursuivies jusqu'à l'atteinte d'un deuxième plateau. Ajout de 2 μL de phosphoglucose isomerase (PGI). Lors de l'atteinte du troisième plateau, ajout de 2 μL d'invertase. La fin du dosage correspond à l'atteinte du quatrième plateau (Fig. M12).

Le dosage des GFS correspond en réalité à un dosage du NADH formé lors des différentes réactions enzymatiques et dont la longueur d'onde d'absorption est de 340nm. Les différentes étapes de dosage sont illustrées sur la Fig. M9.

Les concentrations en glucose, fructose, saccharose sont obtenues en utilisant une équation de calcul qui rend compte de la formation de NADPH proportionnelle à la quantité de GFS de chaque échantillon (§ 5.6.1.2).

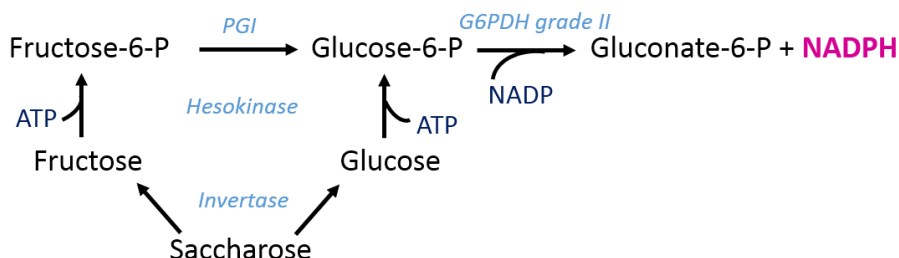


Figure M9| Schéma réactionnel pour le dosage du glucose - fructose - saccharose

5.4.2 Dosages effectués à partir des culots d'extraction

5.4.2.1 Dosage des protéines

La méthode utilisée pour doser les protéines est le dosage Bradford (Bradford, 1976). Il s'agit d'un dosage colorimétrique basé changement de couleur du bleu de Coomassie après liaison avec les acides aminés basiques (arginine, histidine, lysine) et les résidus hydrophobes des acides aminés présents dans les protéines. Le changement d'absorbance (passage de 465-470 nm à 595-600 nm) est proportionnel à la quantité de protéines de l'extrait.

Avant d'effectuer ce dosage, préparation des culots par l'ajout de 400 µL/échantillon de NaOH 0.1 M. Incubation 30 min à 95°C puis centrifugation 5 min à 4000 rpm à 4°C. Réalisation du dosage dans des plaques ELISA-96 à l'aide des lecteurs de microplaques SAFAS MP96 à partir de 3 µL de chaque extrait mélangés à 180 µL de solution de Bradford (Sigma-Aldrich®). Incubation 5 min à température ambiante à l'obscurité avant la lecture des plaques à 600 nm. Les concentrations en protéines de chaque extrait sont quantifiées à partir d'une gamme de dilution BSA (Bovine Serum Albumine) 0-80-160-240-320-400 µg/mL (§ 5.6.1.1) : 3 µL de chaque point de dilution sont déposés et répétés quatre fois par dosage.

5.4.2.2 Dosage de l'amidon

Le dosage de l'amidon s'effectue en deux étapes, sur deux jours consécutifs. La première étape (Jour1) correspond à l'hydrolyse de l'amidon. En effet, après le dosage des protéines, le culot d'extraction des extraits est neutralisé avec 80 µL/échantillon d'une solution de HCl 0.5 M + Acetate NaOH 0.1 M – pH 4.9. Ajout du mix de dégradation composé de (pour 25 mL) :

- 3 mL d'amyloglucosidase
- 30 µL d'α-amylase
- Acetate/NaOH 0.5 M, pH4.9

L'ajout de ce mix permet la dégradation enzymatique de l'amidon après une incubation « over-night » à 37°C (Fig. M10).

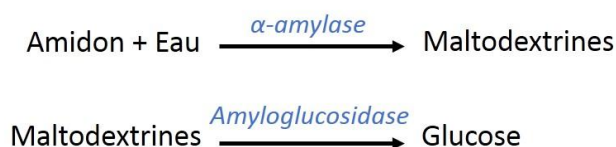


Figure M10 | Schéma réactionnel de la première étape du dosage de l'amidon (dégradation de l'amidon)

La deuxième étape consiste dans le dosage du glucose formé suite à l'hydrolyse de l'amidon. Ce dosage s'effectue sur le même principe que celui décrit pour le glucose avec l'action de l'hexokinase et de la G6PDH grade II. Il s'agit d'une analyse en cinétique effectuée à l'aide des lecteurs de microplaques SAFAS MP96 à 340nm. Après une première étape de plateau avec le mix réactionnel + échantillon, ajout de 2 µL d'hexokinase puis arrêt des mesures lors de l'atteinte d'un deuxième plateau. La concentration en amidon est ensuite obtenue en utilisant la même méthode de calcul que celle des GFS (§ 5.6.1.2).

5.5 Dosage effectués à partir des échantillons des extractions au TCA 6%

Les échantillons obtenus (surnagés) par l'extraction au TCA 6% permettent d'effectuer les dosages métaboliques du glutathion et de l'ascorbate dont le détail des protocoles est présenté ci-dessous.

5.5.1 Dosage du glutathion

Le glutathion est dosé à partir de 10 µL de chaque extrait dans 150 µL/échantillon du mix réactionnel composé de (pour une plaque ELISA-96) :

- 16 mL Buffer 1X = 5.2 mL de tampon phosphate 0.5 M, pH 7 + 0.24 mL EDTA 0.5 M, pH8 + 18.56 H₂O milliQ
- 460 µL de glutathione reductase (préparée dans du Buffer 1X)
- 460 µL de DTNB (acide 5,5'-dithio-bis-(2-nitrobenzoïque)) (1.5mg/mL dilué dans du DMSO – diméthylsulfoxyde)

Le dosage des échantillons s'effectue dans des plaques ELISA-96 à l'aide des lecteurs de microplaques SAFAS MP96 après 5 min d'incubation des plaques à température ambiante. Il s'agit d'une lecture cinétique avec des mesures effectuées toutes les 30 sec à 405nm. Après atteinte d'un plateau (extraits + mix réactionnel), ajout de 5 µL/échantillon de NADPH 2 mM. Suite des lectures jusqu'à l'atteinte d'un deuxième plateau. Les concentrations en glutathion de chaque extrait sont quantifiées à partir d'une gamme de dilution glutathion 0-3.12-6.25-12.5-25-50 µM (§ 5.6.1.1) : 3 µL de chaque point de dilution sont déposés et répétés quatre fois par dosage.

5.5.2 Dosage de l'ascorbate

Le dosage de l'ascorbate s'effectue à partir de 20 µL d'extrait puis ajout de 20 µL/échantillon de DTT 5 mM (dithiothréitol - dilué dans du tampon phosphate pH 7.4). Incubation 20 min à 37°C. Ajout de 10 µL de NEM (N-ethylmaleimide - dilué dans H₂O milliQ) puis bien mélanger et laisser reposer 1 min à T° ambiante. Ajout de 80 µL du mix A + mix B (Tab. M5) puis incubation 40 min à 37°C. Les lectures s'effectuent à 520nm. Les concentrations en ascorbate de chaque extrait sont quantifiées à partir d'une gamme de dilution ascorbate 0-5.05-10.1-15.15-20.2-

30.5 nmoles/ μL (§ 5.6.1.1) : 20 μL de chaque point de dilution sont déposés et répétés quatre fois par dosage.

Tableau M5 | Description de la composition du mix A & B pour le dosage de l'ascorbate

MIX A	H ₃ PO ₄	H ₂ O	TCA	FeCl ₃
1 plaque	2.65 mL	4.5 mL	0.33 g	0.043 g
MIX B	2,2 dipyridil	Ethanol 70%		
1 plaque	0.1 g	2.5 mL		

Proportion (1 plaque) = 6.6 mL de Mix A + 2.4 mL de mix B

5.6 Méthode de détermination des concentrations en métabolites primaires

Après avoir effectué l'ensemble des dosages ciblés des 12 métabolites primaires, les données obtenues sont traitées en utilisant l'application web PLATO DataBase (<https://github.com/cbib/Plato>). Pour cela, il faut dans un premier temps calculer pour chaque métabolite dosé une valeur de standard. Cette valeur est obtenue à partir des échantillons d'*A. thaliana* Col-0 (non infectés – non inoculés) qui ont été placés aléatoirement au moins 5 fois dans chaque Roborack-96 (§ 5.1). Le calcul de ces standards peut être effectué de deux façons différentes en fonction du métabolite primaire analysé.

5.6.1 Obtention des valeurs de standards pour chaque métabolite

5.6.1.1 Calcul des concentrations en métabolites primaires des standards en utilisant une gamme de dilution

Pour 7 /12 des métabolites primaires (acides aminés, protéines, malate, fumarate, glutamate, ascorbate, glutathion) une gamme de dilution a été utilisée lors du dosage. A partir de cette gamme, obtention d'une équation de droite du type $y = ax + b$, où y correspond aux valeurs de densités optique obtenues lors des dosages et x correspond à la concentration en métabolite de chaque échantillon. Grâce à cette équation détermination de la concentration de chaque échantillon à partir des DO obtenues lors des dosages ($x = (y-b)/a$). Cette concentration obtenue correspond en réalité à la concentration contenue dans le volume d'extrait prélevé pour la mesure. Il faut donc effectuer les transformations pour obtenir **la concentration en métabolite exprimée en concentration/g matière fraîche (FW)** en tenant compte de la quantité de poudre pesée pour chaque échantillon (Fig. M11).



Figure M11 | Méthode permettant d'obtenir les concentrations pour chaque métabolite des standards biologiques

5.6.1.2 Calcul des concentrations en métabolites primaires des standards sans gamme de dilution

Pour 5/12 des métabolites (glucose, fructose, saccharose, amidon, chlorophylles), il n'y a pas d'utilisation d'une gamme de dilution lors du dosage. La concentration en métabolites primaires est déterminée à partir de deux équations différentes.

Ainsi, pour obtenir les concentrations en chlorophylles a, utilisation de la formule empirique de Dejana Pankovic : **Chlorophylle a ($\mu\text{g/puits}$) = $5.21A_{665} - 2.07A_{645}$.**

En ce qui concerne les concentrations de glucose, fructose, saccharose et amidon, utilisation de la même équation : **$\mu\text{mol NADPH} = \Delta\text{DO}/(2.85*6.22)$.** Le ΔDO correspond à la différence de densité optique qui existe entre deux plateaux de mesure. Par exemple, pour le dosage du glucose cela correspond à la différence entre le 1^{er} plateau du début (extraits + mix réactionnel) et le 2^{ème} plateau obtenu après l'ajout de l'hexokinase (Fig. M12).

Les données de concentrations obtenues avec ces équations doivent – comme pour les autres métabolites primaires – être corrigées afin d'obtenir au final une concentration en métabolites / gFW.

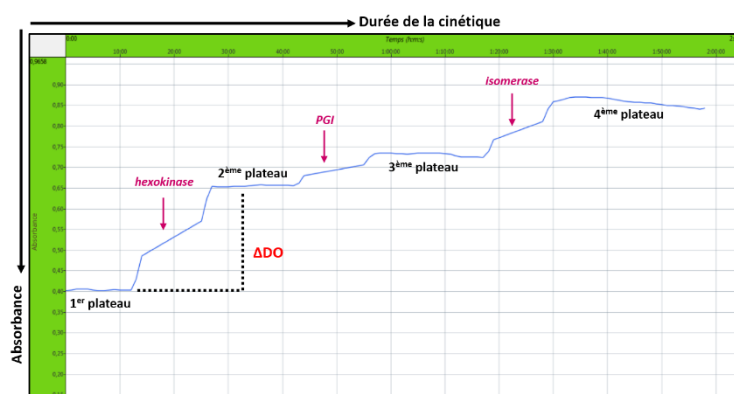


Figure M12 | Principe pour l'obtention des concentrations en métabolites primaires des standards sans gamme de dilution en prenant l'exemple GFS

5.6.2 Utilisation de l'application web PLATO pour déterminer les concentrations en métabolites primaires de chaque échantillon

Après le calcul des standards pour chaque métabolite primaire, il s'agit de déterminer les concentrations en métabolites primaires de chaque échantillon à partir des valeurs brutes mesurées lors des dosages. Cela passe par l'utilisation de l'application web PLATO où les échantillons suivent le même traitement plaque à plaque. Ainsi, la valeur des blancs est enlevée à chaque échantillon puis une normalisation par les standards est effectuée. La variation entre les standards de chaque plaque ne doit pas dépasser 25%. Les données traitées peuvent ensuite être exportées sous la forme d'un classeur excel.

6 Dosage non ciblés des métabolites : Analyses UHPLC-LTQ-Orbitrap-MS

Des dosages non ciblés du contenu en métabolites ont été effectués sur les échantillons des essais extérieurs de 2014 et 2015 déjà utilisés dans le cadre des dosages ciblés de 12 métabolites primaires. Ces analyses ont été effectuées sur la plateforme Metabolome de Bordeaux (BMP) en utilisant à la fois la station robotique de pipetage 96-canaux du plateau

Hit-Me et surtout la chaîne UHPLS-LTQ-Orbitrap-MS (Ultrahigh-performance Liquid Chromatography and Mass Spectrometry).

6.1 Préparation des échantillons avant dosage

L'ensemble des échantillons utilisés dans le cadre de ce dosage non ciblé ont suivi le même traitement avant extraction. Les échantillons avaient déjà été utilisés dans le cadre de dosages ciblés de métabolites primaires, ils étaient donc stockés sous forme de poudre au - 80°C. Les dosages ont été effectués à partir de matière sèche. Les échantillons ont donc dans un premier temps été lyophilisés. Il s'agissait d'une lyophilisation avec une montée de la température par palier de 10°C toutes les 24 heures (-20°C/-10°C/0/10°C/20°C). Après lyophilisation de l'ensemble des échantillons, 10 mg de poudre sèche de chaque échantillon ont été pesés dans des tubes MICRONIC de 1.1 mL. L'ensemble des échantillons ont été disposés dans des Roborack-96.

En parallèle aux échantillons à doser, dans chaque Roborack-96, 3 tubes MICRONIC sont laissés vides pour constituer des blancs. De même, tous les 10 échantillons, un tube MICRONIC est laissé vide afin de déposer par la suite les contrôles qualités (QC).

6.2 Extraction des échantillons

L'extraction des échantillons est effectuée en utilisant une solution d'éthanol 80% - 0.1% d'acide formique. Cette extraction a été effectuée à l'aide de la station robotique de pipetage 96-canaux :

1. Dépôt de 300 µL/échantillon de solution d'éthanol 80% - 0.1% d'acide formique
Sonication pendant 15 min
Centrifugation 5 min à 4°C à 14 000 rpm
Récupération du surnageant dans de nouveaux Roborack-96
2. Dépôt de 300 µL/échantillon de solution d'éthanol 80% - 0.1% d'acide formique
Sonication pendant 15 min
Centrifugation 5 min à 4°C à 14 000 rpm
Récupération à nouveau du surnageant dans les Roborack-96 contenant les surnageants récupérés lors de la première étape d'extraction

Les dosages sont effectués à partir des surnageant récupérés dans des Roborack-96 qui sont stockés à -20°C.

6.3 Filtration des échantillons

Dans le cadre de dosages utilisant la chaîne UHPLS-LTQ-Orbitrap-MS, les échantillons doivent être filtrés afin d'éviter de saturer la colonne avec d'éventuelles impuretés. Utilisation pour cela de plaques de filtration MultiScreen HTS-96 (Merck-Millipore). Il s'agit de filtres de 0,22 µm et 13 mm de diamètre. Cette étape de filtration utilise également la station robotique de pipetage 96-canaux :

1. Humidification des plaques de filtration : dépôt de 80 µL d'éthanol 80% dans les plaques de filtration. Centrifugation 1 min à 1000 g.

2. Filtration des échantillons : dépôt de 150 μL de chaque échantillon dans les plaques de filtration. Centrifugation 1 min à 1000 g. Cette étape est effectuée deux fois de manière à récupérer 300 μL /échantillon du filtrat dans de nouveaux Roborack-96.

6.4 Dépôt du standard interne

Après filtration des échantillons, un standard interne, le **Méthyl vanillate** est rajouté à raison de 2 μL /échantillon. Ce standard interne est préparé à partir de poudre de Méthyl vanillate (Methyl 4-hydroxy-3-methoxybenzoate – Sigma-Aldrich[®]) diluée dans une solution d'éthanol 80% - 0.1% acide formique avec une concentration de 37.5 mg/mL. Cette solution est également filtrée avant le dépôt dans chaque échantillon.

6.5 Dépôt des contrôles qualité (QC)

Dans chaque Roborack-96, dépôt tous les 10 échantillons de 150 μL d'échantillon QC. Ces QC sont obtenus en mélangeant 25 μL de chaque échantillon. Ces QC sont donc représentatifs du contenu en métabolites de l'ensemble des échantillons analysés. Ces échantillons permettent de corriger la dérive instrumentale lors de la séquence d'injection.

6.6 Dosages sur la chaîne UHPLC-LTQ-Orbitrap-MS

6.6.1 Chromatographie Liquide

La chromatographie en phase liquide a été réalisée à l'aide d'un appareil de type Dionex UHPLC Ulyimate 3000 (Thermo Scientific). La séparation des échantillons a été effectuée en phase inverse à l'aide d'une colonne Gemini C 18 (2 x 150 mm ; 3 μm , Phenomenex) équipée avec un système de protection de colonnes (2 x 4 mm, Phenomenex). La phase mobile est constituée d'eau milliQ avec 0.1% d'acide formique (solvant A) et 100% d'acétonitrile (solvant B). Le gradient d'éluion commence avec 3% du solvant B pendant 0.5 min suivi par un gradient linéaire jusqu'à atteindre 10% du solvant B en 0.5 min et un deuxième gradient linéaire de 8 min jusqu'à atteindre 50% du solvant B. Un troisième gradient linéaire permet d'atteindre 100% du solvant B en 4 min puis la phase mobile est maintenue stable pendant 1 min. La phase mobile est ensuite ramenée en 0.5 min à ses conditions de départ (3% B). Elle est maintenue dans cet état pendant 3.5 min en attendant l'analyse de l'échantillon suivant. Le temps total d'analyse pour chaque échantillon est de 18 min. Le débit est de 0.3 mL/min avec une température de la colonne fixée à 30°C. Pendant l'analyse les échantillons sont maintenus à 4°C. Le volume d'injection de chaque échantillon est de 5 μL .

6.6.2 Spectrométrie de Masse

Le système de chromatographie UHPLC est couplé avec spectromètre de masse de type LTQ-Orbitrap Elite (Thermo Scientific). Les analyses sont conduites en mode positif. L'acquisition a été réalisée en « full-scan » avec une résolution de 240 000 FWHM dans une plage de données m/z comprise entre 50 et 1000. L'appareil est équipé d'une source electrospray ESI (Température de chauffage = 350°C ; Température du capillaire = 350°C, Sheath gas 45 (unité arbitraire), Auxiliary gas 15 (unité arbitraire), S-Lens 60 et voltage du spray : 3.2 kV en ESI⁺). L'ensemble des données sont enregistrées à l'aide du logiciel Xcalibur (Thermo Scientific) et extraites avec XCMS.

6.7 Méthode de traitement des données

Après analyses, les fichiers générés par le spectromètre de masse sont transformés au format de type mzXML (Extensible Markup Language) (Deutsch, 2010). Un pré-traitement de l'ensemble de ces données est réalisé à l'aide du package `xcms` de R afin d'obtenir les caractéristiques des pics dans l'espace tridimensionnel défini par la méthode de dosage (temps de rétention, la masse et l'intensité). Ce pré-traitement des données comporte différentes étapes résumées sur la Figure M13 (les principales fonctions du package `xcms` utilisé ont été notées) (Smith *et al.*, 2006). Les paramètres utilisés pour les différentes étapes de traitement sont ceux définis par Patti et ses collaborateurs (Patti *et al.*, 2012). Au final, obtention d'un fichier de format CSV contenant pour chaque échantillon l'ensemble des variables métaboliques identifiées et caractérisées par leur masse, temps de rétention et l'intensité. Différentes étapes de nettoyage sont ensuite effectuées sur ces données afin d'obtenir un fichier final contenant un ensemble de variables avec lesquelles différentes analyses statistiques pourront être menées en fonction des problématiques étudiées. Ainsi, les étapes de nettoyage comprennent **i)** la suppression des pics présents dans les blancs, **ii)** la conservation des pics présents dans au moins la moitié des QC, **iii)** l'élimination des pics dont la variabilité dans les QC est trop importante, pour lesquels le coefficient de variation (CV) est supérieur à 30%, **iv)** la correction en fonction de l'ordre de passage des échantillons (package `ResourceSelection/R`), **v)** la correction en fonction du poids sec exact de chaque échantillon analysé (en grammes), **vi)** l'élimination des pics redondants en utilisant des méthodes de clustering. En effet, des variables présentant des masses et/ou des temps de rétention proches pourraient en réalité correspondre à une seule et même variable. Ainsi, sont considérées comme semblables deux variables d'un même cluster dont le coefficient de corrélation est supérieur à 0.93. Dans ce cas, seule la variable présentant la plus forte intensité est conservée. Après l'ensemble de ces étapes, obtention du fichier de données final contenant l'ensemble des variables caractéristiques des échantillons analysés et sur lesquelles diverses analyses statistiques peuvent être effectuées.

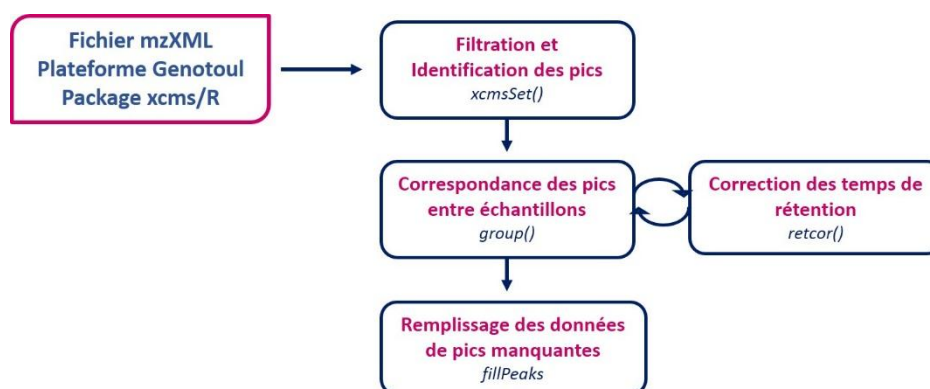


Figure M13 | Différentes étapes du pré-traitement des données mzXML en utilisant le package `xcms` de R

7 Analyses statistiques des données

L'ensemble des données obtenues au cours de ce travail de thèse ont fait l'objet de différentes analyses statistiques à l'aide du **Logiciel R 3.4.0** mais également en utilisant l'**application web BioStatFlow**. Des analyses descriptives ont ainsi été effectuées telles que des analyses en composante principale (ACP) (package R FactoMineR). Afin de confirmer ou infirmer nos observations, différents tests statistiques ont été menés tels que des tests de Student, de Wilcoxon-Mann-Whitney, de Kruskal-Wallis ou encore des ANOVA (« Analysis of Variance »). Le choix de ces tests est fonction de la question biologique posée mais tient compte aussi des caractéristiques des jeux de données analysés en lien notamment avec le respect de la normalité et de l'homocédasticité. Pour aller plus loin, des analyses supervisées de type OPLS-DA (Orthogonal Partial Least Squares – Discriminant Analysis) ont aussi été utilisées notamment dans le cadre de l'analyse des données métaboliques. Les analyses PLS ont été réalisées avec le package R mixOmics. Le calcul de l'héritabilité des différents traits phénotypiques étudiés a été réalisée à l'aide du package lme4. Les analyses de génétique d'association ont été effectuées en utilisant le modèle GEMMA (Zhou & Stephens, 2012), le détail de ces analyses est décrit dans le matériel et méthode de Rubio et *al.*, 2017 (*submitted*). Les analyses QTL menées sont décrites dans le matériel et méthode relatif au chapitre 2 de cette thèse.

Bibliographie

- AbuQamar S, Luo H, Laluk K, Mickelbart MV, Mengiste T. 2009.** Crosstalk between biotic and abiotic stress responses in tomato is mediated by the *AIM1* transcription factor. *The Plant Journal* **58**: 347–360.
- Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KM, Cao J, Chae E, Dezaan TM, Ding W, et al. 2016.** 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**: 481–491.
- Alvarez-Alfageme F, Maharramov J, Carrillo L, Vandenabeele S, Vercammen D, Van Breusegem F, Smaghe G. 2011.** Potential use of a serpin from *Arabidopsis* for pest control. *PLoS One* **6**: e20278.
- Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. 2004.** Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution* **19**: 535–544.
- Andolfo G, Iovieno P, Frusciante L, Ercolano MR. 2016.** Genome-editing technologies for enhancing plant disease resistance. *Frontiers in Plant Science* **7**: 1813–1820.
- Annunziata MG, Apelt F, Carillo P, Krause U, Feil R, Mengin V, Lauxmann MA, Köhl K, Nikoloski Z, Stitt M, et al. 2017.** Getting back to nature: a reality check for experiments in controlled environments. *Journal of Experimental Botany* **68**: 4463–4477.
- Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, et al. 2005.** Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics* **1(5)**: e60.
- Asselbergh B, De Vleeschauwer D, Höfte M. 2008.** Global switches and fine-tuning—ABA modulates plant pathogen defense. *Molecular Plant-Microbe Interactions* **21**: 709–719.
- Atkinson NJ, Lilley CJ, Urwin PE. 2013.** Identification of genes involved in the response of *Arabidopsis* to simultaneous biotic and abiotic stresses. *Plant Physiology* **162**: 2028–2041.
- Atkinson NJ, Urwin PE. 2012.** The interaction of plant biotic and abiotic stresses: from genes to the field. *Journal of Experimental Botany* **63**: 3523–3543.
- Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, et al. 2010.** Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**: 627–631.
- Bartoli C, Roux F. 2017.** Genome-wide association studies in plant pathosystems: toward an ecological genomics approach. *Frontiers in Plant Science* **8**: 763–778.
- Baulcombe D. 2004.** RNA silencing in plants. *Nature* **431**: 356–363.
- Bazakos C, Hanemian M, Trontin C, Jiménez-Gómez JM, Loudet O. 2017.** New strategies and tools in quantitative genetics: how to go from the phenotype to the genotype. *Annual Review of Plant Biology* **68**: 435–455.

- Beauchemin C, Bougie V, Laliberté J-F. 2005.** Simultaneous production of two foreign proteins from a potyvirus-based vector. *Virus Research* **112**: 1–8.
- Bencharki B, Boissinot S, Revollon S, Ziegler-Graff V, Erdinger M, Wiss L, Dinant S, Renard D, Beuve M, Lemaitre-Guillier C, et al. 2010.** Phloem protein partners of *Cucurbit aphid borne yellows virus*: possible involvement of phloem proteins in virus transmission by aphids. *Molecular Plant-Microbe Interactions* **23**: 799–810.
- Berens ML, Berry HM, Mine A, Argueso CT, Tsuda K. 2017.** Evolution of hormone signaling networks in plant defense. *Annual Review of Phytopathology* **55**: 401–425.
- Bergelson J, Roux F. 2010.** Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nature Reviews Genetics* **11**: 867–879.
- Berger S, Sinha AK, Roitsch T. 2007.** Plant physiology meets phytopathology: plant primary metabolism and plant pathogen interactions. *Journal of Experimental Botany* **58**: 4019–4026.
- Bhadauria V (Ed.). 2016.** *Omics in plant disease resistance*. Norwich, UK: Caister Academic Press.
- Bolton MD. 2009.** Primary metabolism and plant defense—fuel for the fire. *Molecular Plant-Microbe Interactions* **22**: 487–497.
- Bomblies K, Yant L, Laitinen RA, Kim S-T, Hollister JD, Warthmann N, Fitz J, Weigel D. 2010.** Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genetics* **6**: e1000890.
- Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Görlach J. 2001.** Growth stage-based phenotypic analysis of *Arabidopsis* a model for high throughput functional genomics in plants. *The Plant Cell* **13**: 1499–1510.
- Brachi B, Villoutreix R, Faure N, Hautekèete N, Piquot Y, Pauwels M, Roby D, Cuguen J, Bergelson J, Roux F. 2013.** Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Molecular Ecology* **22**: 4222–4240.
- Bradford MM. 1976.** A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72**: 248–254.
- Bradshaw WE, Holzapfel CM. 2006.** Evolutionary response to rapid climate change. *Science* **312**: 1477–1478.
- Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD. 2004.** Epistatic interaction between *Arabidopsis* FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 15670–15675.
- Calil IP, Fontes EPB. 2016.** Plant immunity against viruses: antiviral immune receptors in focus. *Annals of Botany* **119**: 711–723.

- Caplan JL, Mamillapalli P, Burch-Smith TM, Czymmek K, Dinesh-Kumar SP. 2008.** Chloroplastic protein NRIP1 mediates innate immune receptor recognition of a viral effector. *Cell* **132**: 449–462.
- Carr T, Whitham SA. 2007.** An emerging model system : *Arabidopsis* as a viral host plant. *Plant Cell Monographs* **7**: 159-183.
- Chan EKF, Rowe HC, Hansen BG, Kliebenstein DJ. 2010.** The complex genetic architecture of the metabolome. *PLoS Genetics* **6**: e1001198.
- Chandrasekaran J, Brumin M, Wolf D, Leibman D, Klap C, Pearlsman M, Sherman A, Arazi T, Gal-On A. 2016.** Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Molecular plant pathology* **17**: 1140–1153.
- Chia DW, Yoder TJ, Reiter W-D, Gibson SI. 2000.** Fumaric acid: an overlooked form of fixed carbon in *Arabidopsis* and other plant species. *Planta* **211**: 743–751.
- Chisholm ST, Coaker G, Day B, Staskawicz BJ. 2006.** Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* **124**: 803–814.
- Chisholm ST, Mahajan SK, Whitham SA, Yamamoto ML, Carrington JC. 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.
- Choi YH. 2004.** Metabolic discrimination of *catharanthus roseus* leaves infected by phytoplasma using ¹H-NMR spectroscopy and multivariate data analysis. *Plant Physiology* **135**: 2398–2410.
- Chuang C, Prasanth KR, Nagy PD. 2017.** The Glycolytic pyruvate kinase is recruited directly into the viral replicase complex to generate ATP for RNA synthesis. *Cell Host & Microbe* **22**: 639–652.
- Chung BN, Choi KS, Ahn JJ, Joa JH, Do KS, Park K-S. 2015.** Effects of temperature on systemic Infection and symptom expression of *Turnip mosaic virus* in Chinese cabbage (*Brassica campestris*). *The Plant Pathology Journal* **31**: 363–370.
- Chung BN, Koh SW, San Choi K, Joa JH, Kim CH, Selvakumar G. 2017.** Temperature and CO₂ level influence *Potato leafroll virus* infection in *Solanum tuberosum*. *The Plant Pathology Journal* **33**: 522-527.
- Chung BY-W, Miller WA, Atkins JF, Firth AE. 2008.** An overlapping essential gene in the Potyviridae. *Proceedings of the National Academy of Sciences* **105**: 5897–5902.
- Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, et al. 2007.** Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**: 338–342.
- Collier SM, Moffett P. 2009.** NB-LRRs work a ‘bait and switch’ on pathogens. *Trends in Plant Science* **14**: 521–529.

- Collum TD, Culver JN. 2016.** The impact of phytohormones on virus infection and disease. *Current Opinion in Virology* **17**: 25–31.
- Coolen S, Proietti S, Hickman R, Davila Olivas NH, Huang P-P, Van Verk MC, Van Pelt JA, Wittenberg AHJ, De Vos M, Prins M, et al. 2016.** Transcriptome dynamics of *Arabidopsis* during sequential biotic and abiotic stresses. *The Plant Journal* **86**: 249–267.
- Cooley MB, Pathirana S, Wu H-J, Kachroo P, Klessig DF. 2000.** Members of the *Arabidopsis* HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens. *The Plant Cell* **12**: 663–676.
- Cosson P, Schurdi-Levraud V, Le QH, Sicard O, Caballero M, Roux F, Le Gall O, Candresse T, 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, Schurdi-Levraud V, Revers F. 2010.** 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**: 1321–1323.
- Decroocq V, Sicard O, Alamillo JM, Lansac M, Eyquard JP, Garcia JA, Candresse T, Le Gall O, Revers F. 2006.** Multiple resistance traits control *Plum pox virus* infection in *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions* **19**: 541–549.
- Deutsch EW. 2010.** Mass spectrometer output file format mzML. *Methods in Molecular Biology* **604**: 319–331.
- Diamond DL, Syder AJ, Jacobs JM, Sorensen CM, Walters K-A, Proll SC, McDermott JE, Gritsenko MA, Zhang Q, Zhao R, et al. 2010.** Temporal proteome and lipidome profiles reveal hepatitis C virus-associated reprogramming of hepatocellular metabolism and bioenergetics. *PLoS Pathogens* **6**: e1000719.
- Dinant S. 2003.** Diversity of the superfamily of phloem lectins (phloem protein 2) in angiosperms. *Plant Physiology* **131**: 114–128.
- Dixon RA. 2001.** Natural products and plant disease resistance. *Nature* **411**: 843–847.
- Dodds PN, Rathjen JP. 2010.** Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics* **11**: 539–548.
- Doerge RW. 2002.** Multifactorial genetics mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics* **3**: 43–52.
- Eu-ahsunthornwattana J, Miller EN, Fakiola M, Wellcome Trust Case Control Consortium 2, Jeronimo SMB, Blackwell JM, Cordell HJ. 2014.** Comparison of methods to account for relatedness in genome-wide association studies with family-based data. *PLoS Genetics* **10**: e1004445.
- Fabre F, Rousseau E, Mailleret L, Moury B. 2012.** Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytologist* **193**: 1064–1075.

- Feldmann KA, Marks MD. 1987.** Agrobacterium-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach. *Molecular and General Genetics* **208**: 1–9.
- Fernandez O, Urrutia M, Bernillon S, Giauffret C, Tardieu F, Le Gouis J, Langlade N, Charcosset A, Moing A, Gibon Y. 2016.** Fortune telling: metabolic markers of plant performance. *Metabolomics* **12**: 158-171.
- Fernandez-Calvino L, Osorio S, Hernandez ML, Hamada IB, del Toro FJ, Donaire L, Yu A, Bustos R, Fernie AR, Martinez-Rivas JM, et al. 2014.** Virus-induced alterations in primary metabolism modulate susceptibility to *Tobacco rattle virus* in *Arabidopsis*. *Plant Physiology* **166**: 1821–1838.
- Fernie AR, Carrari F, Sweetlove LJ. 2004.** Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Current Opinion in Plant Biology* **7**: 254–261.
- Feussner I, Polle A. 2015.** What the transcriptome does not tell — proteomics and metabolomics are closer to the plants' patho-phenotype. *Current Opinion in Plant Biology* **26**: 26–31.
- Findlay JS, Ulaeto D. 2015.** Semliki forest virus and sindbis virus, but not vaccinia virus, require glycolysis for optimal replication. *Journal of General Virology* **96**: 2693–2696.
- Flor HH. 1971.** Current status of the gene-for-gene concept. *Annual Review of Phytopathology* **9**: 275–296.
- Fontaine KA, Sanchez EL, Camarda R, Lagunoff M. 2015.** Dengue virus induces and requires glycolysis for optimal replication. *Journal of Virology* **89**: 2358–2366.
- Frachon L, Libourel C, Villoutreix R, Carrère S, Glorieux C, Huard-Chauveau C, Navascués M, Gay L, Vitalis R, Baron E, et al. 2017.** Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. *Nature Ecology & Evolution* **1**: 1551-1561.
- Fujita Y. 2005.** AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in *Arabidopsis*. *The Plant Cell Online* **17**: 3470–3488.
- Fusari CM, Kooke R, Lauxmann MA, Annunziata MG, Encke B, Hoehne M, Krohn N, Becker FFM, Schlereth A, Sulpice R, et al. 2017.** Genome-wide association mapping reveals that specific and pleiotropic regulatory mechanisms fine-tune central metabolism and growth in *Arabidopsis*. *The Plant Cell*: doi: 10.1105/tpc.17.00232
- Gadjev I. 2006.** Transcriptomic footprints disclose specificity of reactive oxygen species signaling in *Arabidopsis*. *Plant Physiology* **141**: 436–445.
- García-Arenal F, McDonald BA. 2003.** An analysis of the durability of resistance to plant viruses. *Phytopathology* **93**: 941–952.
- Gibbs A, Ohshima K. 2010.** Potyviruses and the digital revolution. *Annual Review of Phytopathology* **48**: 205–223.

- Gibon Y, Dominique R, Catherine D, Stéphane B, Annick M. 2012.** New opportunities in metabolomics and biochemical phenotyping for plant systems biology. *Metabolomics InTech*: 213-240.
- Gomez G, Pallas V. 2004.** A long-distance translocatable phloem protein from cucumber forms a ribonucleoprotein complex *in vivo* with hop stunt viroid RNA. *Journal of Virology* **78**: 10104–10110.
- Gómez P, Rodríguez-Hernández AM, Moury B, Aranda MA. 2009.** Genetic resistance for the sustainable control of plant virus diseases: breeding, mechanisms and durability. *European Journal of Plant Pathology* **125**: 1–22.
- Gouveia BC, Calil IP, Machado JPB, Santos AA, Fontes EPB. 2017.** Immune receptors and co-receptors in antiviral innate immunity in plants. *Frontiers in Microbiology* **7**: 2139.
- Hämäläinen JH, Kekarainen T, Gebhardt C, Watanabe KN, Valkonen JP. 2000.** Recessive and dominant genes interfere with the vascular transport of *Potato virus A* in diploid potatoes. *Molecular Plant-Microbe Interactions* **13**: 402–412.
- Hashimoto M, Neriya Y, Yamaji Y, Namba S. 2016.** Recessive resistance to plant viruses: potential resistance genes beyond translation initiation factors. *Frontiers in Microbiology* **7**: 1695.
- Hayat S, Hayat Q, Alyemeni MN, Wani AS, Pichtel J, Ahmad A. 2012.** Role of proline under changing environments: A review. *Plant Signaling & Behavior* **7**: 1456–1466.
- Heath MC. 2000.** Nonhost resistance and nonspecific plant defenses. *Current Opinion in Plant Biology* **3**: 315–319.
- Heaton NS, Randall G. 2011.** Multifaceted roles for lipids in viral infection. *Trends in Microbiology* **19**: 368–375.
- Herbers K, Meuwly P, Frommer WB, Mettraux J-P, Sonnewald U. 1996.** Systemic acquired resistance mediated by the ectopic expression of invertase: possible hexose sensing in the secretory pathway. *The Plant Cell* **8**: 793–803.
- Hily J-M, Poulicard N, Mora M-Á, Pagán I, García-Arenal F. 2016.** Environment and host genotype determine the outcome of a plant-virus interaction: from antagonism to mutualism. *New Phytologist* **209**: 812–822.
- Hoffmann MH. 2002.** Biogeography of *Arabidopsis thaliana* (l.) heynh.(*Brassicaceae*). *Journal of Biogeography* **29**: 125–134.
- Hoffmann AA, Sgrò CM. 2011.** Climate change and evolutionary adaptation. *Nature* **470**: 479–485.
- Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Mulyati NW, Platt A, Sperone FG, Vilhjálmsson BJ, et al. 2012.** Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics* **44**: 212–216.

- Ingle RA, Carstens M, Denby KJ. 2006.** PAMP recognition and the plant–pathogen arms race. *BioEssays* **28**: 880–889.
- Ishibashi K, 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**: 7933–7939.
- Ishibashi K, Masuda K, Naito S, Meshi T, Ishikawa M. 2007.** An inhibitor of viral RNA replication is encoded by a plant resistance gene. *Proceedings of the National Academy of Sciences* **104**: 13833–13838.
- Ivanov KI, Eskelin K, Lohmus A, Makinen K. 2014.** Molecular and cellular mechanisms underlying potyvirus infection. *Journal of General Virology* **95**: 1415–1429.
- Jänkänpää HJ, Mishra Y, Schröder WP, Jansson S. 2012.** Metabolic profiling reveals metabolic shifts in *Arabidopsis* plants grown under different light conditions: metabolic profiling under different light regime. *Plant, Cell & Environment* **35**: 1824–1836.
- Jenner CE, Walsh JA. 1996.** Pathotypic variation in *Turnip mosaic virus* with special reference to European isolates. *Plant Pathology* **45**: 848–856.
- Jones JDG, Dangl JL. 2006.** The plant immune system. *Nature* **444**: 323–329.
- Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. 2008.** Efficient control of population structure in model organism association mapping. *Genetics* **178**: 1709–1723.
- Kaul P, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito Mal, Lin X, et al. 2000.** Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M. 2007.** Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics* **39**: 1151–1155.
- Kissoudis C, Chowdhury R, van Heusden S, van de Wiel C, Finkers R, Visser RGF, Bai Y, van der Linden G. 2015.** Combined biotic and abiotic stress resistance in tomato. *Euphytica* **202**: 317–332.
- Kliebenstein DJ. 2004.** Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinted glasses. *Plant, Cell & Environment* **27**: 675–684.
- Kliebenstein D. 2009.** Advancing genetic theory and application by metabolic quantitative trait loci analysis. *The Plant Cell Online* **21**: 1637–1646.
- Kogovšek P, Pompe-Novak M, Petek M, Fragner L, Weckwerth W, Gruden K. 2016.** Primary metabolism, phenylpropanoids and antioxidant pathways are regulated in potato as a response to *Potato virus Y* infection. *PLoS One* **11**: e0146135.

- Koornneef M, Alonso-Blanco C, Vreugdenhil D. 2004.** Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology* **55**: 141–172.
- Kopp M, Matuszewski S. 2014.** Rapid evolution of quantitative traits: theoretical perspectives. *Evolutionary Applications* **7**: 169–191.
- Korte A, Farlow A. 2013.** The advantages and limitations of trait analysis with GWAS: a review. *Plant methods* **9**: 29.
- Kumar J, Gupta DS, Gupta S, Dubey S, Gupta P, Kumar S. 2017.** Quantitative trait loci from identification to exploitation for crop improvement. *Plant Cell Reports* **8**:1187-1213.
- Kumar D, Kumar R, Hyun TK, Kim J-Y. 2015.** Cell-to-cell movement of viruses via plasmodesmata. *Journal of Plant Research* **128**: 37–47.
- Lellis AD, Kasschau KD, Whitham SA, Carrington JC. 2002.** Loss-of-susceptibility mutants of *Arabidopsis thaliana* reveal an essential role for eIF (iso) 4E during potyvirus infection. *Current Biology* **12**: 1046–1051.
- Less H, Angelovici R, Tzin V, Galili G. 2011.** Coordinated gene networks regulating *Arabidopsis* plant metabolism in response to various stresses and nutritional cues. *The Plant Cell* **23**: 1264–1271.
- Lima MRM, Felgueiras ML, Graca G, Rodrigues JEA, Barros A, Gil AM, Dias ACP. 2010.** NMR metabolomics of esca disease-affected *Vitis vinifera* cv. Alvarinho leaves. *Journal of Experimental Botany* **61**: 4033–4042.
- Lisec J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Törjék O, Selbig J, Altmann T, et al. 2007.** Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations: metabolic and biomass QTL in *Arabidopsis*. *The Plant Journal* **53**: 960–972.
- Liu X, Wu S, Xu J, Sui C, Wei J. 2017.** Application of CRISPR/Cas9 in plant biology. *Acta Pharmaceutica Sinica B* **7**: 292–302.
- Llave C. 2016.** Dynamic cross-talk between host primary metabolism and viruses during infections in plants. *Current Opinion in Virology* **19**: 50–55.
- López MG, Zanor MI, Pratta GR, Stegmayer G, Boggio SB, Conte M, Bermúdez L, Leskow CC, Rodríguez GR, Picardi LA, et al. 2015.** Metabolic analyses of interspecific tomato recombinant inbred lines for fruit quality improvement. *Metabolomics* **11**: 1416–1431.
- Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F. 2002.** Bay-0 × Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *TAG Theoretical and Applied Genetics* **104**: 1173–1184.
- Luo J. 2015.** Metabolite-based genome-wide association studies in plants. *Current Opinion in Plant Biology* **24**: 31–38.

- Mäkinen K, Hafrén A. 2014.** Intracellular coordination of potyviral RNA functions in infection. *Frontiers in Plant Science* **5**:110.
- Martín Martín Á, y Poch HLC, Herrera DM, Ponz F. 1999.** Resistances to *Turnip mosaic potyvirus* in *Arabidopsis thaliana*. *Molecular Plant-Microbe interactions* **12**: 1016–1021.
- Matros A, Liu G, Hartmann A, Jiang Y, Zhao Y, Wang H, Ebmeyer E, Korzun V, Schachschneider R, Kazman E, et al. 2016.** Genome–metabolite associations revealed low heritability, high genetic complexity, and causal relations for leaf metabolites in winter wheat (*Triticum aestivum*). *Journal of Experimental Botany* **68(3)**: 415–428.
- Matsuda F, Okazaki Y, Oikawa A, Kusano M, Nakabayashi R, Kikuchi J, Yonemaru J-I, Ebana K, Yano M, Saito K. 2012.** Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis: metabolome QTL analysis of rice grain. *The Plant Journal* **70**: 624–636.
- Mengiste T. 2003.** The *Botrytis* susceptible1 gene encodes an R2R3MYB transcription factor protein that is required for biotic and abiotic stress responses in *Arabidopsis*. *The Plant Cell Online* **15**: 2551–2565.
- Meyer RC, Steinfath M, Lisek J, Becher M, Witucka-Wall H, Torjek O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, et al. 2007.** The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **104**: 4759–4764.
- Meyerowitz EM. 2001.** Prehistory and history of *Arabidopsis* research. *Plant physiology* **125**: 15–19.
- Meyers BC. 2003.** Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *The Plant Cell Online* **15**: 809–834.
- Mittler R. 2006.** Abiotic stress, the field environment and stress combination. *Trends in Plant Science* **11**: 15–19.
- Moffett P. 2009.** Mechanisms of recognition in dominant R gene mediated resistance. *Advances in Virus Research* **75**: 1–33.
- Montarry J, Cartier E, Jacquemond M, Palloix A, Moury B. 2012.** Virus adaptation to quantitative plant resistance: erosion or breakdown? *Journal of Evolutionary Biology* **25**: 2242–2252.
- Mundt CC. 2014.** Durable resistance: a key to sustainable management of pathogens and pests. *Infection, Genetics and Evolution* **27**: 446–455.
- van Munster M, Yvon M, Vile D, Dader B, Fereres A, Blanc S. 2017.** Water deficit enhances the transmission of plant viruses by insect vectors. *PLoS One* **12**: e0174398.
- Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES. 2009.** Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell Online* **21**: 2194–2202.

- Mysore KS, Ryu C-M. 2004.** Nonhost resistance: how much do we know? *Trends in Plant Science* **9**: 97–104.
- Nakahara KS, Masuta C. 2014.** Interaction between viral RNA silencing suppressors and host factors in plant immunity. *Current Opinion in Plant Biology* **20**: 88–95.
- Nam M, Bang E, Kwon T, Kim Y, Kim E, Cho K, Park W, Kim B-G, Yoon I. 2015.** Metabolite profiling of diverse rice germplasm and identification of conserved metabolic markers of rice roots in response to long-term mild salinity stress. *International Journal of Molecular Sciences* **16**: 21959–21974.
- Nancarrow N, Constable FE, Finlay KJ, Freeman AJ, Rodoni BC, Trebicki P, Vassiliadis S, Yen AL, Luck JE. 2014.** The effect of elevated temperature on *Barley yellow dwarf virus-PAV* in wheat. *Virus Research* **186**: 97–103.
- Nejat N, Mantri N. 2017.** Plant immune system: crosstalk between responses to biotic and abiotic stresses the missing link in understanding plant defence. *Current Issues in Molecular Biology* **23**: 1–16.
- Nguyen HD, Tomitaka Y, Ho SYW, Duchêne S, Vetten H-J, Lesemann D, Walsh JA, Gibbs AJ, Ohshima K. 2013.** *Turnip mosaic potyvirus* probably first spread to Eurasian brassica crops from wild orchids about 1000 years ago. *PLoS One* **8**: e55336.
- Nicaise V. 2014.** Crop immunity against viruses: outcomes and future challenges. *Frontiers in Plant Science* **5**: 660.
- Ohshima K, Akaishi S, Kajiyama H, Koga R, Gibbs AJ. 2010.** Evolutionary trajectory of *Turnip mosaic virus* populations adapting to a new host. *Journal of General Virology* **91**: 788–801.
- Ohshima K, Yamaguchi Y, Hirota R, Hamamoto T, Tomimura K, Tan Z, Sano T, Azuhata F, Walsh JA, Fletcher J, et al. 2002.** Molecular evolution of *Turnip mosaic virus*: evidence of host adaptation, genetic recombination and geographical spread. *Journal of General Virology* **83**: 1511–1521.
- Orjuela J, Deless EFT, Kolade O, Chéron S, Ghesquière A, 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* **26**: 1455–1463.
- Otim M, Legg DJ, Kyamanywa S, Polaszek A, Gerling D. 2006.** Population dynamics of *Bemisia tabaci* (Homoptera: *Aleyrodidae*) parasitoids on cassava mosaic disease-resistant and susceptible varieties. *Biocontrol Science and Technology* **16**: 205–214.
- Ouibrahim L, Caranta C. 2013.** Exploitation of natural genetic diversity to study plant-virus interactions: what can we learn from *Arabidopsis thaliana*?: Natural virus resistance in *Arabidopsis thaliana*. *Molecular Plant Pathology* **14**: 844–854.
- Ouibrahim L, Mazier M, Estevan J, Pagny G, Decroocq V, Desbiez C, Moretti A, Gallois J-L, 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* **79**: 705–716.

Owens RA, Blackburn M, Ding B. 2001. Possible involvement of the phloem lectin in long-distance viroid movement. *Molecular plant-microbe interactions* **14**: 905–909.

Pagan I, Fraile A, Fernandez-Fueyo E, Montes N, Alonso-Blanco C, Garcia-Arenal F. 2010. *Arabidopsis thaliana* as a model for the study of plant-virus co-evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1983–1995.

Pagny G, Paulstephenraj PS, Poque S, Sicard O, Cosson P, Eyquard J-P, Caballero M, Chague A, Gourdon G, Negrel L, et al. 2012. Family-based linkage and association mapping reveals novel genes affecting *Plum pox virus* infection in *Arabidopsis thaliana*. *New Phytologist* **196**: 873–886.

Palloix A, Ayme V, 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.

Pandey P, Irulappan V, Bagavathiannan MV, Senthil-Kumar M. 2017. Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Frontiers in Plant Science* **8**: 537.

Parlevliet JE. 2002. Durability of resistance against fungal, bacterial and viral pathogens; present situation. *Euphytica* **124**: 147–156.

Patti GJ, Tautenhahn R, Siuzdak G. 2012. Meta-analysis of untargeted metabolomic data from multiple profiling experiments. *Nature Protocols* **7**: 508–516.

Piasecka A, Jedrzejczak-Rey N, Bednarek P. 2015. Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytologist* **206**: 948–964.

Poorter H, Fiorani F, Pieruschka R, Wojciechowski T, van der Putten WH, Kleyer M, Schurr U, Postma J. 2016. Pampered inside, pestered outside? Differences and similarities between plants growing in controlled conditions and in the field. *New Phytologist* **212**: 838–855.

Poque S, Pagny G, Ouibrahim L, Chague A, Eyquard J-P, Caballero M, Candresse T, Caranta C, Mariette S, 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.

Prasch CM, Sonnewald U. 2013. Simultaneous application of heat, drought, and virus to *Arabidopsis* plants reveals significant shifts in signaling networks. *Plant Physiology* **162**: 1849–1866.

Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, Brkljacic J, Browse J, Chapple C, Colot V, Cutler S, et al. 2016. 50 years of *Arabidopsis* research: highlights and future directions. *New Phytologist* **209**: 921–944.

Pumplin N, Voinnet O. 2013. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nature Reviews Microbiology* **11**: 745–760.

- Quenouille J, Paulhiac E, Moury B, Palloix A. 2014.** Quantitative trait loci from the host genetic background modulate the durability of a resistance gene: a rational basis for sustainable resistance breeding in plants. *Heredity* **112**: 579-587.
- Quenouille J, Saint-Felix L, Moury B, Palloix A. 2016.** 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*: genetic backgrounds shape *R*-gene durability. *Molecular Plant Pathology* **17**: 296–302.
- Quistián-Martínez D, Estrada-Luna AA, Altamirano-Hernández J, Peña-Cabriaes JJ, de Oca-Luna RM, Cabrera-Ponce JL. 2012.** Use of trehalose metabolism as a biochemical marker in rice breeding. *Molecular Breeding* **30**: 469–477.
- Rajamäki M-L, Mäki-Valkama T, Mäkinen K, Valkonen J. 2004.** Infection with potyviruses. In: Annual plant reviews, plant–pathogen interactions. Blackwell Publishing Sheffield, UK, 68–91.
- Rasmussen S, Barah P, Suarez-Rodriguez MC, Bressendorff S, Friis P, Costantino P, Bones AM, Nielsen HB, Mundy J. 2013.** Transcriptome responses to combinations of stresses in *Arabidopsis*. *Plant Physiology* **161**: 1783–1794.
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES. 2001.** Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences* **98**: 11479–11484.
- Revers F, García JA. 2015.** Chapter three-molecular biology of potyviruses. *Advances in virus research* **92**: 101–199.
- Revers F, Guiraud T, Houvenaghel M-C, Mauduit T, Le Gall O, Candresse T. 2003.** Multiple resistance phenotypes to *Lettuce mosaic virus* among *Arabidopsis thaliana* accessions. *Molecular Plant-Microbe Interactions* **16**: 608–616.
- Revers F, Nicaise V. 2014.** Plant resistance to infection by viruses. In: John Wiley & Sons Ltd, ed. eLS. Chichester, UK: John Wiley & Sons, Ltd.
- Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE. 2012.** Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics* **44**: 217–220.
- Robaglia C, Caranta C. 2006.** Translation initiation factors: a weak link in plant RNA virus infection. *Trends in Plant Science* **11**: 40–45.
- Roberts TH, Hejgaard J. 2008.** Serpins in plants and green algae. *Functional & Integrative Genomics* **8**: 1–27.
- Rodriguez M-C, Petersen M, Mundy J. 2010.** Mitogen-activated protein kinase signaling in plants. *Annual Review of Plant Biology* **61**: 621–649.

- Rojas CM, Senthil-Kumar M, Tzin V, Mysore KS. 2014.** Regulation of primary plant metabolism during plant-pathogen interactions and its contribution to plant defense. *Frontiers in Plant Science* **5**: 17.
- de Ronde D, Butterbach P, Kormelink R. 2014.** Dominant resistance against plant viruses. *Frontiers in Plant Science* **5**: 307.
- Roossinck MJ. 2012.** Plant virus metagenomics: biodiversity and ecology. *Annual Review of Genetics* **46**: 359–369.
- Roux F, Bergelson J. 2016.** The genetics underlying natural variation in the biotic interactions of *Arabidopsis thaliana*. *Current Topics in Developmental Biology* **119**: 111–156.
- Ruffel S, Dussault M-H, Palloix A, Moury B, Bendahmane A, Robaglia C, 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* **32**: 1067–1075.
- Sade D, Shriki O, Cuadros-Inostroza A, Tohge T, Semel Y, Haviv Y, Willmitzer L, Fernie AR, Czosnek H, Brotman Y. 2015.** Comparative metabolomics and transcriptomics of plant response to *Tomato yellow leaf curl virus* infection in resistant and susceptible tomato cultivars. *Metabolomics* **11**: 81–97.
- Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K. 2002.** DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochemical and Biophysical Research Communications* **290**: 998–1009.
- Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M. 2014.** Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiology* **165**: 1120–1132.
- Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR. 2008.** Mode of inheritance of primary metabolic traits in tomato. *The Plant Cell Online* **20**: 509–523.
- Scheideler M, Schlaich NL, Fellenberg K, Beissbarth T, Hauser NC, Vingron M, Slusarenko AJ, Hoheisel JD. 2002.** Monitoring the switch from housekeeping to pathogen defense metabolism in *Arabidopsis thaliana* using cDNA arrays. *Journal of Biological Chemistry* **277**: 10555–10561.
- Schmid KJ, Törjék O, Meyer R, Schmuths H, Hoffmann MH, Altmann T. 2006.** Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theoretical and Applied Genetics* **112**: 1104–1114.
- Scholthof K-BG. 2007.** The disease triangle: pathogens, the environment and society. *Nature Reviews Microbiology* **5**: 152–156.

Scholthof K-BG, Adkins S, Czosnek H, Palukaitis P, Jacquot E, Hohn T, Hohn B, Saunders K, Candresse T, Ahlquist P, et al. 2011. Top 10 plant viruses in molecular plant pathology: Top 10 plant viruses. *Molecular Plant Pathology* **12**: 938–954.

Sessions A. 2002. A high-throughput *Arabidopsis* reverse genetics system. *The Plant Cell Online* **14**: 2985–2994.

Sharbel TF, Haubold B, Mitchell-Olds T. 2000. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology* **9**: 2109–2118.

Sharma N, Sahu PP, Puranik S, Prasad M. 2013. Recent advances in plant–virus interaction with emphasis on small interfering RNAs (siRNAs). *Molecular Biotechnology* **55**: 63–77.

Sharma M, Sasvari Z, Nagy PD. 2010. Inhibition of sterol biosynthesis reduces tombusvirus replication in yeast and plants. *Journal of Virology* **84**: 2270–2281.

Shindo C, Bernasconi G, Hardtke CS. 2007. Natural genetic variation in *Arabidopsis*: tools, traits and prospects for evolutionary ecology. *Annals of Botany* **99**: 1043–1054.

Sivasubramanian R, Mukhi N, Kaur J (Eds.). 2015. *Arabidopsis thaliana* : a model for plant research. New Delhi: Springer India.

Slootweg EJ, Spiridon LN, Roosien J, Butterbach P, Pomp R, Westerhof L, Wilbers R, Bakker E, Bakker J, Petrescu A-J, et al. 2013. Structural determinants at the interface of the ARC2 and leucine-rich repeat domains control the activation of the plant immune receptors Rx1 and Gpa2. *Plant Physiology* **162**: 1510–1528.

Smith CA, Want EJ, O’Maille G, Abagyan R, Siuzdak G. 2006. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry* **78**: 779–787.

Sofer L, Cabanillas DG, Gayral M, Téplier R, Pouzoulet J, Ducouso M, Dufin L, Bréhélin C, Ziegler-Graff V, Brault V, et al. 2017. Identification of host factors potentially involved in RTM-mediated resistance during potyvirus long distance movement. *Archives of Virology* **162**: 1855–1865.

Soltis NE, Kliebenstein DJ. 2015. Natural variation of plant metabolism: genetic mechanisms, interpretive caveats, evolutionary and mechanistic insights. *Plant Physiology* **169**: 1456–1468.

Soosaar JLM, Burch-Smith TM, Dinesh-Kumar SP. 2005. Mechanisms of plant resistance to viruses. *Nature Reviews Microbiology* **3**: 789–798.

Stare T, Ramšak Ž, Blejec A, Stare K, Turnšek N, Weckwerth W, Wienkoop S, Vodnik D, Gruden K. 2015a. Bimodal dynamics of primary metabolism-related responses in tolerant potato-*Potato virus Y* interaction. *BMC Genomics* **16**: 716–732.

Steinfath M, Strehmel N, Peters R, Schauer N, Groth D, Hummel J, Steup M, Selbig J, Kopka J, Geigenberger P, et al. 2010. Discovering plant metabolic biomarkers for phenotype

prediction using an untargeted approach: Discovering plant metabolic biomarkers. *Plant Biotechnology Journal* **8**: 900–911.

Sulpice R, Pyl E-T, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques MC, et al. 2009. Starch as a major integrator in the regulation of plant growth. *Proceedings of the National Academy of Sciences* **106**: 10348–10353.

Suzuki N, Rivero RM, Shulaev V, Blumwald E, Mittler R. 2014. Abiotic and biotic stress combinations. *New Phytologist* **203**: 32–43.

Thiémélé D, Boissard A, Ndjioudjop M-N, Chéron S, Séré Y, Aké S, Ghesquière A, Albar L. 2010. Identification of a second major resistance gene to *Rice yellow mottle virus*, RYMV2, in the African cultivated rice species, *O. glaberrima*. *Theoretical and Applied Genetics* **121**: 169–179.

Thordal-Christensen H. 2003. Fresh insights into processes of nonhost resistance. *Current Opinion in Plant Biology* **6**: 351–357.

Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Young K, Taylor NE, et al. 2003. Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome research* **13**: 524–530.

Trębicki P, Nancarrow N, Cole E, Bosque-Pérez NA, Constable FE, Freeman AJ, Rodoni B, Yen AL, Luck JE, Fitzgerald GJ. 2015. Virus disease in wheat predicted to increase with a changing climate. *Global Change Biology* **21**: 3511–3519.

Tyler AL, Asselbergs FW, Williams SM, Moore JH. 2009. Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. *BioEssays* **31**: 220–227.

Urcuqui-Inchima S, Haenni A-L, Bernardi F. 2001. Potyvirus proteins: a wealth of functions. *Virus research* **74**: 157–175.

Verma V, Ravindran P, Kumar PP. 2016. Plant hormone-mediated regulation of stress responses. *BMC Plant Biology* **16**: 86–95.

Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics* **12**: 204–213.

Walsh JA, Jenner CE. 2002. *Turnip mosaic virus* and the quest for durable resistance. *Molecular Plant Pathology* **3**: 289–300.

Wang A. 2015. Dissecting the molecular network of virus–plant interactions: the complex roles of host factors. *Annual Review of Phytopathology* **53**: 45–66.

Werner K, Friedt W, 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 SA, Yamamoto ML, Carrington JC. 1999. Selectable viruses and altered susceptibility mutants in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **96**: 772–777.

Win J, Chaparro-Garcia A, Belhaj K, Saunders DGO, Yoshida K, Dong S, Schornack S, Zipfel C, Robatzek S, Hogenhout SA, et al. 2012. Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harbor Symposia on Quantitative Biology* **77**: 235–247.

Wittmann S, Chatel H, Fortin MG, Laliberté J-F. 1997. Interaction of the viral protein genome linked of *Turnip mosaic potyvirus* with the translational eukaryotic initiation factor (iso) 4E of *Arabidopsis thaliana* using the yeast two-hybrid system. *Virology* **234**: 84–92.

Worley B, Powers R. 2016. PCA as a practical indicator of OPLS-DA model reliability. *Current Metabolomics* **4**: 97–103.

Wu S, Alseekh S, Cuadros-Inostroza Á, Fusari CM, Mutwil M, Kooke R, Keurentjes JB, Fernie AR, Willmitzer L, Brotman Y. 2016. Combined use of genome-wide association data and correlation networks unravels key regulators of primary metabolism in *Arabidopsis thaliana*. *PLOS Genetics* **12**: e1006363.

Wylie S, Adams M, Chalam C, Kreuze J, Lopez-Moya JJ, Ohshima K, Praveen S, Rabenstein F, Stenger D, Wang A, et al. 2017. ICTV virus taxonomy profile: *Potyviridae*. *Journal of General Virology* **98**: 352–354.

Xin M, Wang Y, Yao Y, Xie C, Peng H, Ni Z, Sun Q. 2010. Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (*Triticum aestivum* L.). *BMC plant biology* **10**: 123.

Xu K, Nagy PD. 2015. RNA virus replication depends on enrichment of phosphatidylethanolamine at replication sites in subcellular membranes. *Proceedings of the National Academy of Sciences* **112**: 1782–1791.

Xu K, Nagy PD. 2016. Enrichment of phosphatidylethanolamine in viral replication compartments via co-opting the endosomal Rab5 small GTPase by a positive-strand RNA virus. *PLOS Biology* **14**: e2000128.

Xu P, Chen F, Mannas JP, Feldman T, Sumner LW, Roossinck MJ. 2008. Virus infection improves drought tolerance. *New Phytologist* **180**: 911–921.

Xu Y, Li P, Yang Z, Xu C. 2017. Genetic mapping of quantitative trait loci in crops. *The Crop Journal* **5**: 175–184.

Yamaji Y, Maejima K, Komatsu K, Shiraishi T, Okano Y, Himeno M, Sugawara K, Neriya Y, Minato N, Miura C, et al. 2012. Lectin-mediated resistance impairs plant virus infection at the cellular level. *The Plant Cell Online* **24**: 778–793.

- Yamanaka T, Imai T, Satoh R, Kawashima A, Takahashi M, Tomita K, Kubota K, Meshi T, Naito S, Ishikawa M. 2002.** Complete inhibition of tobamovirus multiplication by simultaneous mutations in two homologous host genes. *Journal of Virology* **76**: 2491–2497.
- Yang P, Lüpken T, Habekuss A, Hensel G, Steuernagel B, Kilian B, Ariyadasa R, Himmelbach A, Kumlehn J, Scholz U, et al. 2014.** Protein disulfide isomerase like 5-1 is a susceptibility factor to plant viruses. *Proceedings of the National Academy of Sciences* **111**: 2104–2109.
- Yvon M, Vile D, Brault V, Blanc S, van Munster M. 2017.** Drought reduces transmission of *Turnip yellows virus*, an insect-vectored circulative virus. *Virus Research* **241**: 131-136.
- Zabotina T. 2013.** *Metabolite-based biomarkers for plant genetics and breeding. Diagnostics in Plant Breeding*: 281-309
- Zhang N, Gibon Y, Wallace JG, Lepak N, Li P, Dedow L, Chen C, So Y-S, Kremling K, Bradbury PJ, et al. 2015.** Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *Plant Physiology* **168**: 575–583.
- Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, et al. 2011.** Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications* **2**: 467.
- Zhou X, Stephens M. 2012.** Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**: 821–824.
- Zvereva A, Pooggin M. 2012.** Silencing and innate immunity in plant defense against viral and non-viral pathogens. *Viruses* **4**: 2578–2597.

Réponse d'*Arabidopsis thaliana* au *Turnip mosaic virus* (TuMV) en conditions extérieures et en conditions contrôlées : phénotypage fin de traits de maladie et métaboliques et architecture génétique associée.

Résumé

Les plantes sont des organismes immobiles qui doivent répondre et s'adapter à des contraintes abiotiques et biotiques. Parmi les stress biotiques, les maladies virales, établies ou émergentes, peuvent être responsables de pertes de rendement majeures aux conséquences économiques importantes. Face aux phytovirus la lutte génétique constitue le moyen de lutte le plus efficace, le plus respectueux de l'environnement et du consommateur. Comprendre l'interaction entre les plantes et les virus reste indispensable pour rechercher de nouvelles sources de résistances. Ce travail de thèse s'intéresse à l'étude du pathosystème naturel *Arabidopsis thaliana*/*Turnip mosaic virus* (TuMV). Les essais ont été menés majoritairement en conditions extérieures permettant une analyse de l'interaction dans un environnement multistress. La réponse d'*A. thaliana* a été explorée par l'étude de traits liés à la maladie et par la variation en métabolites primaires et secondaires. Ce travail a permis **i)** de caractériser de façon fine la réponse d'*A. thaliana* au TuMV en conditions multistress en exploitant la diversité naturelle d'une population mondiale et française **ii)** de déterminer l'architecture génétique de cette interaction par des approches de génétique d'association et de QTL mapping. Plusieurs nouveaux loci potentiellement impliqués dans la réponse ont été identifiés **iii)** de montrer l'intérêt du phénotypage métabolique pour discriminer les accessions en fonction de leur sensibilité au TuMV. La multidisciplinarité des approches constitue la richesse de ce travail de thèse qui contribue à une meilleure caractérisation et compréhension de la réponse des plantes lors d'une infection virale.

Mots clés : *Arabidopsis thaliana* – Diversité naturelle - Potyvirus – *Turnip mosaic virus* (TuMV) – Multistress – Métabolisme primaire – Métabolisme secondaire – Génétique d'association – Quantitative Trait Locus (QTL)

***Arabidopsis thaliana* – *Turnip mosaic virus* (TuMV) interaction in common garden and controlled condition experiments: disease and metabolic traits phenotyping and genetic architecture**

Abstract

Plants are immobile organisms which have to adapt to abiotic and biotic constraints. Among biotic stress, established or emerging viral diseases, may be responsible for major yield losses with significant consequences. Genetic control is the most effective, environmentally and consumer-friendly way to control viral infections. Understanding plant/virus interactions remains essential to search for new sources of resistance. This work, focuses on the study of the natural pathosystem *Arabidopsis thaliana*/*Turnip mosaic virus* (TuMV). Most of the trials were conducted in common garden conditions allowing the analysis of the interaction in a multistress environment. *A. thaliana*'s response was explored through the study of disease-related traits and the variations in primary and secondary metabolites. This work allows **i)** the fine characterization of *A. thaliana*'s response to TuMV in multistress conditions through the exploration of the natural diversity of a world and French population **ii)** to determine the genetic architecture of this interaction by genome wide associations and QTL mapping. Several new loci potentially involved in the response have been identified **iii)** to highlight the interest of metabolic phenotyping to discriminate accessions according to their susceptibility to TuMV. The multidisciplinary approaches contribute to a better characterization and understanding of plant-virus interaction.

Keywords : *Arabidopsis thaliana* – Natural Diversity – Potyvirus – *Turnip mosaic virus* (TuMV) – Multistress – Common garden experiments – Primary metabolism – Secondary metabolism – Genome wide association – Quantitative Trait Locus (QTL)
