



**HAL**  
open science

# Voies d'adaptation d'Aphis gossypii, puceron du melon, aux Cucurbitacées ainsi qu'au gène NLR Vat du melon

Pierre Sadon

► **To cite this version:**

Pierre Sadon. Voies d'adaptation d'Aphis gossypii, puceron du melon, aux Cucurbitacées ainsi qu'au gène NLR Vat du melon. Sciences du Vivant [q-bio]. INRAE Avignon, 2023. Français. NNT : . tel-04066792v1

**HAL Id: tel-04066792**

**<https://hal.inrae.fr/tel-04066792v1>**

Submitted on 12 Apr 2023 (v1), last revised 24 May 2023 (v2)

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

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

# THÈSE DE DOCTORAT D'AVIGNON UNIVERSITÉ

**École Doctorale ED536  
AGROSCIENCES ET SCIENCES**

**Mention de doctorat :  
Sciences, Technologies, Santé**

**Laboratoire INRAE – GAFL - AVIGNON**

Présentée par  
**Pierre Olivier Sadon**

---

**Voies d'adaptation d'*Aphis gossypii*,  
puceron du melon, aux Cucurbitacées  
ainsi qu'au gène NLR *Vat* du melon**

---

Soutenue publiquement le 31/03/2023 devant le jury composé de :

M. Bruno Favery, DR, INRAE PACA

M. Hubert Charles, PR, INSA de Lyon

Mme Karin Posthuma, Responsable Phytopathologie Enza Zaden

M. Yvan Rahbé, DR, INRAE Lyon

Mme Desbiez Cecile, CR/HDR, INRAE UR Pathologie Végétale

Mme Nathalie Boissot, CR/HDR, INRAE UR GAFL

Rapporteur

Rapporteur

Examinatrice

Examineur

Examinatrice

Directrice de thèse

# Résumé

Bien que constituant un groupe relativement petit parmi les insectes, les pucerons constituent un sérieux problème en agriculture et la résistance des plantes à ces ravageurs reste peu étudiée. De manière générale, les pucerons doivent manipuler les défenses de la plante pour établir une relation compatible avec le système vasculaire de leur hôte, le phloème, et ainsi s'alimenter durablement. Cette stratégie des pucerons correspond au paradigme des interactions plante-pathogène. Ce paradigme repose sur la reconnaissance, par des récepteurs de la plante codés par des gènes de type NLR (Nucleotide-binding domain Leucine-rich Repeat containing), de molécules produites par les agents pathogènes (effecteurs). *Aphis gossypii* est une espèce de pucerons extrêmement polyphage avec des populations structurées en races hôtes spécialisées sur différentes cultures, dont une spécialisée sur les cucurbitacées. Chez le melon, le gène NLR *Vat* est unique puisqu'il confère à la fois une résistance à *A. gossypii* et aux virus qu'il transmet.

De façon surprenante, l'étude de la diversité des interactions *Vat/A. gossypii* a montré que, d'une manière générale, l'adaptation des pucerons à une résistance de type NLR n'était pas expliquée par un polymorphisme dans un effecteur salivaire. Ce point soulève la question d'une diversité potentielle des voies d'adaptation d'*A. gossypii* à la résistance NLR *Vat*. De manière originale, *A. gossypii* pourrait exploiter l'architecture unique du système vasculaire du melon, partitionné en deux phloèmes aux propriétés divergentes pour contourner la résistance *Vat*.

Dans un premier temps, nous avons cherché à déterminer si les clones d'*A. gossypii* adaptés sur les cucurbitacées et ceux adaptés à *Vat* utilisent le phloème particulier des cucurbitacées comme source d'alimentation alternative lorsqu'ils sont confrontés à des melons sensibles ou résistants. Nous avons montré, au travers d'une approche basée sur la détection de métabolites représentatifs de chacun des deux phloèmes des cucurbitacées, que les clones d'*A. gossypii* appartenant à la race hôte spécialisée sur les cucurbitacées exploitent le phloème classique lorsqu'ils s'alimentent sur melon non-*Vat* et que le contournement de la résistance *Vat* par certains clones n'impliquait pas l'usage du phloème particulier des cucurbitacées.

Dans un second temps, nous avons tenté d'identifier, dans le cadre de référence des interactions plante-pathogène, l'effecteur salivaire d'*A. gossypii* qui interagit directement ou indirectement avec la protéine VAT du melon. Pour cela, nous avons effectué des analyses comparatives de données d'expression de clones déclenchant ou non la résistance *Vat* afin d'identifier des candidats au rôle d'effecteur déclenchant cette résistance. Notamment, un candidat se démarque mais sa validation *in planta* reste à faire.

Ce travail de recherche traite d'une voie d'adaptation non prise en compte dans le paradigme de référence pour les interactions plante-pathogène et offre en particulier la possibilité d'étudier le phénotype *Vat* comme clé de lecture des mécanismes de la résistance héritable.

Mots clés : *Aphis gossypii*, *Vat*, adaptation, effecteurs, avirulence, phloème

# Abstract

Although a relatively small group among insects, aphids are a serious problem in agriculture and plant resistance to these pests remains understudied. In general, aphids must manipulate the plant's defenses to establish a compatible relationship with the vascular system of their host, the phloem, and thus feed sustainably. This strategy of aphids corresponds to the paradigm of plant-pathogen interactions. This paradigm is based on the recognition, by plant receptors encoded by NLR (Nucleotide-binding domain Leucine-rich Repeat containing) type genes, of molecules produced by pathogenic agents (effectors). *Aphis gossypii* is an extremely polyphagous aphid species with populations structured into host races specialized in different crops, including one specialized in cucurbits. In melon, the NLR *Vat* gene is unique since it confers both resistance to *A. gossypii* and to the viruses it transmits.

Surprisingly, the study of the diversity of *Vat/A. gossypii* showed that, in general, the adaptation of aphids to an NLR-type resistance was not explained by a polymorphism in a salivary effector. This point raises the question of a potential diversity of adaptation pathways of *A. gossypii* to the resistance NLR *Vat*. In an original way, *A. gossypii* could exploit the unique architecture of the melon vascular system, partitioned into two phloems with divergent properties to circumvent *Vat* resistance.

First, we sought to determine whether the clones of *A. gossypii* adapted on cucurbits and those adapted to *Vat* use the particular phloem of cucurbits as an alternate food source when faced with susceptible or resistant melons. We have shown, through an approach based on the detection of metabolites representative of each of the two phloems of cucurbits, that clones of *A. gossypii* belonging to the host race specialized in cucurbits exploit the classic phloem when they feed on non-*Vat* melon and that the circumvention of *Vat* resistance by certain clones did not imply the use of the particular phloem of cucurbits.

Secondly, we tried to identify, in the frame of reference of plant-pathogen interactions, the salivary effector of *A. gossypii* which interacts directly or indirectly with the melon VAT protein. For this, we performed comparative analyzes of expression data of clones triggering or not triggering *Vat* resistance in order to identify candidates for the role of effector triggering this resistance. Notably, one candidate stands out but its validation *in planta* remains to be done.

This research work deals with an adaptation pathway not considered in the reference paradigm for plant-pathogen interactions and in particular offers the possibility of studying the *Vat* phenotype as a key to reading the mechanisms of heritable resistance.

Keywords : *Aphis gossypii*, *Vat*, adaptation, effectors, avirulence, phloem

# Remerciements

Je tiens tout d'abord à remercier **Bruno Favery**, **Hubert Charles**, **Karin Posthuma**, **Cécile Desbiez** et **Yvan Rahbé** (que je remercie également pour m'avoir fait découvrir la stylectomie en live !) pour avoir accepté de faire partie de mon jury de thèse et pour le temps consacré à évaluer ce travail.

Un grand merci à toi, **Nathalie**, alias « la cheffe », pour la qualité de ton encadrement, ta patience, ta pédagogie et la disponibilité dont tu as fait preuve à mon égard. J'ai tant appris à tes côtés, sur le plan scientifique et sur moi-même. Tu as réussi à m'apprendre à concilier créativité et pragmatisme, ce qui n'était pas une mince affaire. Je n'oublierai jamais tes imitations de pucerons s'acharnant à piquer des feuilles de melon *Vat* et ton mantra « Kimbe Red ». Ce qui va me manquer particulièrement, c'est cette façon que tu avais de faire irruption dans mon bureau à l'annonce d'une autre de tes phrases fétiches : « Pierre, j'ai eu une idée ... ». D'ailleurs, c'est comme cela que toute cette aventure a commencé.

A toi **Pascale**, un grand merci pour ton soutien technique sans faille et tout particulièrement pour ton implication décisive dans la préparation et la mise en œuvre de mes manip pucerons et bien entendu, pour ton café corse et corsé agrémenté de tes délicieuses pâtisseries ! Je garderai le très bon souvenir de ton accueil, tant au GAFL que chez toi à de multiples occasions y compris la visite que tu avais organisée pour moi à Saint-Rémy et aux alentours. Un grand merci à toi, **Jacques**, pour m'avoir fait découvrir et m'avoir accompagné dans cet univers fascinant qu'est la bioinfo, mais tout autant pour m'avoir laissé pirater ta super tasse à café (je l'ai probablement remplie un bon millier de fois) et aussi pour m'avoir approvisionné en loukoums lorsque j'avais besoin de sucreries. **Vincent**, je tiens à te témoigner ma profonde gratitude pour avoir partagé avec moi ton expérience dans le domaine fascinant qu'est la biologie moléculaire et pour tout le temps que tu as accordé à mon projet. Un grand merci pour toutes les manip que tu as réalisées et qui ont été essentielles à la réussite de ma thèse. Ta zénitude et toutes nos pauses café ont beaucoup compté pour me permettre de relâcher la pression lorsque j'en avais besoin. Je tiens également à remercier mes trois autres anges gardiens de la bioinfo, **Frédérique**, **Sonia** et **Emmanuel** pour leur appui technique et les connaissances qu'ils m'ont apportés ! A vous, **Cécile** et **Sylvie**, pour m'avoir dopé avec tous vos encouragements et votre bienveillance lorsque j'en ai eu besoin.

J'associe à ces remerciements **Raphaël Lugan** et **Olivier Chevallier** de la plateforme de métabolomique de l'Université d'Avignon sans qui ce travail n'aurait pu être mené à bien.

A toi, **Rafa** (a.k.a Tio), mon premier colloc' de bureau, avec qui je partage une amitié indéfectible qui est née instantanément entre nous, *muchas gracias amigo!* Je n'oublierai jamais les 1001 moments collecteurs que nous avons partagés : nos soirées *Explo*, ciné, restau, pauses café, parties de babyfoot, et travail en bioinfo en binôme de choc ... pour n'en citer que quelques-uns ! A toi, **Séverine**, ma deuxième colloc' de bureau, pour ton enthousiasme, ta passion communicative et ta profonde gentillesse au quotidien. Dans ce long marathon qu'est une thèse, nous avons été présents l'un pour l'autre, aussi bien dans les moments difficiles que fun : je t'en remercie sincèrement ! A toi, **Aimeric**, pour ce lien de solidarité entre thésards que nous avons tout de suite tissé ! Ça a été un réel plaisir de te côtoyer, tant dans les moments studieux que dans ceux de loisir. Et surtout, je n'oublierai pas tes rengaines collecteur : « café ? », « quand est-ce qu'on mange ? », « Beer 'o' clock tonight ? » pour ton humour de degré n et avoir fait de moi un killer du babyfoot !

Merci à tous les membres de l'équipe Melon pour ces années passées dans la bonne humeur : **Nat**, **Séverine**, **Pascale**, **Véro**, **Karine**, **Vivi** et **Vincent**. Sachez que tous ces moments de rigolade, en pause-déjeuner ou à l'heure de l'apéro, vont me manquer ! Merci également aux collègues du CPER, du CRB, d'Arbo et de Patho : **Marie-Noelle** (mille mercis à toi pour ta contribution essentielle à la partie métabolomique !), **Rémy**, **Mathilde**, **Renaud**, **Laure**, **Rebecca**, **Frédérique**, **Esther**, **Maxi**, **Nasradine**, **Lorraine**, **Najma** et **Louis** pour les moments conviviaux autour d'un gâteau à la pause-café, lors du déjeuner ou à l'extérieur.

Je tiens également à remercier l'ensemble de l'équipe administrative **Evelyne**, **Astrid**, **Carine**, **Sébastien** et **Philippe** pour m'avoir aidé dans la gestion des aspects administratifs. Merci à toi, **Lalia**, pour toutes nos discussions sur tant de sujets qui rendent la vie plus belle et aussi toutes tes délicates attentions !

Ma thèse s'inscrit dans le cadre d'un contrat CIFRE, en partenariat avec les entreprises semencières ENZA ZADEN et RIJK ZWANN qui ont contribué financièrement à la création de ce projet et qui m'ont donné l'opportunité de réaliser cette thèse. Je les en remercie chaleureusement. Je témoigne toute ma reconnaissance à **Karin**, **Ingrid**, **Dominique** et **Laurent** et pour leur soutien, leurs encouragements et l'intérêt qu'ils ont porté à ce projet. Je remercie aussi tous les membres

de la station d'ENZA d'Almería et tout particulièrement **Laura, Jake, Jorge, Frédéric** et **Pédro** pour votre accueil chaleureux (et en pleine période de canicule) dans votre belle ville et ce petit paradis qu'est Aguadulce (ou Sweeet Woooter comme le disait si bien Pédro). Je garderai un beau souvenir de ma visite de votre impressionnante station expérimentale et de ces paysages si singuliers tout droit sortis des films de Sergio Leone (ou alors serait-ce l'inverse ?).

Spéciale dédicace à toute la team des thésards, du GAFL, de Patho et de Saint-Paul ! **Isidore** et **Mariem.O** (pour m'avoir accueilli dans la team dès mon arrivée et vos accompagnements à la guitare), **Estelle** (et nos concours d'autodérision et ta bienveillance toute particulière), **Delyan** et **Mamoudou** (avec votre zénitude à toute épreuve), **Stéphanie** et **Zoé** (le duo de choc tant dans le fun que dans la Science avec un grand S), **Hussein** (monsieur fallafel, le meilleur cuisinier que j'ai rencontré... pardon maman !), mais également **Mariem.N, Alexandre, Gaëtan, Jiantao, Anne-Sophie, Thomas, Enrico, Dario, Thibaud** et **William**. A toutes nos soirées passées à l'Explo.

Un immense merci à mes amis de Tournefeuille (forza 31 !), **Thomas, Aurélie, Thomas, Christelle, Lucile** et **Agnès** ! Même si quelques 341,9 km (via A9 et A61 n'est-ce pas ?) nous séparent, vous avez toujours été là. A notre longue et belle amitié ... et à nos années bahut !

A toi, **Alban Jacques**, grand gourou mi-chercheur, mi-prof et mi-metalhead (si...si, trois-demi c'est possible !) pour m'avoir accueilli dans ton équipe le temps d'un stage qui sur le papier semblait anodin mais qui m'a fait découvrir et m'a propulsé dans l'univers de la recherche. Merci de m'avoir accordé ta confiance, c'est avec toi que tout a commencé. Pour te citer : « coooooool ! ». Sincèrement, j'espère un jour devenir un aussi bon prof que toi.

Je terminerai par ma famille, ma fierté et mon soutien infaillible, j'ai nommé **la tribu Sadon** ! **Maman, Papa**, mon frère **Rémi, Papy** et **Mamie, Thierry, Anaïs, Laura, Franck, Christelle** et **Louis**. Dans ce marathon et toutes ces épreuves, ce sont des ailes que vous m'avez données. Cette thèse est pour vous !





# Liste des figures et des tables (hors articles)

<b>Figure 1</b> - Représentation schématique du complexe cellules compagnes – élément criblé, unité fonctionnelle du phloème .....	20
<b>Figure 2</b> - Schéma de la dynamique du transport phloémien .....	23
<b>Figure 3</b> - Schéma d'une coupe transversale d'une tige de citrouille .....	26
<b>Figure 4</b> – Le cadre conceptuel de référence pour la description du système immunitaire des plantes .....	34
<b>Figure 5</b> – Confrontation de la stratégie d'invasion des pucerons avec le système immunitaire des plantes à l'échelle moléculaires dans le cas d'une interaction compatible ou incompatible .....	36
<b>Figure 6</b> – Huit effecteurs notables caractérisés chez 4 espèces de pucerons .....	41
<b>Figure 7</b> – Modèle des interactions entre une lignée de melon transgénique Vat-1 et différents biotypes d' <i>A. gossypii</i> montrant des comportements contrastés vis-à-vis de cette résistance .....	51
<b>Figure 8</b> – Caractéristiques des 12 clones d' <i>A. gossypii</i> utilisés pour la recherche de candidats à l'AVR .....	92
<b>Figure 9</b> – Trois hypothèses pour la transition AVR-VIR .....	94
<b>Figure 10</b> - Principales caractéristiques du génome de C9 .....	99
<b>Figure 11</b> – Table de vérité logique déterminant le profil des SNP bialléliques pour les candidats d'AVR .....	104
<b>Figure 12</b> – Construction de la base de données de protéines salivaires d'hémiptères .....	109
<b>Figure 13</b> - Principales caractéristiques des transcriptomes de têtes de pucerons assemblés à partir de données RNA-seq .....	110
<b>Figure 14</b> - Analyse BUSCO des transcriptomes de têtes de pucerons assemblés à partir de données RNA-seq .....	111
<b>Figure 15</b> - Principales caractéristiques des transcriptomes de têtes de pucerons assemblés à partir de données Isoseq .....	113
<b>Figure 16</b> - Analyse BUSCO des transcriptomes de têtes de pucerons assemblés à partir de données Isoseq ..	114
<b>Figure 17</b> - Récapitulatif des stratégies de production de listes de candidats d'AVR .....	116

<b>Figure 18</b> – Recherche de candidats ‘de non expression’ partir de données RNAseq de têtes.....	118
<b>Figure 19</b> - Recherche de candidats ‘de polymorphisme’ partir de données RNAseq de têtes.....	120
<b>Figure 20</b> - Recherche de candidats ‘de non expression’ et ‘de polymorphisme’ avec les données Isoseq de têtes de pucerons.....	122
<b>Figure 21</b> – Recherche de candidats ‘de polymorphisme’ en combinant des données Isoseq et RNAseq de têtes de pucerons.....	124
<b>Figure 22</b> - Confirmation du candidat ‘55596’ .....	125
<b>Figure 23</b> - Recherche de candidats ‘de non expression’ et ‘de polymorphisme’ en combinant des données RNAseq de têtes de pucerons et le génome de référence annoté pour le clone C9. ....	127
<b>Figure 24</b> – Profil attendu d'un gène candidat à l'AVR de type ‘épissage alternatif’ .....	129
<b>Figure 25</b> - Recherche de candidats avec des évènements d'épissage alternatifs en combinant des données Isoseq et RNAseq de têtes de pucerons et le génome de référence annoté pour le clone C9. ....	130
<b>Figure 26</b> – Bilan des candidats d'AVR identifiés par approches in silico .....	134
<b>Figure 27</b> – Nombre de candidats d'AVR identifiés par différents usages de données in silico .....	135
<b>Table 1</b> - Résultat du séquençage Hiseq2000 des ARN de têtes de pucerons .....	97
<b>Table 2</b> – Résumé du traitement des données de séquençage SMRT des ARN de têtes de pucerons.....	97
<b>Table 3</b> – Bilan des candidats d'AVR restant à confirmer .....	139

# Liste des abréviations

**AA** : Acide Aminé

**ACE** : Angiotensin-converting enzyme

**AChE** : Acétylcholinestérase

**ADN** : Acide désoxyribonucléique

**Armet**: Arginine rich, mutated in early stage of tumors

**ARN** : Acide ribonucléique

**AUX** : Auxine

**AVR** : Avirulent/Avirulence

**BAM**: Binary Alignment Map

**BED**: Browser Extensible Data

**BLAST**: Basic Local Alignment Search Tool

**BUSCO**: Benchmarking Universal Single-Copy Orthologs

**CABYV** : Virus de la jaunisse des Cucurbitacées transmise par puceron

**CC-SE** : Complexe cellules compagnes – élément criblé

**CCS** : Séquences consensus circulaires

**CDS** : Séquence Codante

**CMV** : Virus de la mosaïque du concombre

**DAMP** : Damage-Associated Molecular Patterns

**EA** : Epissage Alternatif

**EEA** : Evènements d'Epissage Alternatif

**EFP** : Phloème Extrafasciculaire

**EPG** : Electrical penetration graph (Electropénétrographie)

**ES** : Exons Skipping

**EST** : Marqueurs de Séquence Exprimées

**ETI**: Effector-triggered immunity

**ETS**: Effector-triggered susceptibility

**FASTA** : Format de fichier texte utilisé pour stocker des séquences nucléique ou protéique.

**FLNC** : Sub-reads complets non chimériques

**FP** : Phloème Fasciculaire

**GATK** : Genome Analysis Toolkit

**GC-MS** : Chromatographie gazeuse couplée à une analyse par spectrométrie de masse

**GST**: Glutathion S-transférases

**GFF:** General feature format

**HAMP :** Herbivory-associated molecular patterns

**HR :** Réponse hypersensible

**INDEL :** INsertion/DELétion

**IR :** Intron Retention

**JA :** Acide jasmonique

**LC-MS :** Chromatographie liquide couplée à une analyse par spectrométrie de masse

**MIF1:** Macrophage Migration Inhibitory Factor 1

**NLR:** Nucleotide-binding domain Leucine-rich Repeat containing

**NCBI:** National Center for Biotechnology Information

**ONT:** Oxford Nanopore Technology

**P450:** Cytochrome p450 monooxygénase

**Pb:** Paire de bases

**PCR:** Polymerase chain reaction

**PE:** Paired-End

**PintO1:** Progeny increase to Overexpression 1

**PRR:** Pattern recognition receptor

**PSM:** Plant Secondary Metabolites

**PTI:** Pattern-triggered immunity

**QTL :** Locus de caractère quantitatif

**RNAi:** ARN interference

**ROS:** Reactive oxygen species

**RT:** Retention Time

**SA:** Acide salicylique

**SNP:** Singular Nucleotide Polymorphism

**TC:** Tubes criblés

**UTR:** Untranslated Regions

**VAT:** Virus Aphid Transmission

**VCF:** Variant Calling Format

**VIR :** Virulent/Virulence

**VPS52 :** Vacuolar protein sorting-associated protein 52

**WMV :** Virus de la mosaïque de la pastèque

**ZYMV :** Virus de la mosaïque jaune de la courgette

# Table des matières

<b>RESUME</b> .....	<b>1</b>
<b>ABSTRACT</b> .....	<b>2</b>
<b>REMERCIEMENTS</b> .....	<b>3</b>
<b>LISTE DES FIGURES ET DES TABLES (HORS ARTICLES)</b> .....	<b>7</b>
<b>LISTE DES ABREVIATIONS</b> .....	<b>9</b>
<b>INTRODUCTION GENERALE</b> .....	<b>15</b>
<b>1. Préambule : des cultures et des pucerons</b> .....	<b>16</b>
<b>2. Le phloème : un réseau vasculaire polyvalent ciblé par des pucerons</b> .....	<b>19</b>
2.1. Structure et dynamique du réseau phloémien .....	19
2.1.1. Mécanique du transport phloémien .....	21
2.2. Implication dans diverses fonctions vitales des plantes .....	23
2.2.1. Le phloème comme réseau d'alimentation .....	23
2.2.2. Le phloème comme réseau de communication intégratif .....	24
2.3. Le système phloémien particulier des cucurbitacées .....	26
2.4. Qualités nutritives et potentiel délétère de la sève pour les pucerons .....	28
<b>3. Les interactions plante-puceron : de l'échelle tissulaire à l'échelle moléculaire</b> .....	<b>31</b>
3.1. Une stratégie furtive pour atteindre le phloème .....	31
3.1.1. La salive des pucerons : pièce maîtresse d'une effraction furtive .....	31
3.2. Le dialogue moléculaire entre les effecteurs aphidiens et le système immunitaire de la plante .....	33
3.2.1. Un modèle de référence pour les interactions plante-puceron .....	33
3.2.2. Déviations par rapport au modèle de référence : vers une conception plus intégrative .....	38
3.3. Les effecteurs aphidiens : état de l'art .....	40
3.3.1. Des connaissances très parcellaires .....	40
3.3.2. Focus sur les effecteurs les mieux connus .....	42
3.3.3. Quelques caractéristiques notables des effecteurs aphidiens .....	45
3.3.4. Les effecteurs aphidiens face au paradigme de référence .....	46
<b>4. La résistance <i>Vat</i> du melon : une résistance à <i>A. gossypii</i> et aux virus qu'il transmet</b> .....	<b>47</b>
4.1. Physiologie associée à <i>Vat</i> et conséquences pour les pucerons et les virus .....	47
4.2. Un spectre de résistance en désaccord avec le modèle de référence .....	49
4.3. Effets de <i>Vat</i> à l'échelle des systèmes agricoles .....	52
4.3.1. Effets de <i>Vat</i> sur les populations de pucerons et sa durabilité .....	52
4.3.2. Effet de <i>Vat</i> sur les virus transmis par pucerons et sa durabilité .....	53
<b>5. Ma thèse, son contexte et ses objectifs</b> .....	<b>54</b>

## **CHAPITRE 1 (ARTICLE) - RECHERCHE DU PHLOEME SOURCE POUR *A. GOSSYPHII*... 56**

<b>1. Contexte de l'étude</b> .....	<b>57</b>
<b>2. Aphid adaptation to cucurbits: sugars, cucurbitacin and phloem structure in resistant and susceptible melons</b> .....	<b>58</b>
<b>Abstract</b> .....	<b>58</b>
<b>Background</b> .....	<b>60</b>
<b>Results</b> .....	<b>62</b>
<b>Discussion</b> .....	<b>64</b>
<b>Materials and methods</b> .....	<b>67</b>
<b>References</b> .....	<b>72</b>
<b>Figures</b> .....	<b>77</b>

## **CHAPITRE 2 – RECHERCHE DE L'EFFECTEUR APHIDIEN INTERAGISSANT AVEC LA PROTEINE NLR VAT-1 DU MELON..... 89**

<b>1. Contexte de l'étude</b> .....	<b>90</b>
<b>2. Hypothèses de travail</b> .....	<b>93</b>
<b>3. Matériels et Méthodes</b> .....	<b>96</b>
3.1 Les clones de pucerons.....	96
3.2 Construction de la base donnée salivaire.....	96
3.3 Données transcriptomiques/génomiques mobilisées pour ma thèse.....	96
3.4 Production de listes de candidats d'AVR <i>in silico</i> .....	100
3.4.1 Utilisation de données RNAseq.....	100
3.4.2 Utilisation de données Isoseq .....	101
3.4.3 Intégration du génome de référence pour le clone C9 .....	102
3.5 Confirmation des gènes candidats en laboratoire.....	107
<b>4. Résultats</b> .....	<b>108</b>
4.1. Construction d'une base de données de protéines salivaires d'hémiptères.....	108
4.2 Génération de transcriptomes <i>de novo</i> de têtes de pucerons .....	110
4.3 Transcriptomes vs base de données salivaires .....	115
4.4 Production de listes de candidats d'AVR <i>in silico</i> .....	116
4.4.1 Stratégie 1 : utilisation des données RNA-seq .....	117
4.4.2 Stratégie 2 : combinaison de données RNAseq et Isoseq .....	122
4.4.3 Stratégie 3 : combinaison de données RNA-seq, PacBio Isoseq et génome de référence.....	126
<b>5. Discussion</b> .....	<b>133</b>
5.1. Vue d'ensemble des candidats d'AVR identifiés.....	133
5.2 Pertinence des stratégies et cohérence entre elles .....	134
5.3 Intérêt et usage de la base de données salivaire .....	137
5.4 Un candidat pour la race d'hôte des cucurbitacées .....	137
5.5 Identification de candidats additionnels .....	138

<b>DISCUSSION GENERALE ET CONCLUSION .....</b>	<b>143</b>
<b>1. Vue d'ensemble .....</b>	<b>144</b>
<b>2. L'adaptation d'<i>A. gossypii</i> aux cucurbitacées .....</b>	<b>144</b>
<b>3. Le déclenchement de la résistance <i>Vat-1</i> .....</b>	<b>145</b>
3.1 Objectif : validation <i>in planta</i> du gène d'AVR.....	146
3.2 D'autres hypothèses pour expliquer la transition AVR-VIR ? .....	147
3.3 Identification de protéomes salivaire AVR/VIR : vers une nouvelle stratégie de recherche de candidats.....	150
<b>4. Le contournement de la résistance <i>Vat-1</i>.....</b>	<b>151</b>
4.1 L'EFP n'est pas impliqué dans le contournement de <i>Vat-1</i> .....	151
4.1.1 Des effecteurs pour expliquer le contournement de <i>Vat-1</i> ? .....	152
4.1.2 Portrait-robot d'un effecteur de contournement.....	153
<b>5. Conclusion de ma thèse .....</b>	<b>154</b>
.....	156
 <b>ANNEXES .....</b>	 <b>156</b>
<b>1. Article : Tables supplémentaires .....</b>	<b>157</b>
<b>2. Scripts .....</b>	<b>159</b>
 <b>RÉFÉRENCES BIBLIOGRAPHIQUES .....</b>	 <b>166</b>





# **Introduction générale**

# 1. Préambule : des cultures et des pucerons

Les pucerons sont un problème sérieux pour l'agriculture bien qu'ils soient un groupe d'insectes relativement restreint (4000 espèces dans le monde, environ 600 en France) face à 10000 espèces de sauterelles, 12000 espèces de papillons géométriques et 60 000 espèces de charançons (Dedryver et al., 2010). Une centaine d'espèces de pucerons sont capables d'exploiter les systèmes agricoles. Le déséquilibre climatique aura des conséquences sur les performances des pucerons et leurs impacts sur les paysages agricoles mondiaux, ainsi que sur leurs organismes associés (Finlay & Luck, 2011; Hullé et al., 2010; Van Baaren et al., 2010; Van Munster et al., 2017).

Les paysages ruraux offrent une multitude d'espèces végétales, mais un nombre limité d'entre elles ont une valeur écologique satisfaisante pour les pucerons. Pour trouver un hôte convenable dans un court délai, les pucerons sélectionnent leur plante hôte sur la base d'un ensemble de **signaux** variés et appréciés par différents organes spécialisés dans la **réception d'informations visuelles, olfactives, tactiles et gustatives** (Döring, 2014). Ainsi, la recherche d'une plante convenable comme site d'alimentation, mais aussi de vie et de multiplication s'effectue lors d'une **séquence comportementale précise**. Lors d'une première phase de repérage, les pucerons vont s'orienter vers une plante répondant positivement à un ensemble de **stimuli visuels** et de composés volatils. Suite à leur atterrissage sur la plante à première vue intéressante, les pucerons réalisent une **appréciation kinésique** de la surface des feuilles notamment en utilisant leur système antennaire comme récepteurs tactiles. Dans une ultime phase d'appréciation gustative, les pucerons effectuent un **sondage des couches cellulaires de la plante**.

Lorsqu'une plante a été sélectionnée par les pucerons, ceux-ci s'y installent de façon durable et affaiblissent leurs plantes hôtes de différentes façons : **(i)** en tant qu'**insectes phloémophages**, ils détournent à leur profit une partie des nutriments nécessaires aux plantes pour leur croissance, leur développement et leur reproduction; **(ii)** le mécanisme d'alimentation des pucerons s'accompagne d'injections de **composés salivaires** potentiellement phytotoxiques; **(iii)** le **miellat**, condensé de sucres excrété par les pucerons, constitue un milieu favorable au développement d'agents pathogènes et **(iv)**, la **transmission de phytovirus**, à l'origine des dégâts les plus importants sur le plan économique (Dedryver et al., 2010).

Face à la menace représentée par les pucerons pour l'économie agricole, l'usage de pesticides a longtemps été plébiscité. Cependant, l'émergence de **populations résistantes à différentes classes d'insecticides** (e. g. carbamates, organophosphates, néonicotinoïdes) a été détectée à travers le monde et ce, depuis plusieurs décennies (Carletto et al., 2009; Furk et al., 1980). Les mécanismes moléculaires impliqués dans la résistance, ou dans la perte de sensibilité aux insecticides incluent des mutations du gène de structure du système nerveux central AChE (acétylcholinestérase) ciblé par les principes actifs de plusieurs de ces insecticides ou encore une élévation accrue du métabolisme de détoxification conférée par des carboxylestérases (Andrews et al., 2004; C. W. Cao et al., 2008), (Bass et al., 2014; J. S. Ramsey et al., 2010). **L'inefficacité de ses insecticides à long terme** ainsi que la prise de conscience collective de leurs **impacts délétères sur l'environnement** entraînent un abandon progressif de la lutte chimique. Les travaux de recherches s'orientent désormais vers **la compréhension et le transfert opérationnel des ressources génétiques végétales offertes par la diversité des populations naturelles** (Boissot, Schoeny, et al., 2016; Catherine Dogimont et al., 2010).

***Aphis gossypii***, puceron du melon ou du coton, est un ravageur de cultures à l'échelle mondiale. Cette espèce de puceron **extrêmement polyphage** a été identifiée sur plus de 900 espèces de plantes (Thomas, Boissot, et al., 2012). Son spectre d'hôtes remarquable inclut des espèces d'intérêt agronomique et économique appartenant à des groupes botaniques variés incluant les *Cucurbitaceae* (e.g. melon, courgette, pastèque, concombre), les *Malvaceae* (e. g. coton, hibiscus, gombo), les *Solanaceae* (piment, pomme de terre, aubergine) et les *Rutaceae* (e. g. divers agrumes). Au sein de l'espèce *A. gossypii*, les populations sont structurées en plusieurs **rares d'hôtes** spécialisées sur différentes cultures (Carletto et al., 2009). Chaque race d'hôte est représentée par un nombre limité de génotypes qui se reproduisent principalement voire exclusivement par de manière clonale (parthénogénèse). L'une d'entre elles est connue pour sa capacité à contourner spécifiquement la résistance non hôte des Cucurbitacées. Cette **race d'hôte spécifique des cucurbitacées** est formée par **trois groupes génétiques différents** (Fuller et al., 1999; Thomas et al., 2016).

*A. gossypii* transmet **plus de 50 phytovirus** appartenant majoritairement à trois familles différentes (les Potyviridae, les Bromoviridae et les Luteoviridae) (Hervé Lecoq & Desbiez, 2012; Van Emden & Harrington, 2007) et constitue un vecteur efficace pour quatre virus fréquemment observés dans les bassins de production français : le virus de la mosaïque du concombre (**CMV**, *Cucumovirus*, *Bromoviridae*), le Virus de la mosaïque de la pastèque (**WMV**, *Potyvirus*, *Potyviridae*), le Virus de la mosaïque jaune de la courgette (**ZYMV**, *Potyvirus*, *Potyviridae*) et le virus de la jaunisse des Cucurbitacées transmise par puceron (**CABYV**, *Polerovirus*, *Luteoviridae*).

**Le phloème, réseau vasculaire des plantes, est un espace clé dans les interactions plante-puceron. A l'interface entre ces deux protagonistes, le flux salivaire sécrété par les pucerons est susceptible d'interagir avec les défenses inductibles de son hôte et joue un rôle décisif dans l'issue de cette interaction.** Le cas d'étude *A. gossypii*-melon présente des singularités utiles à la production de connaissances scientifiques concernant : **(i)** l'adaptation des clones appartenant à la race d'hôte des cucurbitacées, ainsi qu'à **(ii)** l'adaptation des clones de cette race d'hôte à une résistance génétique chez le melon.

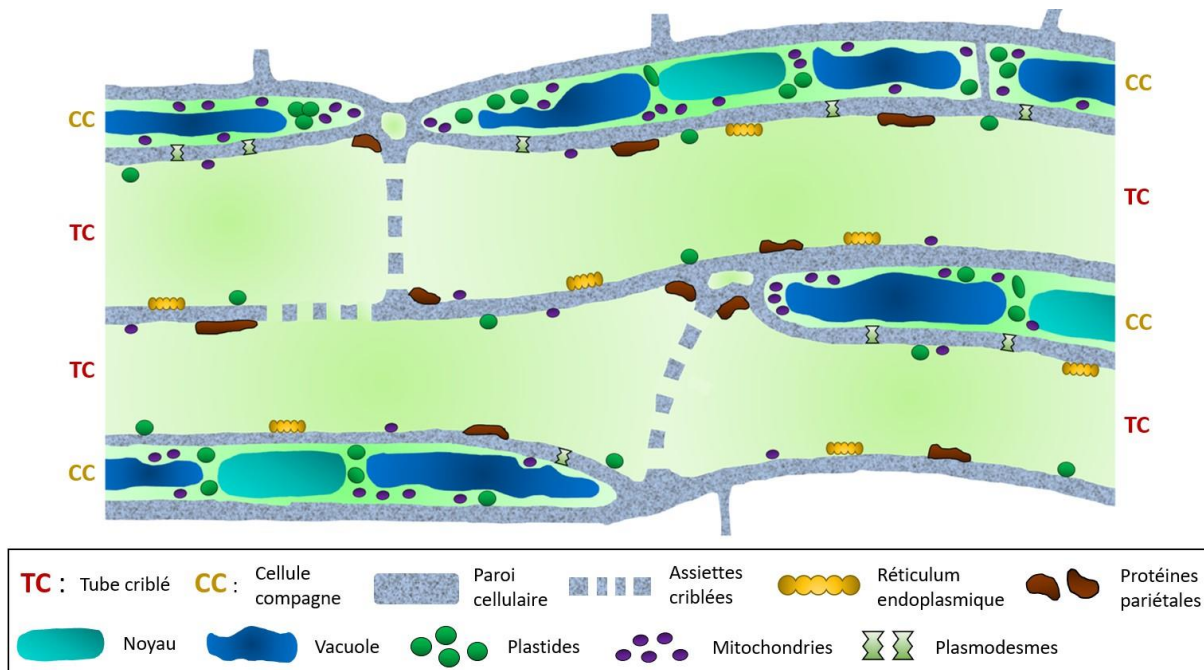
## 2. Le phloème : un réseau vasculaire polyvalent ciblé par des pucerons

### 2.1. Structure et dynamique du réseau phloémien

L'émergence d'un système vasculaire chez les plantes a fortement contribué à l'adaptation et au développement des végétaux terrestres dans des environnements changeants et occasionnant divers stress (Fukuda & Ohashi-Ito, 2019). Le système vasculaire assure le **transport de l'eau et des nutriments** nécessaires à la croissance et au développement des plantes. Au cours de l'évolution, sa sophistication progressive a rendu possible la complexification de la structure des plantes.

Ce réseau systémique se compose de deux types de tissus de transports spécialisés, **le xylème et le phloème**, tous deux issus de la différenciation des cellules méristématiques du procambium (Carlsbecker & Helariutta, 2005), sous le contrôle de programmes de différenciations cellulaires spécifiques. Le xylème et le phloème jouent des **rôles complémentaires** au sein de la plante et assurent le transport de l'eau et des solutés dans toute la plante (Aart J.E. Van Bel, 1999). Ils sont organisés de façon collatérale et interconnectés au sein des tissus de soutien de la plante (Schulz, 1998).

Le phloème est un **tissu composite** formé de différents types de cellules vivant en interaction et dont les fonctions sont complémentaires (Jiang et al., 2019). Les cellules conductrices majeures du phloème, ou **éléments criblés (SE)**, sont issues de la différenciation de cellules méristématiques modifiées par l'exercice d'un **programme de spécialisation cellulaire spécifique** (Blob et al., 2018; Heo et al., 2017; Jiang et al., 2019; Lewis et al., 2022). Leur **architecture hautement spécialisée [figure 1]** donne son efficacité à la **conductivité du flux de sève** circulant dans les différents tissus (Carlsbecker & Helariutta, 2005).



**Figure 1** - Représentation schématique du complexe cellules compagnes – élément criblé, unité fonctionnelle du phloème

Les tubes criblés (TC) et les cellules compagnes (CC) sont connectés par l'intermédiaire de nombreux plasmodesmes. Les plastides, les mitochondries et les réticulums endoplasmiques empilés ont une position pariétale et sont distribués de façon uniforme le long de la membrane plasmique des TC, alors que les protéines pariétales sont agrégées localement. Les protéines pariétales et les réticulums endoplasmiques peuvent être localisés sur ou à proximité des assiettes criblées sans perturber le flux de masse. Adaptation de (A. J.E. Van Bel, 2003).

Une dégradation du contenu cytoplasmique par **autolyse** entraîne un appauvrissement considérable de la machinerie de biosynthèse des cellules du méristème. Ce **processus autophagique** est semblable à de nombreux programmes de morts cellulaires. Il est toutefois incomplet et de nature sélective, causant la dégradation de plusieurs types d'organelles sans affecter les autres (Cronshaw et al., 1981; Oparka & Turgeon, 1999). La désagrégation de la vacuole centrale est suivie par la dissolution du cytosquelette et l'élimination des ribosomes cytoplasmiques, des appareils de Golgi et du noyau cellulaire (Cayla et al., 2015; Knoblauch et al., 2018; Oparka & Turgeon, 1999; G. A. Thompson & van Bel, 2012). L'espace intracellulaire résultant de ce processus, ainsi que la relocalisation pariétale des quelques éléments cytoplasmiques demeurés intacts (Heo et al., 2017), minimise les frottements causés par la circulation du flux de masse.

En complément de ce réaménagement cellulaire, les éléments criblés voient leur structure changer. Leur élongation, ainsi que l'épaississement de leur paroi permettent d'assurer la résistance à la pression exercée par le flux de masse circulant (Blob et al., 2018). La formation de pores spécialisés dans la paroi cellulaire renforcée, **les assiettes criblées**, permet l'interconnexion des éléments criblés. Le déplacement à **longue distance** du flux de masse est assuré par la **continuité symplastique** entre les éléments criblés (Mullendore et al., 2010). Les canaux ainsi formés constituent l'une des unités du réseau de transport phloémien : **les tubes criblés** (Knoblauch & Oparka, 2012).

L'architecture modulaire des tubes criblés est une **innovation majeure** des plantes terrestres qui permet une extension du réseau phloémien adaptée à la croissance des plantes et au développement de nouveaux organes. Cette flexibilité de structure assure la connexion des différents organes tout au long de la vie des plantes. (Schulz, 1998; A. J.E. Van Bel, 2003).

Si le processus de différenciation confère aux éléments criblés les propriétés physiques nécessaires pour conduire le flux de sève, l'appauvrissement associé de leur machinerie de biosynthèse est **incompatible avec une autonomie de fonctionnement** et *a fortiori* de survie. C'est leur association intime avec un autre type de cellules spécialisées, **les cellules compagnes**, qui répond à leurs besoins vitaux (De Rybel et al., 2016; Knoblauch & Oparka, 2012; A. J.E. Van Bel, 2003). Avec un cytoplasme particulièrement riche en organelles, notamment des mitochondries hautement actives, les cellules compagnes assurent le **rôle de pourvoyeurs en énergie** pour les tubes criblés (Cayla et al., 2015; Heo et al., 2017). **Le complexe cellules compagnes – élément criblé (CC-SE)**, issu de la même cellule mère (Aart J.E. Van Bel, 1999), est **la véritable unité fonctionnelle du transport phloémien** (A. J.E. Van Bel, 2003) **[figure 1]**. La connexion interne entre éléments criblés et cellules compagnes du complexe se fait via les plasmodesmes, qui régulent le flux de masse de façon bidirectionnelle.

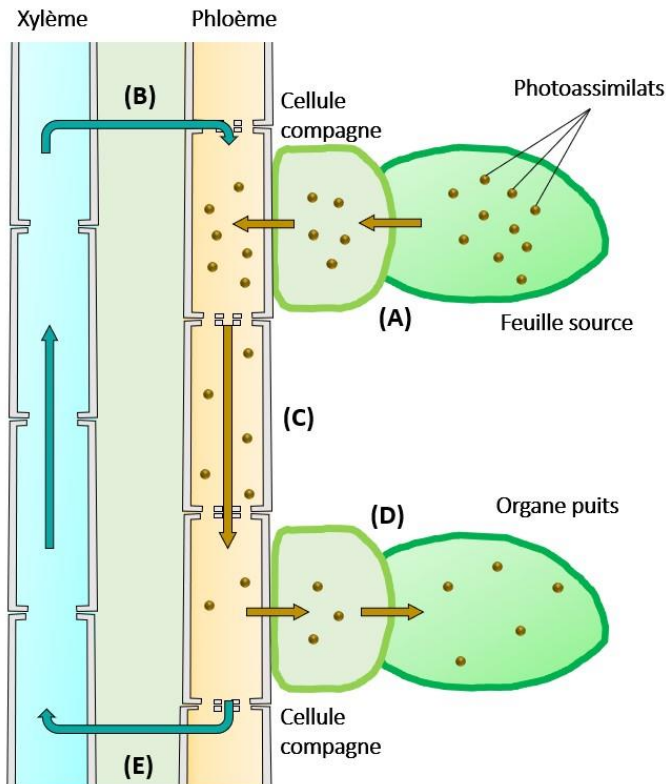
Les CC-SE sont incrustés dans un tissu de soutien formé par des cellules peu différenciées, **le parenchyme**. L'ensemble de ces différents types de cellules sont en contact physique via un espace intercellulaire, communément appelé **apoplaste** (Jiang et al., 2019).

### 2.1.1. Mécanique du transport phloémien

Le phloème des plantes supérieures transporte les produits de la photosynthèse et d'autres nutriments sur de longues distances, **des feuilles sources aux tissus puits**, comme les jeunes feuilles vers les fruits ou encore les racines où ils sont utilisés pour la croissance ou le stockage (Robert Turgeon & Oparka, 2010). L'hypothèse pour le transport longue distance des photoassimilats a été proposée

par Ernst Münch dès 1930. Münch établit que le **flux de masse** circulant dans le phloème résulte du **différentiel de pression osmotique** entre les organes sources et les organes puits, sans qu'aucune force additionnelle ne soit requise (Knoblauch et al., 2016; Mullendore et al., 2010). Ce transport requiert un **couplage hydraulique** avec le flux d'eau circulant dans le xylème (Dinant & Lemoine, 2010) **[figure 2]**. Ces flux d'eau, bidirectionnels entre le xylème et le phloème, sont rendus possible grâce à leur organisation collatérale, ainsi qu'à de nombreuses interconnexions entre cellules proches. Dans les tubes criblés du phloème, le mouvement de la sève (*i.e.* le sucrose, les polyols, les acides aminés, etc.) est généré par un flux de masse, lui-même induit par les variations de pression osmotique au niveau des zones de chargement (*i.e.* les feuilles sources) et des zones de déchargement (*i.e.* les organes puits). Cette hypothèse, établie depuis maintenant presque un siècle, est toujours valide et a été reprise comme postulat de départ dans de nombreuses études. Cependant, les études consacrées à la preuve de ce concept sont relativement peu nombreuses et tardives, son caractère intuitif et plausible ayant largement contribué à son acceptation (Knoblauch et al., 2016; Knoblauch & Peters, 2010). Néanmoins des alternatives et des compléments à l'hypothèse de Münch ont été proposés. De manière générale, l'ensemble de ces travaux de recherche ont contribué à étayer l'hypothèse initiale de Münch (Knoblauch & Peters, 2010).





**Figure 2** - Schéma de la dynamique du transport phloémien

Les photoassimilats accumulés dans les feuilles sources lors de la photosynthèse sont chargés dans le phloème via les cellules compagnes (A). Le différentiel positif de pression osmotique dans les tubes criblés créé à la suite du chargement des photoassimilats entraîne un appel d'eau mécanique depuis le xylème (B). L'appel d'eau depuis le xylème génère un flux de masse qui entraîne les photoassimilats à distance de leur point de chargement (C). Les photoassimilats sont déchargés au niveau des organes puits (e.g. jeunes feuilles, fruits ou racines) où ils sont utilisés pour la croissance ou le stockage (D). Le différentiel négatif de pression osmotique dans les tubes criblés crée à la suite du déchargement des photoassimilats entraîne un mouvement d'eau vers le xylème (E). L'eau ainsi recyclée dans le xylème est de nouveau disponible pour entretenir le flux de masse dans le système phloémien. Adaptation de (Dinant & Lemoine, 2010; Windt et al., 2006).

## 2.2. Implication dans diverses fonctions vitales des plantes

### 2.2.1. Le phloème comme réseau d'alimentation

La conception fonctionnelle initiale du phloème renvoie à un réseau de transport des nutriments et métabolites mobilisés pour **l'homéostasie, la croissance et le développement des plantes**. Depuis plusieurs décennies, l'étude du contenu de la sève phloémienne a permis l'identification et la quantification des nutriments transportés, notamment dans les Poacées (Fukumorita & Chino, 1982; Hayashi & Chino, 1986), les Fabacées (Sandström & Pettersson, 1994), les

Cucurbitacées (Haritatos et al., 1996; B. Zhang et al., 2010), les Crucifères (Deeken et al., 2002), les Rosacées (Nadwodnik & Lohaus, 2008), et dans diverses conditions biotiques et abiotiques. La vision d'ensemble offerte par ces travaux montre que les profils des nutriments transportés par le phloème sont variables en fonction des espèces et de l'environnement. Les principaux nutriments identifiés dans le phloème sont des composés organiques, majoritairement **des sucres**, issus de la photosynthèse, et **des acides aminés** (Dinant et al., 2010; Dinant & Lemoine, 2010). D'autres molécules nécessaires à la vie des plantes ont été détectées, notamment des composés organiques comme **des polyamides, des acides organiques** ainsi que **des vitamines et des ions inorganiques** (*e.g.* potassium, calcium et phosphore) et des micronutriments (*e.g.* le fer, le manganèse et le zinc) (Dinant et al., 2010).

### 2.2.2. Le phloème comme réseau de communication intégratif

Le rôle du phloème ne se limite pas à l'allocation systémique de nutriments. Le phloème intègre également une variété de **voies de signalisation**, établissant ainsi une communication interorganes coordonnée et permettant la régulation de différents processus physiologiques, ayant trait à la **croissance**, au **développement** et aux **mécanismes de défense** (Lewis et al., 2022; Robert Turgeon & Wolf, 2009; Twayana et al., 2022). Les voies de signalisation sont déclenchées par la perception de stimuli endogènes et exogènes ; elles mobilisent **des phytohormones** (Davies, 2010; Gray, 2004), **des ARN** (Brant & Budak, 2018; Yoo et al., 2004), **des protéines** (De Schepper et al., 2013; Turnbull & Lopez-Cobollo, 2013), **des ions** (Fromm & Lautner, 2007; Hedrich et al., 2016) **des espèces réactives de l'oxygène** (W. G. Choi et al., 2017; Gilroy et al., 2016).

Grâce à des expérimentations reposant sur la transmission du flux de sève chez plantes greffées ainsi que sur des applications exogènes de phytohormones, une multitude de travaux de recherche a prouvé que le phloème était une voie de transport notable pour plusieurs phytohormones majeures, les cytokinines (CK), l'acide jasmonique (JA), les gibbérellines, l'auxine (AUX), l'acide abscissique (ABA) et l'acide salicylique (SA) (Baker, 2000; Lacombe & Achard, 2016; Aart J.E. Van Bel & Gaupels, 2004). La présence de l'ABA, du JA et du SA dans la sève phloémienne renforce l'hypothèse du rôle du phloème dans la réponse à des stress environnementaux et dans la défense contre les insectes herbivores et les agents pathogènes (Davies, 2010; Gray, 2004). En effet ces hormones sont largement connues pour leurs rôles dans la médiation des mécanismes de défense en réponse à des stress biotiques et abiotiques. L'ABA, le JA et le SA sont respectivement connus pour leur implication dans la réponse à des stress environnementaux, dans la défense contre les insectes herbivores et certains agents pathogènes et dans la réponse immunitaire à différents agents pathogènes (Davies, 2010; Gray, 2004).

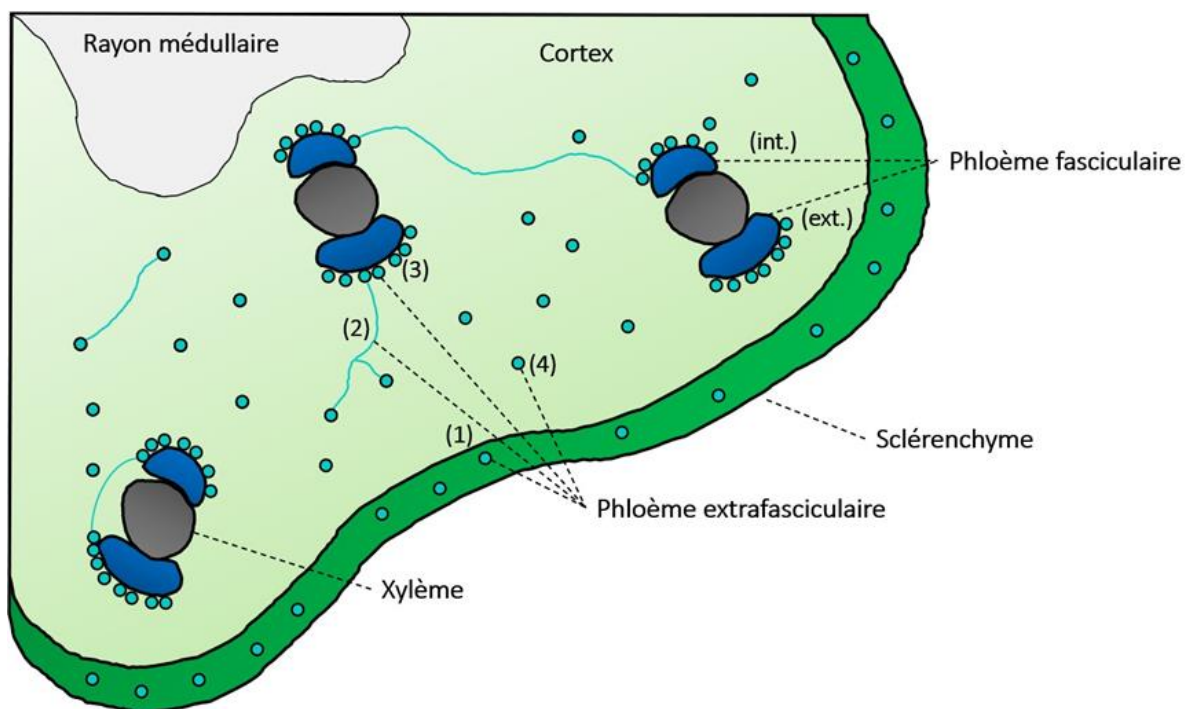
Le phloème intègre aussi des voies de signalisation basées sur la **transduction de signaux électriques** (Fromm & Lautner, 2007; Hedrich et al., 2016). Différents éléments mobiles, notamment des phytohormones, des ions calciques et des espèces réactives de l'oxygène sont recrutés (Gilroy et al., 2016). La transduction des signaux électriques a lieu sous la forme d'ondes de potentiel électrique, via divers **canaux ioniques** (Dinant & Lemoine, 2010). Les signaux électriques participent à des **processus physiologiques variés** (régulation de la respiration, de la photosynthèse, de la croissance, stress biotiques et abiotiques, etc.) et sont déclenchés consécutivement à la perception de stimuli également variés (Fromm & Lautner, 2007) comme des chocs thermiques et mécaniques (W.-G. Choi et al., 2016), des blessures infligées par des insectes (Toyota et al., 2018), etc. L'émergence d'un tel mode de communication chez les végétaux reflète très probablement une adaptation des plantes en réponse à des stress exogènes et nécessitant une **réaction très rapide**, nettement supérieure à celle des signaux chimiques (Fromm & Lautner, 2007), pour préserver leur intégrité physique.

Les premiers travaux de recherches ayant fait état de la **présence d'ARN** dans des échantillons de sève phloémienne remontent à plus de trente ans, mais durant de nombreuses années la détection d'ARN était souvent interprétée comme une contamination des échantillons de sève phloémienne. Cependant, de récentes analyses chez de nombreuses espèces ont apporté la preuve que ces populations d'ARN sont **des éléments authentiques** de la sève phloémienne (Kehr & Buhtz, 2008; Omid et al., 2007). Les caractérisations des profils transcriptionnels d'extraits de sève phloémienne chez des espèces végétales divergentes ainsi que chez des plantes greffées montrent clairement que les populations d'ARN du phloème sont propres à ce tissu et peuvent être transportées sur de longues distances, à **l'échelle systémique** (Ham & Lucas, 2017; Kehr & Buhtz, 2008; Yoo et al., 2004). Ces découvertes ont conduit à la proposition d'un nouveau paradigme pour le transport phloémien, où des protéines et différentes espèces d'ARN sont **des supports d'information**, au même titre que les phytohormones ou les ions calciques, et intègrent des réseaux de communication sur de longues distances (Omid et al., 2007). A titre d'exemple, les ARN CmGAIP et CmSUTP1, identifiés chez la citrouille (*Cucurbita maxima*), sont respectivement impliqués dans le développement des feuilles et dans le transport de sucrose (Kehr & Buhtz, 2008). Ces découvertes ont indéniablement joué un rôle décisif dans l'amélioration de la compréhension du système phloémien et ont ouvert la voie à un nouveau champ de recherche. La caractérisation complète d'un ARN du phloème est une entreprise complexe puisqu'elle consiste à retracer l'intégralité de son parcours, de l'organe qui l'a produit à celui qui va le réceptionner. L'identification de l'ensemble des étapes sous-jacentes et des mécanismes associés (*e.g.* le transport et l'adressage de ces ARN) sont peu connus. Chez la citrouille, la protéine *C. maxima* Phloem SMALL RNA BINDING PROTEIN1 (CmPSRP1) est la seule protéine de transport de petits ARN non codants connue chez les plantes (Zhu et al., 2019). Les connaissances actuelles sur ce sujet

restent très parcellaires et aucun ARN du phloème n'a été complètement caractérisé.

### 2.3. Le système phloémien particulier des cucurbitacées

En raison des dimensions remarquables de leurs éléments criblés et de leurs exsudations de sève copieuses, les cucurbitacées sont, depuis des décennies, utilisées comme modèles pour étudier l'anatomie et la physiologie du phloème (Cronshaw et al., 1981; Mullendore et al., 2010; Walz et al., 2004). **Remarquablement, les cucurbitacées possèdent deux systèmes phloémiens [figure 3] et jusqu'à ce jour, cette caractéristique anatomique n'a été observée qu'au sein de cette famille (R Turgeon, 2016).**



**Figure 3** - Schéma d'une coupe transversale d'une tige de citrouille

Le phloème fasciculaire (bleu foncé) est organisé de façon bicollatérale autour du xylème (gris) et présente des tubes criblés internes (int.) et externes (ext.). Le phloème extrafasciculaire (bleu clair) intègre : des tubes criblés ectocycliques (1), des éléments criblés transversaux formant des jonction latérales (traits bleu clair) reliant les différents éléments du phloème extrafasciculaire (2), des éléments criblés périphériques et adjacents aux faisceaux du phloème fasciculaire (3), des éléments criblés entocycliques à l'intérieur anneau du sclérenchyme (4). Adaptation de (C. Zhang et al., 2012).

**Le phloème fasciculaire** (Fascicular Phloem, FP) est composé d'un ensemble de faisceaux regroupant des tubes criblés suivant une **organisation radiale et bi-collatérale** autour du xylème, avec des faisceaux conducteurs internes et externes. **Le phloème extrafasciculaire** (Extrafascicular Phloem, EFP) est une composante anatomique propre aux cucurbitacées. L'EFP consiste en un réseau complexe de tubes criblés, **dispersés à travers les tissus pariétaux** de la plante formant le cortex. **Différentes configurations spatiales** de ce réseau ont émergé au cours de l'évolution. Selon les espèces, une ou plusieurs de ces configurations spatiales peuvent être présentes et représentées dans des proportions différentes (R Turgeon, 2016). Dans sa globalité, l'EFP est composé de tubes criblés en périphérie proche des faisceaux du FP, de tubes criblés entocycliques et ectocycliques, respectivement situés en périphérie interne et externe de l'anneau du sclérenchyme, et de tubes criblés transversaux reliant ces différents éléments entre eux ainsi qu'aux faisceaux du FP. Les unités fonctionnelles du transport du FP et de l'EFP, les complexes CC-SE, sont similaires en termes de forme et de diamètre (C. Zhang et al., 2012).

Des récentes études s'accordent sur le fait que les deux systèmes phloémiens des cucurbitacées ont des **compositions différentes** et que, par conséquent, leurs fonctions divergent (Lopez-Cobollo et al., 2016; B. Zhang et al., 2010; C. Zhang et al., 2012). Sur la base de cet **isolement fonctionnel**, il est établi que le FP et l'EFP sont essentiellement indépendants, bien que physiquement adjacents (Lopez-Cobollo et al., 2016). Le FP est clairement établi comme étant le réseau spécialisé dans le **transport longue distance des photoassimilats**. L'EFP est largement reconnu comme étant une **structure de défense** spécialisée contre les attaques des insectes herbivores (Gaupels et al., 2012; Gaupels & Ghirardo, 2013; Konno, 2011; Robert Turgeon & Oparka, 2010; B. Zhang et al., 2010). Contrairement au FP, l'EFP ne transporte pas son contenu à travers la plante (Schnieder et al., 2022). Suite aux blessures infligées par des insectes herbivores, l'EFP maintenu sous haute pression exsude abondamment un latex contenant un cortège de molécules répulsives ou toxiques pour ces bioagresseurs, notamment des **métabolites secondaires** (*e.g.* cucurbitacines, alcaloïdes, terpénoïdes) (R Turgeon, 2016) et des **protéines** (*e.g.* diverses enzymes, lectines) (Konno, 2011). Une quantité relativement faible de monosaccharides (hexoses) est transférée dans l'EFP, vraisemblablement requis pour la synthèse de composés secondaires (R Turgeon, 2016). Le fait que les faisceaux de l'EFP soient absents des principaux sites de chargement des photoassimilats, situés dans les nervures principales des feuilles sources, corrobore l'hypothèse selon laquelle l'EFP n'est pas impliqué dans le transport des sucres (Robert Turgeon & Oparka, 2010). Très récemment, un autre travail de recherche a abouti à la même conclusion (Schnieder et al., 2022).

En dépit d'un amoncellement conséquent d'études autour du système phloémien des cucurbitacées, la configuration spatiale de ce réseau vasculaire reste étonnamment **vague**. Remarquablement, la **représentation graphique de référence** pour l'anatomie vasculaire des cucurbitacées est une coupe transversale d'une tige de *C. maxima* réalisée lors de la première moitié du siècle dernier par Crafts et schématisée par Turgeon environ soixante-dix ans plus tard (Robert Turgeon & Oparka, 2010; C. Zhang et al., 2012). Toujours chez *C. maxima*, une observation du même type de coupe par stéréo microscopie en lumière visible et en fluorescence a été documentée assez récemment et conforte la représentation de référence (B. Zhang et al., 2010). La spatialisation du système phloémien dans les feuilles des cucurbitacées **est peu documentée**. Cependant, chez le melon, une visualisation des veines mineures par microscopie confocale suggère que l'organisation des vaisseaux phloémiens dans les feuilles : **(i)** est similaire à la représentation de référence et **(ii)** se simplifie en progressant vers la pointe des feuilles (Kanvil et al., 2017). Toutefois et très récemment, une entreprise de caractérisation de l'anatomie du système phloémien dans les racines et la tige du concombre abonde dans le même sens (Schnieder et al., 2022).

## 2.4. Qualités nutritives et potentiel délétère de la sève pour les pucerons

La sève phloémienne est l'unique ressource alimentaire pour de nombreuses espèces de pucerons, dont les **exigences nutritionnelles** sont plus ou moins strictes. La ressource alimentaire offerte par le phloème présente un **déséquilibre nutritionnel** important pour les pucerons puisqu'elle est riche en sucres et relativement pauvre en acides aminés (Sandström et al., 2000). De plus, elle présente une **variabilité considérable** en fonction de l'espèce végétale, du stade de développement de la plante, de la photopériode, des changements de saisons ainsi que des stress biotiques et abiotiques mais aussi à l'échelle des éléments criblés (Dinant et al., 2010; Douglas, 2006; Sun et al., 2020). Les pucerons sollicitent différents mécanismes pour **introduire une stabilité** dans la composition de leur ressource alimentaire qui leur soit favorable. La **symbiose** bien connue avec la bactérie *Buchnera aphidicola* permet aux pucerons d'acquérir différents **acides aminés essentiels** que leur plante hôte ne peut pas synthétiser. Remarquablement, les pucerons peuvent aussi **manipuler localement**, au niveau de leur site d'alimentation, le métabolisme de leur plante hôte et **modifier sa composition** pour améliorer la couverture de leurs besoins nutritionnels. Ce mécanisme original a été décrit par un ensemble conséquent de travaux de recherches sur des systèmes plantes-pucerons variés (H. H. Cao et al., 2016; Eleftherianos et al., 2006; Giordanengo et al., 2010; Sandström et al., 2000; Telang et al., 2002).

La sève phloémienne contient une variété de **métabolites secondaires** (PSM, acronyme de *Plant Secondary Metabolites*) potentiellement délétères pour des pucerons **non adaptés** à leur plante hôte. Par exemple, au sein des espèces d'intérêt économique notable, les **glucosinolates**, les **alcaloïdes** et les **cucurbitacines** sont autant d'exemples de composés défensifs associés intimement et respectivement aux *Brassicaceae*, aux *Solanaceae* et aux *Cucurbitaceae* (Gaupels & Ghirardo, 2013; Nalam et al., 2019; J. S. Ramsey et al., 2010). Chez les légumineuses, une variété de PSM ont fait l'objet d'études pour leurs propriétés répulsives vis-à-vis des insectes, notamment des **flavonoïdes** et des **saponines**. Cependant l'état actuel de la littérature ne présente pas clairement de PSM spécifiques des légumineuses (Goławska et al., 2010; Goławska, Kapusta, et al., 2008; Goławska, Łukasik, et al., 2008; Kamphuis et al., 2013).

Le processus d'ingestion de la sève phloémienne par les pucerons est **de nature passive** et suggère une absence de sélectivité des composés absorbés, et donc une absorption systématique et inévitable des PSM en même temps que les nutriments. Cette stratégie défensive des plantes visant à rendre l'unique source nutritive des pucerons impropre à leur consommation exerce une pression de sélection considérable sur ces derniers. La majorité des pucerons connus se sont **spécialisés sur une seule ou sur un nombre restreint d'espèces végétales apparentées** (*e.g.* *A. pisum* est une espèce de pucerons spécialisée sur quelques membres de la famille des fabaceae comme le pois, le trèfle ou encore la luzerne), il est très probable que ces derniers aient développé des mécanismes de tolérances aux PSM efficaces et variés. Malgré un état des connaissances assez limité sur ce sujet, il est possible de discriminer plusieurs mécanismes d'adaptation originaux des pucerons aux PSM dans leur globalité. Certains mécanismes **contrecarrent la synthèse des PSM** qui a été initiée lors de l'invasion du puceron et d'autres confèrent une **tolérance accrue aux PSM** (Züst & Agrawal, 2016).

L'usage d'enzymes de **détoxification**, comme les **glutathion S-transférases** (GST) ou les **cytochromes p450 monooxygénases** (P450) est la stratégie la plus répandue et également la plus étudiée (Kettles & Kaloshian, 2016; Laskowska et al., 1999; J. S. Ramsey et al., 2010; Thorpe et al., 2016). Le répertoire d'enzymes de détoxification des espèces de pucerons généralistes (*e.g.* *M. persicae*) devrait être plus conséquent et plus varié que celui des pucerons spécialistes puisqu'ils sont confrontés à un éventail de PSM lui aussi plus conséquent et plus diversifié. Une analyse protéomique comparative entre le puceron spécialiste *A. pisum* et le puceron généraliste *Aulacorthum solani* (puceron de la digitale ou de la pomme de terre) a révélé une différence remarquable dans l'activité enzymatique de leur GST respectives et conforte l'hypothèse que le répertoire des enzymes de détoxification des pucerons reflète leur préférence d'hôte (Francis et al., 2001). Une analyse transcriptomique comparative entre *A. pisum* et *M. persicae* abonde dans le même sens et conforte

l'hypothèse d'un répertoire enzymatique de détoxification des PSM plus étendu chez les pucerons généralistes (J. S. Ramsey et al., 2010). D'autres stratégies de tolérances comme le développement de leurres moléculaires ayant une affinité avec les toxines végétales ou leur séquestration sous une forme non toxique (*e.g.* la neutralisation des glucosinolates par *Brevicoryne brassicae* ou *Lipaphis erysimi*) ont été identifiées mais restent moins connues (Züst & Agrawal, 2016).



## 3. Les interactions plante-puceron : de l'échelle tissulaire à l'échelle moléculaire

### 3.1. Une stratégie furtive pour atteindre le phloème

Les pucerons sont des **insectes phloémophages** et l'accès à leur source d'alimentation est un véritable challenge. Leurs performances biologiques sont asservies à leur capacité à accéder au phloème, en évitant ou en sabotant les réponses de défense des plantes, et à maintenir cet accès aux cellules phloémiennes sans compromettre la survie de ces dernières (Giordanengo et al., 2010). Pour atteindre leur objectif, les pucerons utilisent de façon coordonnée leurs **pièces buccales spécialisées**, ainsi qu'un **répertoire moléculaire original**.

Les pièces buccales des pucerons ont une architecture complexe. Elles renferment un **faisceau de stylets flexibles** que l'insecte enfonce dans le tissu végétal pour atteindre le phloème. Une gaine protectrice et télescopique, ou **rostre**, renferme le faisceau de stylets quand ils sont compactés. Ce faisceau comporte deux stylets mandibulaires entourant deux stylets maxillaires et constitue le **canal alimentaire**. Le **canal salivaire**, parallèle au canal alimentaire communique avec les glandes salivaires situées dans la tête du puceron. Les canaux alimentaires et salivaires sont séparés sur la quasi-totalité de la longueur des stylets et fusionnent à la pointe en un **canal commun** (Leonhardt, 1940; Mittler, 1957). Lorsque le puceron sonde les différentes couches cellulaires à la recherche de sa source nutritive, les stylets se déploient et coulissent dans une gouttière dorsale du rostre, le **labium** (Balch, 1952; Bradley et al., 1962; Ponsen, 1972).

#### 3.1.1. La salive des pucerons : pièce maîtresse d'une effraction furtive

Préalablement à la phase de sondage, les pucerons sécrètent un **bouchon salivaire** puis créent une brèche entre deux cellules épidermiques végétales à l'aide de leur rostre. Une fois dans les tissus de la plante, les stylets progressent par des mouvements coordonnés entre les cellules vers les couches cellulaires plus profondes. Tout au long de leur chemin vers le phloème, les stylets **se courbent** et réalisent un **parcours apoplastique** pendant que le canal salivaire éjecte une **salive gélifiante** qui forme une gaine autour des stylets. Au cours de leur chemin, les stylets de pucerons perforent brièvement les cellules, dans lesquelles le canal salivaire éjecte une **salive aqueuse** (W. Fred Tjallingii, 2006). La stratégie de pénétration des stylets des pucerons dans les tissus végétaux est révélée par la **technique d'électropénétrographie** (EFP). Cette technique consiste à intégrer l'insecte et la plante dans un circuit électrique par l'intermédiaire de deux électrodes, rendant possible le suivi de la pénétration des stylets

dans les tissus végétaux grâce à l'enregistrement des variations de résistance électrique induites par les différentes activités du puceron sur sa plante hôte (McLean & Kinsey, 1964; W F Tjallingii, 1978).

**La salive gélifiante** des pucerons est sécrétée tout au long de la progression des stylets dans l'espace intercellulaire. Au contact de l'air circulant dans l'apoplaste, la salive gélifiante s'oxyde et la salive gélifiée forme **une gaine tubulaire continue qui enveloppe les stylets** sur toute leur longueur au fur et à mesure de leur progression vers le phloème (Will & Vilcinskas, 2015). La salive gélifiée est un outil sophistiqué et polyvalent des pucerons (W. Fred Tjallingii, 2006; Walling, 2008; Will et al., 2012, 2013). La gaine rigide autour constitue une **couche protectrice** des stylets, relativement souples et fragiles, au contact des couches cellulaires qu'ils rencontrent lors de leur progression. Les **propriétés lubrifiantes** de cette gaine facilitent les mouvements des stylets, limitant ainsi l'abrasion voire la rupture des cellules végétales rencontrées. Lors du sondage du contenu des cellules, la salive gélifiée forme un **joint d'étanchéité** au niveau des différents sites de ponction, empêchant le déversement du contenu cytoplasmique des cellules blessées dans l'apoplaste, dont certains composés sont associés aux **mécanismes de défenses**. La salive gélifiante est essentiellement composée de **protéines enzymatiques** (dont des phénoloxydases, des peroxydases, des pectinases et des béta-glucosidases) qui facilitent la dégradation des parois cellulaires, des phospholipides et des hydrates de carbone conjugués (Giordanengo et al., 2010). La protéine majoritairement présentée est la **protéine de structure SHP** (*structural sheath protein*). Elle assure vraisemblablement l'intégrité de la gaine salivaire. Son rôle a été démontré chez *A. pisum* par extinction de l'expression du gène *shp* et sa présence est avérée chez plusieurs espèces de pucerons (Will & Vilcinskas, 2015).

**La salive aqueuse** des pucerons est sécrétée **durant les piqûres** des cellules du mésophylle et **lors de la phase d'ingestion**, lorsque le contact entre les stylets des pucerons et les cellules conductrices du phloème est établi (W. Fred Tjallingii, 2006; Will et al., 2013). A l'inverse de la salive gélifiée, **la composition protéique de cette salive est variable**. La salive aqueuse contient plusieurs protéines présentant une variété d'activités biochimiques, dont des pectinases, des cellulases, des polyphénoloxydases, des peroxydases, des lipases mais également des protéines chélatrices des ions calciques qui **favorisent le long processus d'infestation** (Miles & Peng, 1989; Nalam et al., 2019). Parmi les différentes propriétés de la salive aqueuse, sa capacité à séquestrer les ions calciques pour interférer avec les signaux physiologiques à l'origine de **l'occlusion des tubes criblés** du phloème est la mieux documentée (Carolan et al., 2009; Will et al., 2007, 2013).

## 3.2. Le dialogue moléculaire entre les effecteurs aphidiens et le système immunitaire de la plante

### 3.2.1. Un modèle de référence pour les interactions plante-puceron

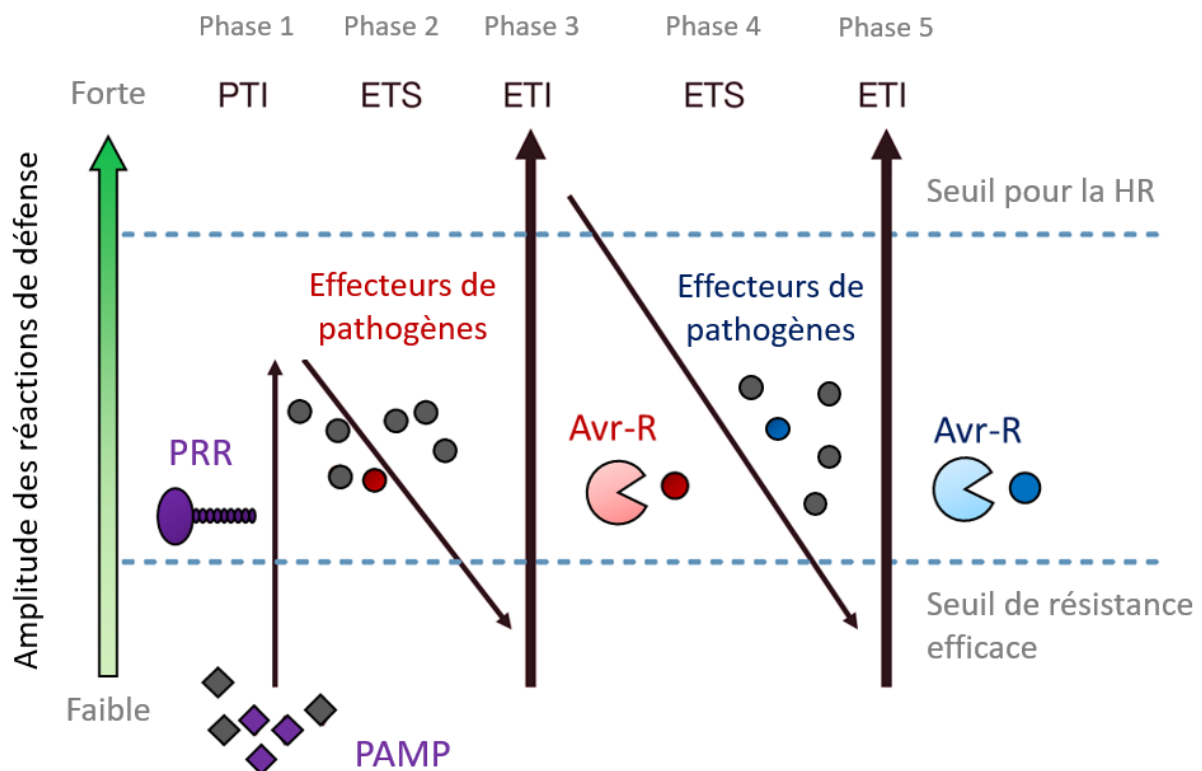
La stratégie furtive mise en place par les pucerons, c'est-à-dire principalement un cheminement apoplastique pour atteindre leur source d'alimentation, bien que sophistiquée, se solde généralement **par un échec**. L'intrusion du puceron, et en particulier les composés salivaires, est perçue par le **système immunitaire** de la plante. Un **dialogue moléculaire** complexe s'établit qui présente des similitudes avec le dialogue moléculaire majoritairement décrit pour des relations pathogènes/plantes (Kaloshian & Walling, 2005).

**Différents cadres conceptuels** ont été proposés pour décrire de façon simplifiée les mécanismes orchestrant le dialogue moléculaire entre les plantes et leurs bioagresseurs, et inspirer des travaux de recherches de plus en plus pertinents.

Le premier cadre décrit la **théorie des interactions gène-pour-gène**. Il a été établi par Flor au cours de la première moitié du siècle dernier et propose une description de phénotypes de l'interaction dans le cadre de la génétique mendélienne (Flor, 1971). Sur la base d'observations de la répartition de phénotypes de résistance à la rouille chez le lin, Flor postule que l'issue de l'interaction moléculaire entre une plante et un agent pathogène repose sur la reconnaissance d'**un seul gène dominant de cet agent pathogène** (ou gène d'avirulence) par **un seul gène de résistance de l'hôte** (gène de résistance ou gène-R). Conformément à cette théorie, l'adaptation d'un agent pathogène à une plante reflète la **perte de reconnaissance** d'un gène-R à l'échelle moléculaire.

Le modèle suivant, proposé par Jones et Dangl en 2006, est fondé sur le postulat que le système immunitaire est contrôlé par des récepteurs codés par des gènes et impliqués dans la perception d'une invasion [figure 4]. Ce modèle d'évolution des interactions plante-pathogène, usuellement désigné comme « **modèle zigzag** » ou encore « **molecular arm race** », unifie la théorie de Flor et la perception des éliciteurs de agents pathogènes sous un même paradigme avec une **perspective évolutive** (Jones & Dangl, 2006). En 2013, Dangl propose un **élargissement du modèle** en introduisant les nématodes et les **insectes piqueurs suceurs** (Dangl et al., 2013). **Chez les pucerons**, la caractérisation au **niveau moléculaire** de **trois gènes appartenant à la famille NLR** (Nucleotide-binding domain Leucine-rich Repeat containing) : **(i) Mi1-2** de la tomate qui confère la résistance à la fois à

*Macrosiphum euphorbiae* ainsi qu'à d'autres hémiptères (Rossi et al., 1998) et à plusieurs espèces de nématodes à galles (Vos et al., 1998), (ii) **Vat-1** du melon qui confère la résistance à *A. gossypii* ainsi qu'aux virus qu'il transmet (Catherine Dogimont et al., 2014) et (iii), **Adnr1** du blé qui confère la résistance à *Diuraphis noxia* (Nicolis & Venter, 2018) appuient ce choix (Boissot, Schoeny, et al., 2016; Rossi et al., 1998). Egalement, le gène **NLR Bph26** du riz qui confèrent la résistance à la cicadelle *Nilaparvata lugens* (Tamura et al., 2014) renforcent davantage le choix de l'intégration des insectes piqueurs suceurs au modèle zigzag (Dangl et al., 2013).



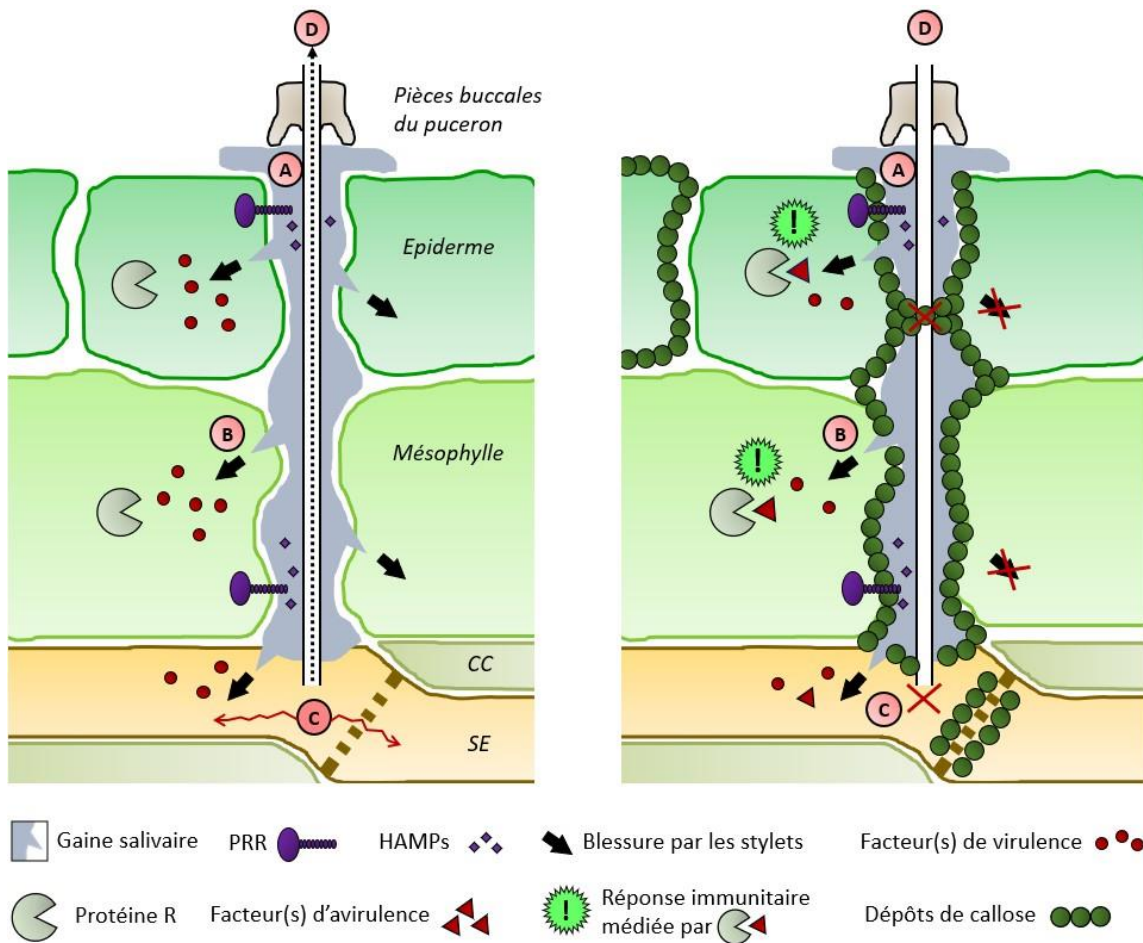
**Figure 4** – Le cadre conceptuel de référence pour la description du système immunitaire des plantes

**Lors de la première phase**, la perception par la plante de motifs moléculaires associés à des agents pathogènes (PAMP, acronyme de « Pathogen-associated molecular patterns ») via des récepteurs de reconnaissances transmembranaires ou PRR déclenche la première vague de réponses immunitaires ou PTI (PAMP-triggered immunity). **Lors de la seconde phase**, des agents pathogènes adaptés délivrent des effecteurs qui interfèrent avec la PTI ou favorisent des performances biologiques, rétablissant les défenses des plantes à leur niveau d'expression basal. Ces effecteurs, dits effecteurs de virulence, établissent un état d'interaction plantes-pathogènes favorable aux agents pathogènes ou ETS (Effector-triggered susceptibility). **Lors de la troisième phase**, la perception par la plante d'un effecteur d'avirulence (sphère rouge) via une protéine de résistance NLR (camembert rouge, désignée 'Avr-R') active la seconde vague de réponses immunitaires ou ETI (Effector-triggered immunity). Cette seconde vague de défense est une version amplifiée et prolongée de la PTI, réputée pour franchir le seuil d'induction d'une réaction d'hypersensibilité (HR) pour enrayer la progression des

agents pathogènes. **Lors de la quatrième phase**, la sélection favorise les agents pathogènes non reconnus par la protéine R (adaptation par la perte de l'effecteur d'avirulence) ou capables d'interférer avec l'ETI (émergence d'un nouvel effecteur de virulence acquis via un flux de gènes horizontal, représenté par une sphère bleue) et rétablissent l'ETS. **Lors de la cinquième phase**, la sélection favorise les plantes dont un isoforme de la protéine de résistance NLR (camembert bleu, désignée 'Avr-R') peut percevoir un nouvel effecteur d'avirulence des agents pathogènes, rétablissant ainsi l'ETI. La mécanique d'alternance entre l'ETI et l'ETS est entretenue par l'exercice de pressions de sélections mutuelles au sein du système plantes-pathogènes. Adapté de *Jones & Dangl, 2006*.

Le modèle zigzag [Figure 4] introduit une avancée conceptuelle majeure où les mécanismes de défenses induits par la **reconnaissance des éliciteurs** d'agents pathogènes peuvent être **contrecarrés par l'action d'effecteurs**. Ces effecteurs sont sécrétés à l'intérieur des cellules de l'hôte et induisent des changements physiologiques chez la plante dont les agents pathogènes sont les bénéficiaires. Actuellement, le modèle zigzag est le paradigme de référence pour la description du dialogue moléculaire entre les plantes et les agents pathogènes. Sa **valeur descriptive élégante et pédagogique** a fortement contribué à son acceptation générale par les communautés scientifiques liées à l'étude de l'immunité des plantes.

Au sens immunitaire du terme, l'issue de l'interaction entre un effecteur et les défenses de leur plante hôte est binaire et correspond à **deux réactions extrêmes et opposées** : une réaction dite **compatible** et une réaction dite **incompatible**. Dans le cas d'une interaction compatible [figure 5, à gauche], les modulations physiologiques induites par un effecteur constituent une manipulation et une altération des défenses de la plante hôte et facilitant l'installation durable de l'agent pathogène. **Dans ce premier cas**, l'effecteur contribue à une issue d'interaction plante-pathogène favorable au pathogène, alors désigné par le terme de **facteur de virulence**. Dans le cas d'une interaction incompatible [figure 5, à droite], les modulations physiologiques induites par un effecteur déclenchent des mécanismes de défense de l'hôte et pénalise la survie, la croissance, et la multiplication du pathogène. **Dans ce deuxième cas**, l'effecteur contribue à une issue d'interaction plante-pathogène défavorable au pathogène, alors désigné par le terme de **facteur d'avirulence**.



**Figure 5** – Confrontation de la stratégie d'invasion des pucerons avec le système immunitaire des plantes à l'échelle moléculaire dans le cas d'une interaction compatible ou incompatible

Les stylets des pucerons pénètrent entre deux cellules de la surface des feuilles. La salive gélifiée est sécrétée tout le long de la progression apoplastique des stylets en direction du phloème. Cette salive enrobe les stylets et scelle les sites de sondage des cellules pour empêcher le déclenchement des mécanismes de défenses de la plante. La salive aqueuse est sécrétée dans le cytoplasme des cellules sondées. **Dans le cas d'une interaction compatible (gauche)**, la salive aqueuse contient des éliciteurs qui annulent la PTI, la situation est dite ETS (Effector-triggered susceptibility). Aucun effecteur du flux salivaire injecté dans le cytoplasme des cellules sondées n'est reconnu par la protéine R cytoplasmique. Ces pucerons atteignent les tubes criblés du phloème (SE). La traversée des cellules compagnes (CC) adjacentes aux SE n'est pas systématique et n'est pas représentée dans ce schéma dans un souci de clarté. Les pucerons alternent des phases de sécrétions salivaires dans la lumière du SE (C) et d'ingestion de la sève phloémienne (D). **Dans le cas d'une interaction incompatible (droite)**, un effecteur salivaire est reconnu par une protéine R cytoplasmique. Cette reconnaissance moléculaire déclenche l'ETI (Effector-triggered immunity). La réponse immunitaire se manifeste dans un délai très court (de l'ordre de la minute) et sous plusieurs formes : des dépôts de callose dans la paroi cellulaire des cellules blessées mais aussi des cellules adjacentes et probablement des assiettes criblées. Adaptation de (Boissot, Schoeny, et al., 2016; Hogenhout & Bos, 2011; Will et al., 2013). Graphismes inspirés de (W. Fred Tjallingii, 2006).

La première ligne de défense active des plantes est formée par les récepteurs de reconnaissance de motifs moléculaires associés à une invasion ou **PRR** (*Pattern Recognition Receptors*), [figure 4]. Il s'agit de récepteurs transmembranaires capables de percevoir des motifs moléculaires caractéristiques des agents pathogènes ou **PAMP** (*Pathogen-associated molecular patterns*) en surface des cellules végétales. Par analogie, les **HAMP** (*herbivory-associated molecular patterns*) pour les insectes phytophages (Acevedo et al., 2015; Bonaventure, 2018; Stahl et al., 2018; Yates & Michel, 2018) concernent l'ensemble des composés de signalisation des insectes ravageurs (issus de la salive, des sécrétions orales, de l'exosquelette, des œufs et du miellat) pouvant entrer en contact avec leurs plantes hôtes quel que soit leur stade de développement et déclenchant des réactions de défense chez ces dernières (Mithöfer & Boland, 2008). **Dans la suite de mon manuscrit et en accord avec le modèle de Jones et Dangl, je désignerai par « HAMP » les éliciteurs de pucerons putativement attendus comme des composés de la salive gélifiante [figure 5]** (Dangl et al., 2013; Hogenhout & Bos, 2011; Kaloshian & Walling, 2016a).

Les PAMP/HAMP correspondent à des structures moléculaires qui assurent des fonctions essentielles et qui sont **largement conservées** chez une même classe de d'agents pathogènes/insectes phytophages. La plante reconnaît une invasion via la perception des PAMP par ses PRR et déclenche une variété de réactions basales de défense immunitaire participant à la résistance non-hôte et désignée communément par l'acronyme **PTI** (*PAMP-triggered immunity*). L'éventail de réponses standards associées à cette première vague de défense comprend une **production de ROS** (reactive oxygen species), un **dépôt de callose** autour des parois cellulaires et une **activation de gènes de défense** (Kaloshian & Walling, 2016b; Richards et al., 2010).

Si la littérature fait état d'une gamme remarquable de PAMP chez les champignons (*e.g.* la chitine et l'ergostérol), les oomycètes (*e.g.* les bêta-glucanes et les élicitines) et les bactéries (*e.g.* la flagelline, les protéines de choc froid et protéine d'élongation G EF-Tu), elle renseigne rarement les PRR en correspondance (Harvey et al., 2019; Nürnberger et al., 2004). Les rares couples PAMP/PRR caractérisés incluent la protéine bactérienne flg22 et le récepteur FLS2 (Flagelline sensitive 2) chez *A. thaliana*, la chitine de la paroi des cellules fongiques et les récepteurs CEBiP (chitin-elicitor binding protein) chez *Oryza sativa* (Hayafune et al., 2014), ou encore la protéine G EF-Tu et le récepteur EFR (Harvey et al., 2019; Zipfel et al., 2006). Dans le modèle zigzag, la formalisation de la PTI repose notamment sur le **couple flg22/FLS2**. En revanche, **aucun couple HAMP/PRR n'est actuellement caractérisé**. Toutefois, la protéine de structure SHP évoquée en **3.1.1** correspond à la définition d'un HAMP (Will & Vilcinskas, 2015).

La **seconde ligne de défense active des plantes** est formée par des **récepteurs intracellulaires** impliqués dans la reconnaissance spécifique de peptides, **injectés dans le cytoplasme** aussi appelés **effecteurs [figure 4]**. Chez les pucerons, les effecteurs sont contenus dans la **salive aqueuse** et injectés dans les cellules sondées lors du trajet des stylets vers le phloème, ainsi que dans le phloème lui-même **[figure 5]**. Le déclenchement de l'ETI est souvent associé à une **réponse d'hypersensibilité (HR)** locale, correspondant à une mort cellulaire programmée, affectant les cellules à proximité de la zone où la reconnaissance de la protéine d'avirulence a eu lieu (Heath, 2000). D'autres réponses physiologiques caractéristiques incluent l'activation d'autres gènes de défense en aval, ainsi qu'une production de ROS accrue et d'acide salicylique (Meyers et al., 2003). La réponse immunitaire déclenchée consécutivement à la perception d'agents pathogènes avirulents, appelée **ETI** (*Effector-triggered immunity*) dans le modèle d'évolution des interactions plantes-pathogènes, repose sur l'interaction hautement spécifique entre un effecteur d'avirulence et une **protéine de résistance** (protéine-R) de plante codée par un gène de résistance (R-gène) (Boller & Felix, 2009). La reconnaissance d'un effecteur par une protéine-R est à **caractère spécifique** et s'effectue de manière directe, ou via l'interaction avec un facteur tiers (Van Der Hoorn & Kamoun, 2008). Dans la majorité des cas, les gènes-R codent pour des **protéines cytoplasmiques** de la classe **NLR**. De manière générale, la manifestation de l'ETI est cohérente avec la théorie gène-pour-gène établie par Flor. Actuellement, un nombre limité d'effecteurs sont caractérisés et concernent majoritairement des champignons, des bactéries et des oomycètes ; et dans une moindre mesure des nématodes et des virus (Kapos et al., 2019). Chez les hémiptères et notamment **chez les pucerons, une vingtaine d'effecteurs putatifs** sont mentionnés dans la littérature. Cependant, leur caractérisation demeure partielle et les démonstrations de leur rôle d'effecteur repose exclusivement sur des preuves indirectes (voir **Partie 3.3**). En revanche, chez les plantes, **une centaine de gènes NLR sont bien étudiés** (dont 29 chez *Arabidopsis thaliana*, 18 chez le riz et 8 chez le blé) : **(i)** plus des deux tiers ont un effecteur identifié, **(ii)** l'interaction avec un effecteur a été effectivement démontrée dans la moitié des cas. Aucun des cas **(i)** ou **(ii)** ne concerne les hémiptères (Kapos et al., 2019).

### 3.2.2. Déviations par rapport au modèle de référence : vers une conception plus intégrative

La description de l'immunité des plantes proposée modèle zigzag présente **un certain nombre de limites** : **(i)** conceptuellement, le modèle zigzag **impose des termes strictes** qui donne une vision de l'immunité des plantes comme un **processus séquentiel** faisant intervenir **deux formes de résistance distinctes**, **(ii)** la PTI apparaît comme une **phase immunitaire statique** sur le plan évolutive



(i.e. théoriquement, les couples PAMP-PRR sont très conservés au sein de leurs espèces respectives), (iii) les éliciteurs endogènes, correspondant à des motifs moléculaires caractéristiques du soi modifié, ou **DAMP** (*Damage-Associated Molecular Patterns*) ne sont pas pris en compte par ce modèle, (iv) l'ETI apparaît comme une forme d'immunité **hautement spécifique** des pathosystèmes, (v) l'explicitation d'une discrimination stricte entre les PAMP et les effecteurs, (vi) la coexistence de deux points de vue différents dans le même modèle entretient la **conception erronée** discutée d'une immunité scindée en deux formes strictement différentes (i.e. les PAMP sont définis du point de vue du bioagresseur alors que les effecteurs définis du point de vue de la plante).

Il apparaît de plus en plus évident que les couples PAMP-PRR sont soumis à des variations dynamiques et que les PRR ne sont **pas nécessairement des récepteurs « anciens »** qui persistent au sein de leurs familles de plantes respectives. Le fait qu'un nombre conséquent de ces récepteurs montrent une variabilité au sein de la même espèce ou à une échelle taxonomique limitée suggère qu'ils sont **le produit d'une évolution récente** (Cook et al., 2015). De plus, la résistance conférée par des gènes codant pour des protéines NLR n'est pas nécessairement restreinte à un seul pathosystème (Białas et al., 2018; Cook et al., 2015). Le cas du **gène NLR *Mi-1.2*** présent chez la tomate constitue un exemple pertinent pour illustrer cette déviation par rapport au modèle de référence. En effet, *Mi-1.2* confère une résistance à quatre espèces d'insectes phloémophages phylogénétiquement très distants : les pucerons (*M. euphorbiae*), les nématodes à galles (*Meloidogyne spp*), les aleurodes (*Bemisia tabaci*) et les psylles (*Bactericerca cockerelli*) (Casteel et al., 2006; Nombela et al., 2003; Rossi et al., 1998; Vos et al., 1998). D'autres exemples notables sont **détaillés dans la revue de Cook, Mesarich et Thomma (2015)**. La détection d'une variété d'effecteurs par des gènes NLR est conceptuellement similaire à la reconnaissance des PAMP par les PRR, et la littérature relève de nombreux cas **où la discrimination stricte entre les PAMP et les effecteurs ne peut être maintenue**. La mécanique du dialogue moléculaire entre les plantes hôtes et leurs agents pathogènes respectifs intègre une variété de voies de signalisations qui ont lieu et évoluent simultanément. Les PAMP et PRR d'une part, et effecteurs et les récepteurs NLR d'autres part **coévoluent de façon parallèle**. Les deux branches de l'immunité des plantes, la PTI et l'ETI, **ne sont pas séquentielles** (Cook et al., 2015).

Un nouveau modèle, dit **modèle d'invasion**, a été proposé pour intégrer la multitude de résultats expérimentaux en déviation par rapport au modèle de référence. Le modèle d'invasion illustre l'immunité des plantes comme un système de surveillance pouvant percevoir une invasion. Ce modèle présente une vision plus intégrative et applicable à des systèmes d'étude plus contrastés du parasitisme à la symbiose (Cook et al., 2015). Il est explicitement décrit comme une vision moins résolutive du modèle zigzag permettant d'inclure les champignons nécrotrophes, les virus et les insectes broyeurs.

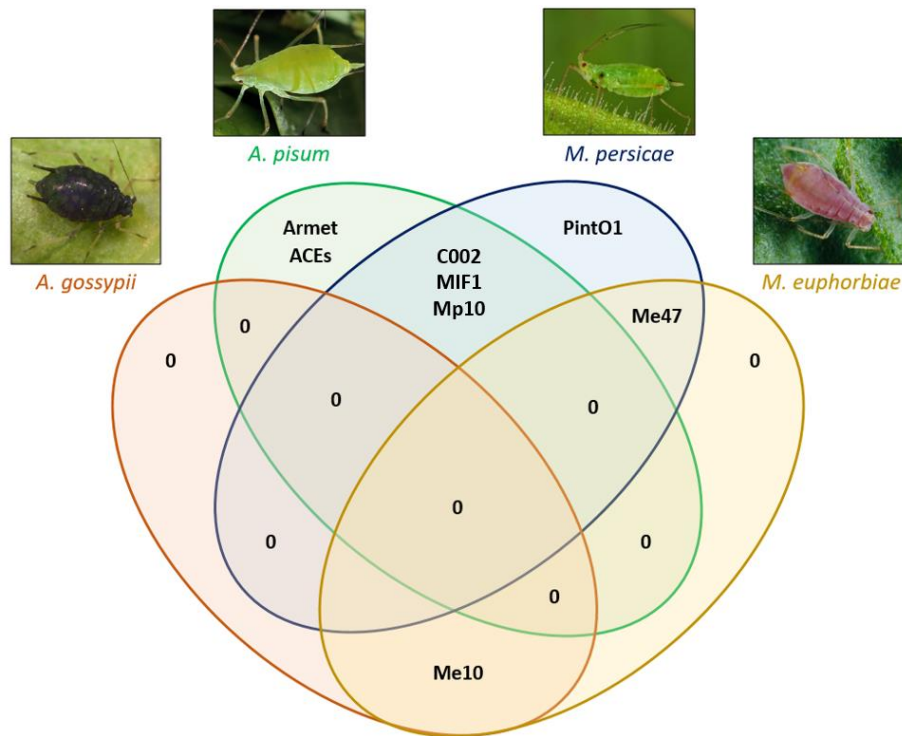
Les **DAMP** sont reconnus par des mécanismes similaires (Bigeard et al., 2015). Typiquement, ces signaux de dangers sont des fragments de parois cellulaires issus de la dégradation enzymatique et/ou mécanique des tissus végétaux par les bioagresseurs ou encore des éléments du contenu cytosolique déversé dans l'apoplaste suite à l'endommagement de la structure des cellules, comme des peptides (Boller & Felix, 2009). Jusqu'ici, les quelques DAMP mentionnés dans la littérature ont été identifiés sur la cohérence des réponses physiologiques consécutives à leur application sur le végétal avec le modèle des interactions plantes-pathogènes établi, sans que les PRR correspondant ne soient identifiés. Le seul couple DAMP/PRR bien caractérisé à ce jour est AtPEP1 et PEPR1, chez *A. thaliana* (Boller & Felix, 2009). L'échec de la plupart des tentatives d'identification de ces PRR, qu'il s'agisse des PAMP ou de DAMP, pourrait refléter une déviation par rapport aux modèles de reconnaissance établis.

### 3.3. Les effecteurs aphidiens : état de l'art

Dans ce manuscrit, je propose une vision intégrative et large du terme effecteur qui ici **désigne toute molécule sécrétée par un bioagresseur pouvant interagir avec les défenses inductibles de son hôte (en modulant la structure et/ou la fonction de leurs cellules) et jouant un rôle décisif, en positif ou en négatif, dans l'issue de l'interaction** (Hogenhout et al., 2009; Kamoun, 2006).

#### 3.3.1. Des connaissances très parcellaires

Durant la dernière décennie, une **combinaison d'approches génomiques, transcriptomiques et protéomiques** a permis d'établir **des catalogues d'effecteurs candidats conséquents**. Les **principales espèces de pucerons** concernées sont *A. pisum*, *M. persicae* et *M. euphorbiae* (Atamian et al., 2013; Bos et al., 2010; Boulain et al., 2018, 2019; Carolan et al., 2011; Harmel et al., 2008; Naessens et al., 2015). La grosseur des individus de ces espèces rend possible la récupération de salive ou l'extraction des glandes salivaires sans contraintes techniques majeures. De nombreux effecteurs potentiels ont été recensés mais **peu ont fait l'objet d'une étude fonctionnelle [figure 6]**. A une exception près, ces effecteurs n'ont pas d'annotations fonctionnelles et/ou ne présentent aucune similarité de séquences avec d'autres organismes déjà renseignés dans des bases de données.



**Figure 6** – Huit effecteurs notables caractérisés chez 4 espèces de pucerons

Sur ces 8 effecteurs, 7 ont été caractérisés chez les deux principales espèces modèles *A. pisum* ou *M. persicae*, 3 sont spécifiques d'une seule espèce et un seul a été identifié chez *A. gossypii*.

**L'étude fonctionnelle** des effecteurs candidats mobilise principalement les techniques d'**expression *in planta*** de façon stable ou transitoire, ou la **répression de la cible putative *in planta*** (Atamian et al., 2013; Bos et al., 2010; Rodriguez et al., 2014) ou encore la répression de l'expression de ces candidats via l'**interférence par ARN (RNAi)** directement dans le corps du puceron (Y. Chen et al., 2020; Mutti et al., 2006, 2008; Naessens et al., 2015; Y. Zhang, Liu, Francis, et al., 2022; Y. Zhang, Liu, Fu, et al., 2022). Les résultats obtenus par ces méthodes constituent des **preuves indirectes** de l'activité et la fonction des effecteurs étudiés. Actuellement, le recours à la modification génétique n'est **pas envisageable pour les pucerons**.

L'activité biochimique des effecteurs aphidiens identifiés reste **généralement inconnue** (ApC002, MpC002, Mp10, Mp42, Mp57, Mp58) ou à l'état de **prédictions** basées sur des homologies de séquences (ACE1, ACE2, MpMIF1). Ces valeurs prédictives restent précieuses dans la mesure où elle permet de réduire la taille conséquente des banques d'effecteurs potentiels à un niveau acceptable pour des validations fonctionnelles, mais également dans la mesure où elles permettent et de formuler et d'étayer des hypothèses pertinentes sur des catégories fonctionnelles putatives pour orienter de futurs travaux de recherches.

Davantage d'informations sont disponibles sur les phénotypes d'interactions plantes-pucerons et de ce fait, ces effecteurs sont souvent catégorisés dans la littérature selon leur impact sur diverses performances caractéristiques des pucerons (comme leur potentiel reproductif, leur capacité de survie ou encore leur activité d'alimentation). Certains d'entre eux élicitent les défenses des plantes et ont un **impact négatif** sur les performances des pucerons comme Mp10, Mp42, ACE1, ACE2, Mp57, Mp58 et GroEL (Chaudhary et al., 2014; Elzinga et al., 2014; Rodriguez et al., 2014; Wang, Luo, et al., 2015). Au contraire, d'autres effecteurs contribuent à la virulence des pucerons et **favorisent les performances biologiques** comme ApC002, MpC002, Me10, Me23, Armet et ApMIF1 (Atamian et al., 2013; Mutti et al., 2008; Naessens et al., 2015; Pitino & Hogenhout, 2013; Wang, Dai, et al., 2015; Wang, Luo, et al., 2015).

### 3.3.2. Focus sur les effecteurs les mieux connus

**C002** est le premier effecteur aphidien identifiée et décrit du point de vue fonctionnel dans la littérature (Mutti *et al.*, 2008). La réduction de l'expression de la protéine C002 d'*A. pisum* (ApC002) via la méthode d'interférence par ARN entraîne la mortalité d'*A. pisum* sur *Vicia faba*. Les profils EPG (Electropénétrographie) associés à cette réduction de l'expression montrent une perturbation générale du comportement alimentaire des pucerons et suggère fortement que la protéine ApC002 est **nécessaire pour l'établissement d'une phase d'alimentation durable**. Son homologue chez *M. persicae*, MpC002, induit une **augmentation de la fécondité** lorsqu'il est surexprimé de façon transitoire chez *Nicotiana benthamiana* (Bos *et al.*, 2010). Actuellement, **la fonction précise de C002 reste inconnue**.

**Armet** (*Arginine rich, mutated in early stage of tumors*) est l'un des effecteurs aphidiens les mieux décrits à ce jour. Armet est une protéine animale ubiquitaire exploitée de façon vraisemblablement unique par les pucerons (N. Cui *et al.*, 2019; Wang, Dai, *et al.*, 2015). Chez *A. pisum*, la réduction de l'expression d'Armet via la méthode d'interférence par ARN entraîne une perturbation du comportement alimentaire sur *V. faba* et cette perturbation se reflète sur les profils EPG. La durée de la phase de sondage et celle d'ingestion passive **sont affectées** et la durée de vie des pucerons est réduite. Cette observation est cohérente avec les prévisions *in silico* de la fonction d'Armet comme protéine chélatrice des ions calciques. Les mécanismes moléculaires sous-jacents sont encore peu connus mais une étude récente suggère qu'Armet permettrait aux pucerons **d'exploiter à leur profit la diaphonie des voies de signalisation antagonistes** de l'acide jasmonique et de l'acide salicylique (N. Cui et al., 2019). Selon les résultats de cette étude, la suppression de la signalisation défensive basé sur l'acide jasmonique via la modulation de la voie de signalisation de l'acide salicylique par Armet

aurait une influence positive sur le comportement alimentaire et la colonisation d'*A. pisum* sur *N. benthamiana* et *M. truncatula* (N. Cui et al., 2019; Yates & Michel, 2018).

**ACE1/ACE2** (Angiotensin-converting enzymes 1 et 2) sont deux effecteurs aphidiens identifiés et partiellement caractérisés chez *A. pisum* sur *V. faba* (Wang, Luo, et al., 2015). Les ACE sont une famille de **métallopeptidases à zinc conservée à une échelle évolutive très importante**, chez les vertébrés (*e.g.* des mammifères) et invertébrés (*e.g.* des insectes), où elles assurent des rôles physiologiques clés et variés. Chez les insectes ainsi que chez les pucerons, des ACE ont été détectés dans des organes assurant **des rôles divergents** (*e.g.* les glandes salivaires, l'intestin moyen et l'hémolymphe). Chez *A. pisum*, des transcrits associés aux ACE sont exprimées **dans les glandes salivaires** et la présence de ces protéines dans la salive des pucerons a été démontrée (Carolan et al., 2009, 2011). ACE1 et ACE2 moduleraient des performances biologiques des pucerons par un effet **synergique**. La répression simultanée de l'expression d'ACE1 et ACE2, engendrée via la méthode RNAi, est potentiellement liée à un **comportement alimentaire plus efficace** mais semble exercer un **effet négatif sur la survie des pucerons**. D'un point de vue évolutif, l'interprétation de la coexistence d'un effet favorable (un gain d'efficacité du processus d'alimentation) et défavorable (une perte de fitness) lié à la répression simultanée de l'expression d'ACE1 et ACE2 est une tâche **complexe**. Du point de vue de la survie du puceron sur son hôte, ACE1 et ACE2 auraient un **profil de facteur de d'avirulence**.

**MIF1** (*Macrophage Migration Inhibitory Factor 1*) est une protéine cytokine salivaire identifiée chez *A. pisum* et *M. persicae* (Dubreuil et al., 2014; Naessens et al., 2015). Les MIF sont une famille de protéines **conservée à un rang taxinomique élevé** (*e.g.* chez les mammifères, les plantes, les bactéries et les arthropodes) et notamment connus pour être **des régulateurs clés de la réponse immunitaire** (Sparkes et al., 2017). Les MIF aphidiens interfèrent avec le système immunitaire des plantes et leur expression chez les pucerons est requise pour leur permettre d'exploiter leur hôte. Des analyses de complémentation fonctionnelle démontrent que la protéine MIF1 est très probablement **nécessaire et suffisante pour inhiber les mécanismes de défense** de *N. benthamiana* (Naessens et al., 2015). L'expression ectopique des gènes MIF de *M. persicae* dans les tissus foliaires de *N. benthamiana* inhibe plusieurs manifestations physiologiques largement associées à l'immunité des plantes, telles que l'expression de gènes associés à des mécanismes de défense, les **dépôts de callose** et les **réponses d'hypersensibilité**.

**Me47** est une *glutathion S-transférase* (ou *GST*) salivaire fonctionnelle spécifique des isothiocyanates, composés impliqués dans la défense contre les herbivores (Kettles & Kaloshian, 2016). Chez les pucerons, les GST sont **liés à la détoxification** de composés de défenses naturels des plantes

(e.g. les glucosinolates chez les crucifères), mais aussi des insecticides (Francis et al., 2005; Lokeshwari et al., 2016; J. S. Ramsey et al., 2010). L'effet de Me47 *in planta* a été examiné sur les performances (i.e. fécondité) de *M. persicae* et *M. euphorbiae*, respectivement sur *N. benthamiana* et *A. thaliana*, et *Solanum lycopersicum*. Cette analyse a révélé que la contribution de Me47 à la performance des pucerons était **spécifique des espèces d'hôtes testés** et pouvait, selon le système considéré, avoir un **impact positif ou délétère sur la fécondité** des pucerons (Kettles & Kaloshian, 2016). Du point de vue des performances des pucerons, Me47 peut avoir un **profil de facteur de virulence ou d'avirulence**, et ce profil dépend de la spécialisation de pucerons sur leur plantes hôtes.

**Me10** interagit spécifiquement avec la protéine **TFT7** de la tomate, et **améliore la virulence** de *M. euphorbiae* (Chaudhary et al., 2019). TFT7 est un isoforme de la **famille des protéines 14-3-3**, connue pour être ubiquitaires chez les eucaryotes et régulatrices de nombreux processus physiologiques dont la transduction de signaux intégrant des mécanismes de défense des plantes (Denison et al., 2011). La coexpression de l'homologue de Me10 chez *A. gossypii*, Ag10k, et de la protéine TFT7 de la tomate dans des feuilles de *N. benthamiana* en utilisant l'expression transitoire médiée par *Agrobacterium* a montré qu'Ag10k interagissait spécifiquement avec TFT7 (Chaudhary et al., 2019).

**PintO1** (ou Mp1) contribue à la **virulence** de *M. persicae* en **ciblant de façon spécifique** la protéine de transport vésiculaire **VPS52** (Vacuolar Protein Sorting-associated protein 52) chez *N. benthamiana*. L'infestation de *M. persicae* déclenche la réduction du niveau de VPS52. La surexpression de VPS52 dans le phloème a un **impact négatif sur la fécondité**. Ces deux résultats permettent de faire l'hypothèse Mp1 est un facteur de virulence pour les pucerons (Chaudhary et al., 2019; Rodriguez et al., 2017). Le fait que d'une part, l'interaction entre PintO1 et la protéine VPS52 de *N. benthamiana* soit à caractère spécifique, et que d'autre part la famille VPS52 présente un haut niveau de conservation chez les plantes suggère que la spécificité d'interaction entre PintO1 et VPS52 est le résultat des changements structurels mineurs (Rodriguez et al., 2017).

**Mp10**, identifié chez *M. persicae*, présente **une double activité**, pouvant entraîner à la fois le **déclenchement et la suppression des défenses des plantes** (Bos et al., 2010). Lorsqu'il est surexprimé transitoirement dans le cytoplasme chez *N. benthamiana*, Mp10 supprime la production de ROS induite par le MAMP bactérien flg22, suggérant un rôle de **suppresseur de la PTI** (Bos et al., 2010). En revanche, cet effecteur montre également un rôle dans l'activation des réponses de défense des plantes ainsi qu'un **impact négatif sur la fécondité** des pucerons dans les tests de surexpression (Bos et al., 2010; Rodriguez et al., 2014). De manière intéressante, Mp10 est le premier effecteur aphidien

pour lequel un lien direct a été établi **avec l'acrostyle** (Deshoux et al., 2022), un organe cuticulaire situé à l'extrémité des stylets maxillaires (Uzest et al., 2010), et initialement connu pour son rôle dans la rétention transitoire de phytovirus (Webster et al., 2018). L'interaction de Mp10 à l'acrostyle, consistant en une liaison spécifique avec la protéine cuticulaire Stylin-03, a été démontrée chez *M. persicae* et *A. pisum* dans une approche originale combinant microscopie à épifluorescence, interférence par ARN et interaction dans la levure (Deshoux et al., 2022). Il serait intéressant d'étudier s'il existe un lien entre la double activité de Mp10 (*i.e.* déclenchement et la suppression des défenses des plantes), et sa capacité à être retenu par l'acrostyle, et donc, non injecté dans la plante.

### 3.3.3. Quelques caractéristiques notables des effecteurs aphidiens

Les entreprises visant à caractériser la nature et la fonction des effecteurs aphidiens sont relativement récentes par rapport à celles portant sur les effecteurs des agents pathogènes. En dépit d'un intérêt grandissant par la communauté scientifique, les connaissances accumulées restent insuffisantes pour déduire des règles générales sur le fonctionnement des effecteurs aphidiens et établir des profils précis. Néanmoins, plusieurs contributions scientifiques décisives permettent de dégager quelques caractéristiques notables pour les effecteurs aphidiens et ainsi proposer un premier panorama de leur fonctionnement.

J'ai pu dégager **huit caractéristiques notables** pour les effecteurs aphidiens à partir d'un état de l'art limité à une dizaine d'effecteurs relativement connus : **(i)** certains effecteurs ont été clairement identifiés pour leur rôle avirulent, comme ACE1/ACE2 (Wang, Luo, et al., 2015), ou virulent, comme C002 et Mp1 (Chaudhary et al., 2019; Mutti et al., 2008), d'autres semblent cumuler les deux rôles (Bos et al., 2010; Kettles & Kaloshian, 2016; Wang, Luo, et al., 2015), **(ii)** la moitié des effecteurs connus ne présentent aucune similarité avec des protéines connues et sont spécifiques des pucerons comme C002, Mp55 et PintO1 (Mutti et al., 2008) alors que d'autres sont conservés à des échelles évolutives très importantes comme ACE1, ACE2, Armet et MIF1 (Naessens et al., 2015; Wang, Dai, et al., 2015), et connues pour assurer d'autres fonctions, comme les ACE (Kettles & Kaloshian, 2016; Wang, Luo, et al., 2015), **(iii)** à l'exception de l'ARN non codant *Ya1*, tous les autres effecteurs identifiés sont des protéines (Y. Chen et al., 2020), **(iv)** si le mode d'action des effecteurs aphidiens est présumé cytoplasmique, car attendus comme composants de la salive aqueuse, la suggestion d'un mode d'action extracellulaire dans les cas de MpPIntO1 et MpC002 semble indiquer que, contrairement à la représentation de référence, des effecteurs aphidiens sont également présents dans le flux de salive gélifiante (Mugford et al., 2016; Will & Vilcinskas, 2015), **(v)** l'activité de certains effecteurs, tels que Mp1, Mp2 et Me47, est spécifique du couple plante-puceron (Bos et al., 2010; Kettles & Kaloshian, 2016; Pitino & Hogenhout, 2013), **(vi)** pour moduler la physiologie des plantes, certains effecteurs

aphidiens comme ACE1 et ACE2, agissent de concert alors que d'autres ont vraisemblablement une action individuelle suffisante (Wang, Luo, et al., 2015), **(vii)** les effecteurs aphidiens caractérisés ciblent des processus physiologiques généraux et variés qui sont impliqués dans les défenses des plantes : transport vésiculaire, chélation des ions calcium, occlusion des tubes criblés du phloème blessés par les stylets des pucerons (Chaudhary et al., 2019; Rodriguez et al., 2017; Will et al., 2007) ou signalisation. A ce jour, PintO1 et Me10 sont les seuls effecteurs aphidiens dont la molécule cible chez la plante est connue.

### 3.3.4. Les effecteurs aphidiens face au paradigme de référence

Conformément au paradigme de référence pour les interactions plantes-bioagresseurs, il est attendu, pour qu'une réponse immunitaire ait lieu (ETI), qu'un **effecteur d'avirulence aphidien** introduit dans la plante hôte via le flux salivaire soit reconnu par une **protéine de résistance NLR**. Cependant, aucun effecteur aphidien actuellement étudié n'est connu pour interagir avec une protéine NLR. Du côté des plantes, plusieurs résistances aux pucerons héréditaires ont été décrites chez des espèces cultivées (Catherine Dogimont et al., 2010), mais peu ont été **caractérisées au niveau moléculaire** : **(i)** le gène **NLR *Mi1-2*** confère à la tomate une résistance au puceron de la pomme de terre *M. euphorbiae* ainsi qu'à d'autres ravageurs (Q. Li et al., 2006; Rossi et al., 1998), **(ii)** le gène **NLR *Vat-1*** confère au melon une résistance au puceron du melon et du coton, *Aphis gossypii*, ainsi qu'aux phytovirus lorsqu'il est le vecteur (Catherine Dogimont et al., 2014), **(iii)** le gène ***SL11*** confère à *A. thaliana* une résistance au puceron du pêcher *Myzus persicae* (Kloth et al., 2017; Kloth & Kormelink, 2020) et le gène **NLR *Adnr1*** confère au blé la résistance au puceron du blé *Diuraphis noxia* (Nicolis & Venter, 2018). Au total, trois de ces gènes de résistance, *Mi1-2*, *Vat-1* et *Adnr1*, codent pour des protéines NLR. Toutefois, **chez les pucerons**, *i.e.* l'autre partenaire de l'interaction, **aucun gène pour les effecteurs interagissant avec les gènes NLR impliqués dans le cadre ETI n'a été identifié jusqu'à présent**. L'extension de mes recherches à l'échelle des hémiptères (*i.e.* les cicadelles, les aleurodes et les punaises) ne m'a pas permis de relever de telles interactions pour des insectes piqueurs-suceurs.



## 4. La résistance *Vat* du melon : une résistance à *A. gossypii* et aux virus qu'il transmet

La première observation d'une résistance à *A. gossypii* remonte au milieu du siècle dernier en Amérique (Ivanoff, 1944). Quelques décennies plus tard, à la fin des années 70, une équipe française de virologistes et de généticiens des plantes mirent en évidence la coségrégation de ce phénotype, avec un trait d'intérêt agronomique notable : une **résistance complète aux trois virus** : CMV, ZYMV et WMV **lorsque qu'ils sont inoculés par *A. gossypii***. (H. Lecoq et al., 1980; Pitrat & Lecoq, 1980). Ce double phénotype, singulier, est contrôlé par un seul gène dominant, baptisé ***Vat***, pour « **Virus Aphid Triggered** » (Boissot, Schoeny, et al., 2016). Le phénotype *Vat* a été identifié à partir de différentes ressources génétiques provenant de régions géographiques variées et introgressé dans de nombreuses lignées commerciales (Boissot, Thomas, et al., 2016; C Dogimont et al., 2008; E. Garzo et al., 2002; Kishaba et al., 1971; Pitrat & Lecoq, 1980; Thomas, Dogimont, et al., 2012). Le locus *Vat* a été séquencé au début des années 2000 (Catherine Dogimont et al., 2014). Il est situé dans une **région génomique remarquable** qui contient la plus grande concentration de gènes de type NLR chez le melon (González et al., 2014). Outre la résistance élicitée par les pucerons, plusieurs résistances aux agents pathogènes ont été localisées dans la zone: des résistances à l'oïdium, la résistance au *Cucumber vein yellowing virus*, le gène *Fn*, gène d'incompatibilité qui entraîne une nécrose et la mort de la plante avec certaines souches de ZYMV et aussi *FomV-2*, QTL à effet faible de résistance à *Fusarium oxysporum* f. sp. *melonis*. **Dans mes travaux de thèse, j'utilise exclusivement l'allèle *Vat* « historique », noté *Vat-1*, qui a été isolé à partir de l'accession de melon coréenne PI 161375 et avec un groupe de clones d'*A. gossypii* montrant des comportements contrastés vis-à-vis de cette résistance** (Boissot, Thomas, et al., 2016; Catherine Dogimont et al., 2014).

### 4.1. Physiologie associée à *Vat* et conséquences pour les pucerons et les virus

Les modalités du comportement alimentaire d'*A. gossypii* via la technique d'EPG (Electropénétrographie) sont renseignées par un cortège d'études conséquent (Jian Q. Chen et al., 1997; Jian Qun Chen et al., 1997; E. Garzo et al., 2002; Klingler et al., 1997). Ces études mobilisent des **clones originaires de régions géographiques variées et différentes lignées *Vat* avec des fonds génétiques variés**, et au total, permettent d'établir un panorama précis et cohérent des effets de *Vat* sur les différentes phases d'alimentations d'*A. gossypii*. L'ensemble de ces études démontre que le processus d'alimentation des pucerons **est perturbé sur plantes *Vat* : (i) le rapport de temps entre les phases de sondage et de salivation est déséquilibré, (ii) la fréquence des événements où les stylets**

atteignent le phloème connaît une diminution remarquable, **(iii)** dans les quelques cas où les stylets parviennent à atteindre le phloème, la phase d'ingestion n'est pas durable **(iv)**, les piqûres d'essai sont plus nombreuses et reflètent une multiplication de tentatives d'alimentations sans succès.

En complément de ces données d'EPG, des observations à l'échelle cellulaire montrent que la résistance *Vat* est associée à une variété de réponses physiologiques caractéristiques de l'ETI et qui se manifestent à des échelles de temps et des niveaux d'observation différents (Boissot, Schoeny, et al., 2016). Dans un délai très court (de l'ordre de la minute), **une réponse d'hypersensibilité (HR)**, locale, se manifeste à proximité de la zone blessée par les stylets des pucerons. Dans les 10 minutes suivant l'infestation, un pic de **production d'espèces réactives de l'oxygène** et une **dégradation de la membrane cytoplasmique** des cellules proches de la zone blessée ont lieu. Dans le cas d'une transmission de virus lors des piqûres de sondage (*i.e.* non-persistante), la charge virale est entièrement expulsée dès les premières sécrétions salivaires lors du sondage des couches cellulaires en surface (Pirone & Perry, 2002). Le blocage et la destruction rapide de la charge ainsi transmise est aspécifique et principalement attribué à la réponse d'hypersensibilité, **déclenchée par la reconnaissance d'un facteur aphidien**. Cette réponse explique la résistance au CMV, ZYMV, WMV quand ils sont transmis par *A. gossypii*. Respectivement dans un délai de 20 minutes et de 4,5 heures suivant l'infestation, des **dépôts de callose et de lignine** ont lieu dans les parois des cellules sondées par les stylets et probablement dans les cellules adjacentes non blessées (Villada et al., 2009). Des observations récentes suggèrent que la réponse immunitaire liée au déclenchement de *Vat* se manifeste également **au niveau du phloème** et implique la coagulation et l'agrégation de **protéines d'occlusion** ainsi un dépôt **de callose** aient lieu au niveau des **assiettes criblées** (Peng & Walker, 2018). Cette obstruction de l'espace intercellulaire, voire phloémien semble être à la cause des difficultés d'alimentations reflétées par les profils d'EPG (Peng & Walker, 2018; Villada et al., 2009). Cette réponse explique la difficulté pour *A. gossypii* à atteindre le phloème, et entraîne probablement une réduction de la transmission du CABYV, qui nécessite une connexion stylet-phloème durable (Hervé Lecoq & Desbiez, 2012; Schoeny et al., 2017).

Actuellement, une seule étude associe la résistance *Vat* à un **dépôt de callose** au niveau du phloème et suggèrent qu'elle se manifeste uniquement dans le FP (Peng & Walker, 2018). Le phloème spécifique des cucurbitacées, ou EFP, n'est **apparemment pas concerné** par cette réponse physiologique. Toutefois, aucun travail de recherche ayant eu recours à la technique d'EPG n'a pris en compte la spécificité du phloème des cucurbitacées. L'éventualité d'une spécificité des profils d'EPG en association avec chacun des deux phloèmes des cucurbitacées a été évoquée mais n'a jamais été investiguée (Kanvil et al., 2017). Selon moi, le fait que la composition et les dimensions des tubes criblés

sont comparables chez ces deux phloèmes me semble remettre en cause, au moins en partie, de la pertinence d'un tel projet.

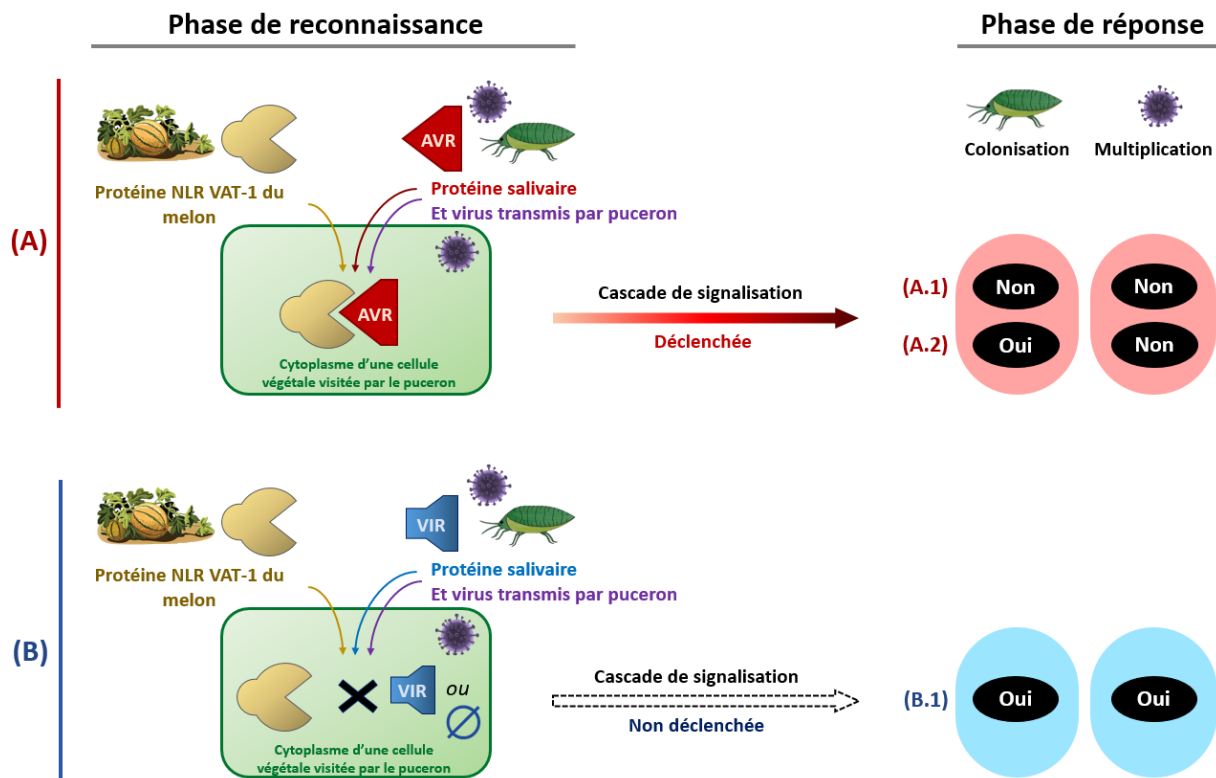
## 4.2. Un spectre de résistance en désaccord avec le modèle de référence

La protéine présumée issue de l'expression de *Vat-1* est **prédite cytoplasmique** et comprend des **domaines CC-NBS-LRR** (coiled-coil, nucleotide-binding site, leucine-rich-repeat) (Catherine Dogimont et al., 2014). Cette catégorie de gènes, maintenant appelé NLR, est largement associée aux mécanismes de défenses contre des agents pathogènes ainsi qu'aux hémiptères (Kaloshian & Walling, 2005, 2016b; Kapos et al., 2019; Rafiqi et al., 2009). La **reconnaissance spécifique** par une protéine NLR d'un motif moléculaire **caractéristique d'une invasion** induit un changement dans la conformation de la protéine NLR et cette modification structurelle est interprétée comme un signal d'activation d'une variété de mécanismes de défense. Lors de la reconnaissance d'un effecteur par une protéine NLR, **le module CC** serait impliqué dans la signalisation aval, **le module NB** déclencherait la mise en place de la conformation compétente de la protéine NLR pour la transmission d'un signal physiologique et **le module LRR** serait le déterminant majeur de la spécificité de reconnaissance et l'initiateur de la reconnaissance (Collier & Moffett, 2009; Maekawa et al., 2011; Rafiqi et al., 2009).

En accord avec le cadre conceptuel établi pour cette catégorie de gènes, le fonctionnement de *Vat* implique un découplage de la phase de reconnaissance de l'invasion et celle de la réponse immunitaire consécutive. Lors de la **phase de reconnaissance**, la protéine cytoplasmique VAT doit reconnaître un éliciteur d'origine aphidienne qui a été injecté via les stylets d'*A. gossypii* à l'intérieur des cellules de l'épiderme ou du mésophylle. La **réponse immunitaire** est déclenchée au niveau de la cellule piquée bloquant la réplication et le mouvement des virus éventuellement transmis avec le flux salivaire et inhibe la colonisation des pucerons. Selon la théorie des interactions gène pour gène de Flor, l'adaptation d'un agent pathogène à une résistance conférée par un gène de plante est due un changement ou à la disparition du gène d'avorulence (Flor, 1971). En accord avec ce modèle général, un clone adapté serait capable de coloniser massivement une plante *Vat* et ne déclencherait pas la résistance au virus qu'il a transmis. Inversement, un clone non adapté serait incapable de coloniser une plante *Vat* et déclencherait la résistance au virus qu'il a transmis.

La **valeur explicative du modèle proposé par Jones et Dangl [figure 4]** pour le double phénotype conféré par *Vat* a été testée sur des lignées transgéniques *Vat-1* infestées par une gamme de six clones d'*A. gossypii* (C6, C9, CUC1, GWD, GWD2, et NM1) ayant préalablement acquis le CMV

(Boissot, Thomas, et al., 2016). A l'exception du phénotype observé pour GWD2, les phénotypes observés pour tous les autres clones étaient cohérents avec ce modèle : ces clones étaient soit incapable de coloniser la lignée transgénique *Vat-1* et déclenchaient la résistance au CMV [figure 7 A.1], soit capables de coloniser massivement la lignée transgénique *Vat-1* et ne déclenchaient pas la résistance au CMV [figure 7 B.1]. **Remarquablement**, le clone GWD2 était capable de coloniser la lignée transgénique *Vat-1* alors qu'il déclenchait résistance au CMV [figure 7 A.2]. Cette **déviatiion par rapport au modèle général** d'adaptation par perte de reconnaissance, où un clone d'*A. gossypii* qui élicite la résistance aux virus est capable de coloniser une plante *Vat*, révèle l'existence d'un mécanisme adaptatif additionnel chez *A. gossypii* (Boissot, Schoeny, et al., 2016; Boissot, Thomas, et al., 2016). Plus généralement, ce cas de figure, en désaccord avec le modèle de Jones et Dangl, a été relevé pour un ensemble de clones en présence d'autres allèles/homologues de *Vat-1*. Sur 74 interactions lignées de melon/clone de puceron où la résistance au CMV était déclenchée par les pucerons, les deux tiers montraient une forte colonisation par *A. gossypii* et ne sont pas explicables avec le modèle de Jones et Dangl (Boissot, Thomas, et al., 2016). Cette expérimentation a montré que l'adaptation des pucerons à l'ETI par un polymorphisme dans un effecteur salivaire n'est **pas la norme** et a permis d'établir un modèle pour les interactions *Vat*-puceron [figure 7] (Boissot, Thomas, et al., 2016). Aussi, la résistance aux virus déclenchée par les piqures d'*A. gossypii* sur plante *Vat* n'a pas pu être contournée expérimentalement (Boissot, Thomas, et al., 2016).



**Figure 7** – Modèle des interactions entre une lignée de melon transgénique *Vat-1* et différents biotypes d' *A. gossypii* montrant des comportements contrastés vis-à-vis de cette résistance

La **phase de reconnaissance** est représentée en accord avec la théorie des interactions gène pour gène établie par Flor (1942) et dont l'application est largement admise pour les gènes NLR. La protéine VAT-1 du melon reconnaît spécifiquement (de manière directe ou indirecte) l'isoforme avirulent d'un effecteur salivaire (AVR) présent chez certains clones d' *A. gossypii* et déclenche une cascade de signalisation élicitant les défenses de la plante. Les clones *A. gossypii* porteurs de l'isoforme virulent de cet effecteur (VIR) ou ayant perdu cet effecteur (ensemble vide) ne sont pas reconnus par la protéine VAT-1 et les défenses de la plante ne sont pas élicitées. La **phase de réponse** illustre les différents phénotypes *Vat* observés en conditions expérimentales contrôlées suite à des infestations par un panel de six clones *A. gossypii* porteurs du CMV (Boissot et al, 2016a). Les phénotypes « colonisation par pucerons » et « multiplication du virus » sont notés : « Non » pour absence, « Oui » pour présence. Différents cas ont été observés : **(A.1)** résistance aux pucerons et au CMV, **(A.2)** sensibilité aux pucerons et résistance au CMV et **(B.1)** sensibilité aux pucerons et au CMV. *Adaptation de Boissot et al (2016a).*

## 4.3. Effets de *Vat* à l'échelle des systèmes agricoles

### 4.3.1. Effets de *Vat* sur les populations de pucerons et sa durabilité

Le gène *Vat-1* appartient à la famille des **gènes de résistance NLR**, famille connue pour exercer une **pression de sélection** notable sur la diversité naturelle des agents pathogènes présents et favoriser l'**émergence de biotypes résistants**. Du fait de la sévérité de la pression sélective exercée mais aussi de la prolongation de la présence aux champs dans de nombreux systèmes agronomiques, les gènes NLR ont une **durée d'efficacité très variable**, allant de quelques années à plusieurs décennies (Ayliffe et al., 2008). En raison de la complexité des agroécosystèmes, il est **difficile de prédire** la durée d'efficacité agronomique des gènes de type NLR.

**A l'échelle de la plante**, la résistance *Vat* réduit fortement l'acceptation et la colonisation des lignées de melon par certains clones d'*A. gossypii*. **A l'échelle de la parcelle**, la résistance *Vat-1* décroît la densité des populations aphidiennes et modifie la structure génétique ainsi que la composition de ces populations (Thomas et al., 2016).

**A l'échelle des bassins de production de melons**, où seule la forme *Vat-1* a été largement déployée (Boissot, Schoeny, et al., 2016), l'évolution de la composition des populations d'*A. gossypii* reste **difficile à anticiper** de par la complexité inhérente aux agroécosystèmes mais aussi en raison de la rareté des travaux de recherche à ce sujet. Hormis le déploiement de la résistance *Vat* au champ, **une variété de paramètres** biotiques et abiotiques influencent la composition de ces populations à l'échelle locale.

**A l'échelle annuelle**, la proportion de l'espèce *A. gossypii* dans le flux printanier total des populations aphidiennes migrantes ainsi que la composition génétique dépend des caractéristiques géographiques (Thomas et al., 2016). La **richesse clonale** des populations colonisatrices est **significativement réduite à deux niveaux** : **(i)** sur les cultures de melon, **(ii)** et encore davantage sur melon porteur du gène *Vat-1*. A l'échelle pluriannuelle, les populations d'*A. gossypii* évoluent à l'échelle régionale en réponse au déploiement de variétés de melons *Vat-1* (Thomas et al., 2016). Des études de terrain ont déterminé **deux goulets d'étranglement** pouvant affecter la dynamique des clones adaptés au cluster *Vat* : **(i)** une faible production des formes dispersantes sur melon *Vat-1* et **(ii)**, l'extinction hivernale.

La durabilité de la résistance *Vat-1* aux populations d'*A. gossypii* est multifactorielle et dépend fortement des caractéristiques des agroécosystèmes en place et tout particulièrement de la présence d'autres cucurbitacées pouvant servir d'hôtes alternatifs pour les clones adaptés, en automne et en hiver (Thomas et al., 2016). Dans certains cas, l'émergence rapide de biotypes virulents dans des agroécosystèmes est liée à une diversité clonale préexistante.

#### 4.3.2. Effet de *Vat* sur les virus transmis par pucerons et sa durabilité

L'effet de *Vat-1* sur les épidémies de virus transmis par pucerons est relativement peu documenté. Les quelques connaissances disponibles proviennent essentiellement d'expériences **en laboratoire**. Le potentiel du gène *Vat-1* pour contrôler les épidémies virales **dépend significativement de la composition des populations aphidiennes**. D'autres espèces aphidiennes récurrentes dans de nombreux agroécosystèmes et n'élicitant pas la résistance *Vat-1*, incluant *M. euphorbiae*, *M. persicae*, *A. craccivora* et *A. fabae*, sont **des vecteurs alternatifs** pour le CMV, le WMV, le ZYMV et le CABYV. L'apparition et la propagation des épidémies virales sur les lignées *Vat-1* s'explique en grande partie par la **prédominance relative d'*A. gossypii*** comme vecteur.

Cependant, *Vat-1* a un **impact significatif** sur les **épidémies de CABYV** (Schoeny et al., 2017). Le succès de *Vat-1* pour maîtriser la propagation de ce virus est dû au fait qu'*A. gossypii* est son vecteur prédominant mais aussi que sa transmission nécessite **un contact prolongé avec le phloème** alors que *Vat-1* en restreint fortement l'accès. *Vat-1* a aussi un **effet partiel** mais toutefois notable sur les **épidémies de CMV**, vraisemblablement dû au fait qu'il soit l'un des vecteurs principaux, l'effet variable serait dû à l'importance relative d'*A. gossypii* (Schoeny et al., 2017). Les expériences en plein champ montrent que la structure des populations de virus n'est **pas modifiée** en présence de *Vat-1*, ce qui laisse penser que son effet est **durable**, comme le suggérait les études au laboratoire. En effet, les **expériences d'évolution en laboratoire** ont montré que les virus CMV, WMV, ZYMV **ne pouvaient pas évoluer pour contourner la résistance *Vat*** élicitée par *A. gossypii* (Boissot, Thomas, et al., 2016). Si l'effet de *Vat* ne permet pas de contrôler totalement les épidémies de virus transmis par pucerons dans les cultures, le **large spectre** de cet effet et l'incapacité des virus à s'y adapter ont fait de ce type de résistance un **atout majeur pour les sélectionneurs** de melons (Boissot, Thomas, et al., 2016).

## 5. Ma thèse, son contexte et ses objectifs

L'intérêt des entreprises semencières pour la résistance *Vat* s'est concrétisé par plusieurs programmes de sélection variétale. Toutes les variétés de melon inscrites au catalogue français de 2022 sont déclarées résistantes aux pucerons et la notoriété des créateurs de variétés repose sur le fait que la résistance *Vat* soit efficace et pérenne. Ma thèse CIFRE a été cofinancée par deux entreprises internationales spécialisées dans la sélection de semences potagères : **ENZA ZADEN** et **RIJK ZWANN**.

**Ma thèse est orientée vers la production de connaissances et représente une opportunité d'étudier : (i)** l'adaptation des clones de pucerons appartenant à la race d'hôte des cucurbitacées, **(ii)** l'adaptation des clones de cette race d'hôte à une résistance génétique chez le melon et mais aussi de manière plus générale **(iii)**, d'étudier le phénotype *Vat* comme clé de lecture des mécanismes de la résistance héritable. **Mes travaux traitent de la diversité potentielle des voies d'adaptation d'*A. gossypii* aux Cucurbitacées et à *Vat*.**

**Dans un premier temps**, je propose un dérivatif au paradigme des interactions plante-bioagresseur où l'adaptation des pucerons ne passe pas par une modification du gène d'avirulence interagissant avec *Vat*. J'é mets ici l'hypothèse que les clones adaptés à la race d'hôte des cucurbitacées ainsi que ceux qui contournent la résistance *Vat* sont capables d'exploiter le phloème particulier des cucurbitacées. Dans cette partie, j'envisage le double phénotype contrôlé par *Vat-1* comme un outil pour la recherche du phloème source de clones montrant des comportements contrastés vis-à-vis de cette résistance.

**Dans un second temps**, je considère la voie d'adaptation canonique des agents pathogènes et bioagresseurs de plantes selon laquelle il est probable que les gènes codants pour les protéines reconnues par un gène NLR soient soumis à une sélection diversifiante (Flor, 1971; Jones & Dangl, 2006; Rouxel & Balesdent, 2010). J'é mets ici l'hypothèse qu'un effecteur aphidien interagit avec la protéine NBS-LRR VAT-1 du melon, que cet effecteur a une hérédité simple, et que son polymorphisme explique le passage de l'état avirulent à virulent.





**Chapitre 1 (article) - Recherche du phloème  
source pour *A. gossypii***

# 1. Contexte de l'étude

Ce chapitre est sous la forme d'un **article soumis** en **Décembre 2022**, chez **BMC Plant Biology**. Cet article traite d'**une voie d'adaptation originale** pour des clones d'*A. gossypii* spécialisés sur les cucurbitacées ne faisant pas intervenir la modification du gène d'avirulence interagissant avec *Vat*. Nous avons testé l'hypothèse que la **structure originale du phloème** intervient dans le **contournement de la résistance *Vat-1*** par **certains clones adaptés**, qui seraient alors capables d'exploiter le phloème particulier des cucurbitacées. L'objectif était d'identifier le **phloème source** pour trois clones d'*A. gossypii*, montrant des comportements contrastés vis-à-vis de la résistance *Vat-1*, à partir de **métabolites fortement associés à l'un des deux phloèmes** du melon qu'ils ont **ingérés**. Cet effort de recherche est le premier à prendre en compte simultanément : **(i)** la résistance NLR VAT-1, **(ii)** l'existence de clones contournant et **(iii)**, la complexité structurale du phloème du melon. Les designs expérimentaux et leur mise en œuvre pour l'identification et la quantification des métabolites phloémiens, ingérés par des pucerons ont été réalisés **en collaboration avec la plateforme métabolomique 'Métaboscope'** sur le site de l'université d'Avignon. Les **tables additionnelles** de l'article sont présentées dans les **annexes référencées à la fin du manuscrit**.

## 2. Aphid adaptation to cucurbits: sugars, cucurbitacin and phloem structure in resistant and susceptible melons

Pierre Sadon<sup>1</sup>, Marie-Noëlle Corre<sup>1</sup>, Raphael Lugan<sup>2</sup> and Nathalie Boissot<sup>1</sup>

1 National Institute for Agriculture, Food and Environment, INRAE, Génétique et Amélioration des Fruits et Légumes, Domaine St-Maurice, 84143, Montfavet Cedex, France

2 National Institute for Agriculture, Food and Environment, INRAE, Plantes et Systèmes de cultures Horticoles, Domaine St Paul 84914, Avignon Cedex, France

Corresponding author: [nathalie.boissot@inrae.fr](mailto:nathalie.boissot@inrae.fr)

### Abstract

#### Background

*Aphis gossypii*, a strictly phloemophagous aphid, colonize hundreds of plant families, and a group of clones formed a cucurbit-specialised host-race. Cucurbits are unique in having evolved a specific extra-fascicular phloem (EFP), which carries defence-related metabolites such as cucurbitacin, whereas the fascicular phloem (FP) is common to all higher plants and carries primary metabolites, such as raffinose-family oligosaccharides (RFOs). Both cucurbitacins (in the EFP) and galactinol (in the FP) have been suggested to be toxic to aphids. We investigated these hypotheses in cucurbit-specialized *A. gossypii* fed on melon plants with or without aphid-resistance conferred by the NLR gene *Vat*. We selected a plant-aphid system with (i) *Vat*-mediated resistance not triggered, (ii) *Vat*-mediated resistance triggered by an aphid clone adapted to the presence of *Vat* resistant alleles and (iii) *Vat*-mediated resistance triggered by a non-adapted aphid clone.

#### Results

We quantified cucurbitacin B, its glycosylated derivative, and sugars, in melon plants and aphids that fed on. The level of cucurbitacin in plants was unrelated to both aphid infestation and aphid resistance. Galactinol was present at higher quantities in plants when *Vat*-mediated resistance was triggered, but its presence did not correlate with aphid performance. Finally, we showed that cucurbit-specialized *A. gossypii* fed from the FP but could also occasionally access the EFP without sustainably feeding from

it. However, the clone not adapted to *Vat*-mediated resistance were less able to access the FP when the *Vat* resistance was triggered.

### **Conclusion**

We concluded that galactinol accumulation in resistant plants does not affect aphids, but may play a role in aphid adaptation to fasting and that Cucurbitacin *in planta* is not a real threat to *Aphis gossypii*. Moreover, the specific phloem of Cucurbits is involved neither in *A. gossypii* cucurbit specialisation nor in adaptation to *Vat*-dependent resistance.

## **Keywords**

*NLR*, plant resistance, specialized metabolites, primary metabolites, *Vat* melon.

## Background

Vascular plants colonized land about 400 million years ago and 300 million years later aphids evolved specialized mouthparts called stylets, which could penetrate plant tissues to establish an intricate relationship with this vascular system that is their feeding source (Powell et al., 2006). About 4000 aphid species are known, including the melon-cotton aphid, *Aphis gossypii*, which diverged about 5 million years ago (<http://timetree.org>). *A. gossypii* is an extremely polyphagous pest with populations structured into host-specific races feeding and reproducing on different crops (Boissot, Schoeny, et al., 2016; Carletto et al., 2009), including a specialist aphid for cucurbitaceous species (Mistral et al., 2021). Cucurbits, which diverged 60 million years ago (<http://timetree.org>) earlier than *A. gossypii*, exhibit an unusual vascular system which is split into two physically and functionally distinct phloems (B. Zhang et al., 2010). The fascicular phloem (FP), composed of bundles of sieve tubes and companion cells on both sides of the xylem (bicollateral), is mainly involved in the transport of photoassimilates. The extrafascicular phloem (EFP) is composed of an array of phloem elements located outside the bundles and is mainly involved in the transport of various metabolites, including defence-related compounds (B. Zhang et al., 2010). The FP is common to higher plants whereas the EFP is only found in the Cucurbitaceae family (R Turgeon, 2016). Different EFP configurations exist in Cucurbitaceae, depending on the species and its phylogeny: *Alsomitra*, for instance, displays only collateral (rather than bicollateral) bundles, while *Cucurbita* has the most complex phloem type with a highly developed extrafascicular phloem, featuring all types of extrafascicular elements (R Turgeon, 2016). Knowledge on the interaction between aphids and the cucurbit vascular system is quite sparse. Electrical-Penetration Graph (EPG) recordings brought direct evidence that aphids feed from the melon vascular system (Jian Qun Chen et al., 1997; Klingler et al., 1997), but no EPG waveform has been able to discriminate between FP or EFP feeding aphids thus far. Despite the suggestion that the EFP might contain a dietary composition more suited to aphids due to its relatively low sugar/amino acid content (Fiehn, 2003, Mattson, 1980), the comparison of sugar profiles from *Myzus persicae* stylet exudates with sieve exudates from FP and EFP indicated that *M. persicae* fed exclusively from FP on *Cucurbita pepo* (Kanvil et al., 2017). Moreover, the location of stylet track endings, as shown using microscopy in *Cucurbita pepo*, suggested that *A. gossypii* fed on FP, and since feeding on artificial diets supplemented with EFP exudates had a negative impact on their growth rate after 72 hours, aphids have been suggested to show EFP avoidance behaviour (Kanvil et al., 2017).

Melon is of particular interest for the plant-aphid interaction given that it is the only cucurbit for which a high level of resistance to *A. gossypii* has been detected. This resistance is mediated by the Nucleotide-binding domain - Leucine-rich Repeat (NLR) gene named *Vat* (Catherine Dogimont et al., 2014). EPG profiles in tissues showed that aphid feeding behaviour is impaired on *Vat*-melon plants,

i.e. plants carrying a resistant allele in the *Vat* gene cluster (Chovelon et al. 2021): aphids reach the phloem less often and feed for a shorter time on *Vat*-plants compared to susceptible plants (Peng & Walker, 2018; Villada et al., 2009). Callose and lignin deposition in the walls of cells adjacent to the stylet path, and putatively also on the sieve-plates in the phloem, occurs faster in *Vat*-melon plants (Elisa Garzo et al., 2018; Peng & Walker, 2018; Villada et al., 2009). Such phloem obstruction was also observed after mechanical wounding in both FP and EFP, but later and to a lesser extent for the EFP (Crafts, 1932; Peng & Walker, 2018; Walker, 2021; B. Zhang et al., 2010). Yet, some clones are adapted to *Vat*-mediated resistance: they can survive, reproduce and cause injury to *Vat*-melon plants (Boissot, Thomas, et al., 2016). Insight into the spectrum of resistance conferred by *Vat* suggests that *A. gossypii* adaptation to *Vat*-mediated resistance occurs by two different mechanisms: (i) following the effector triggered immunity model for NLR resistance proposed by Jones and Dangl (Dangl et al., 2013) whereby an alteration in the avirulence salivary component results in loss of recognition by the VAT NLR protein, preventing the triggering of resistance and (ii) in the absence of alteration to the avirulence salivary component, aphids cope with *Vat* resistance through unknown mechanisms (Boissot, Thomas, et al., 2016).

The timeline of aphid and cucurbit evolutionary history is consistent with the putative co-evolution of the cucurbit-specialised *A. gossypii* host-race feeding from the EFP and adapting to the defence-related compounds present. In the present study, we investigated whether the cucurbit-specialised *A. gossypii* host-race and the *Vat*-adapted clones could use the EFP as an alternative feeding source on susceptible or resistant melon plants. We selected a small set of metabolites typically found in FP and EFP: for FP we chose raffinose-family oligosaccharides (RFOs) (Sengupta et al., 2015; B. Zhang et al., 2010) and trehalose, with galactinol being potentially toxic to aphids (T. Cao, 2010; T. Cao et al., 2013). For EFP, we selected cucurbitacin as being potentially toxic to aphids (Yousaf et al., 2018; Zhao et al., 2021). We developed a GC-MS method to analyse RFOs and trehalose in plants and aphids, and an LC-MS method to analyse cucurbitacin B (Cb) (Gaupels & Ghirardo, 2013; B. Zhang et al., 2010). We looked at the abundance and variability of these metabolites in a variety of melon lines. We employed two sources of *Vat* resistance: one from the Korean melon line PI 161375 and the other from the Zimbabwean line PI 482398, as well as two pairs of lines which differ for the presence of the *Vat* allele derived from the resistant lines: in the first case, the *Vat* gene from PI 161375 was introduced by transgenesis into a Védrentais background, and in the second, the *Vat* resistance from PI 482398 was introduced into a Charentais T background by backcrossing. We named these two pairs of lines Védrentais/[Védrentais]<sup>R-PI161375</sup> and Charentais T/[Charentais T]<sup>R-PI482398</sup>. The metabolic markers were screened in three plant-aphid interaction systems using the pair of isogenic melon lines Védrentais/[Védrentais]<sup>R-PI161375</sup> and three aphid clones: NM1, C6 and GWD2 with the following response on [Védrentais]<sup>R-PI161375</sup>: (i) NM1 triggers *Vat*-mediated resistance and is not adapted to that

resistance, (ii) GWD2 triggers *Vat*-mediated resistance and is adapted to that resistance and (iii), C6 which does not trigger the *Vat*-mediated resistance [Fig. 1] (Boissot, Thomas, et al., 2016).

## Results

### Abundance and variability of phloem-associated metabolites in six melon genotypes

We quantified seven phloem-associated metabolites in the parental lines and the pairs of isogenic and near-isogenic lines [Fig. 2 and Sup\_Tab. 1]. The five FP-associated metabolites were detected in all lines but the concentrations of stachyose, sucrose and trehalose showed greater differences between [Védrantais]<sup>R-PI161375</sup> and Védrantais than between [Charentais T]<sup>R-PI482398</sup> and Charentais T. This contrast was unexpected as Védrantais and its transgenic line [Védrantais]<sup>R-PI161375</sup> differ only for the *Vat* transgene while the near -isogenic lines Charentais T/[Charentais T]<sup>R-PI482398</sup> differ for about 2% of their genome. Remarkably, both resistance donor lines, PI 161375 and PI 482398, contained low amounts of Cb and Cb-Glc, when detectable, which meant they were unsuitable for the investigation into putative ingestion of EFP-associated metabolites by aphids. Pairs of Védrantais and Charentais T lines however, contained higher amounts of Cb and Cb-Glc than the resistant parental lines.

Using Védrantais/[Védrantais]<sup>R-PI161375</sup> we investigated whether the infestation of plants by aphids induced modification of sugar and cucurbitacin content compared to healthy plants [Fig.3 and Sup\_Tab. 2]. Levels of stachyose and sucrose in aphid-infested melon plants did not significantly differ from those of healthy plants, regardless of the aphid clone and the melon genotype. In contrast, galactinol levels were different in infested compared to healthy melon plants. Interestingly, galactinol levels were higher in GWD2-infested [Védrantais]<sup>R-PI161375</sup> melon plants than in healthy [Védrantais]<sup>R-PI161375</sup> while galactinol levels were lower in C6-infested [Védrantais]<sup>R-PI161375</sup> than in healthy [Védrantais]<sup>R-PI161375</sup>. No significant differences were observed in the levels of cucurbitacin B and its glycoside in any of the plant-clone combinations compared to healthy plants [Fig.3].

### Identification of phloem-associated metabolites accumulated by aphids

To ensure that most of the melon phloem-associated metabolites found in aphids were accumulated during the experiment, we purged aphids after mass rearing, either by allowing them to feed on a plant free of the cucurbit marker metabolites, *i.e.* okra, or by fasting aphids in empty petri dishes. As expected, in two-week-old okra leaves, neither stachyose nor Cb were detected [Fig. S1 and Sup\_Tab. 3]. Rearing aphids on okra was unsuccessful however: after two weeks, the aphids had poorly colonized



the plants and showed noticeable weight loss and had laid very few eggs. These dramatic effects on the biological performance of aphids ultimately led to the disqualification of okra as a system for purging aphids. The fasting alternative was tested for 3h, 6h and 23h on batches of 20 to 50 adults of the C6 clone, previously reared on Védraçais melon plants. The metabolite quantities in aphid batches are shown in [Fig. 4 and Sup\_Tab. 4]: average amounts of stachyose, sucrose and galactinol decreased by 94, 80 and 84% respectively after 3 hours of fasting. Thereafter, and until the end of the fasting period, the quantity of these three sugars remained stable and similar to the quantities recorded at t=3h. In contrast, average amounts of raffinose and trehalose remained stable during the entire fasting period. Notably, neither Cb nor Cb-Glc were detected in aphid batches. Therefore, 3-hour fasting appeared to be efficient in purging stachyose, sucrose and galactinol from aphids.

## Search for aphid-accumulated phloem metabolite markers

After 4-hours of fasting, aphids were fed on either Védraçais or [Védraçais]<sup>R\_P161375</sup> for 17 hours. Metabolites were not available in same quantities for aphids according to the melon plant they fed on. To compare quantities of metabolites collected by each aphid batch, we considered the ratio [metabolite quantity in an aphid batch]/[metabolite quantity in the leaf on which they fed].

Each experiment focused on a *Vat*-adapted clone: GWD2 which triggers *Vat*-mediated resistance in [Védraçais]<sup>R\_P161375</sup>, and C6 which does not. In both experiments, NM1 was used as a reference for triggering of *Vat*-mediated resistance in [Védraçais]<sup>R\_P161375</sup>.

The three RFO related metabolites, stachyose, sucrose and galactinol, were detected in all aphid batches feeding on Védraçais and [Védraçais]<sup>R\_P161375</sup> [Fig. 5 and Sup\_Tab. 5; Fig. S2 and Sup\_Tab. 6]. NM1 accumulated significantly less sugar on [Védraçais]<sup>R\_P161375</sup> than on Védraçais (on average 6.2, 3.1 and 2.6 times less for stachyose, sucrose and galactinol respectively in the first experiment and 4.5, 4.2 and 3.4 times less in the second). On the contrary, both *Vat*-adapted clones, GWD2 in experiment 1 and C6 in experiment 2, accumulated similar amounts of sugar on [Védraçais]<sup>R\_P161375</sup> and Védraçais.

Remarkably, no Cb-Glc was detected in the aphids from either experiment. In NM1 and C6, no Cb was detected after feeding on Védraçais or [Védraçais]<sup>R\_P161375</sup>, therefore, NM1 and C6 did not accumulate Cb after fasting, regardless of the plant genotype. In GWD2, Cb was detected only in 2/9 batches which fed on [Védraçais]<sup>R\_P161375</sup> and only in very small amounts [Sup\_Tab. 7]. Therefore, few GWD2 aphids putatively accumulated Cb after feeding on [Védraçais]<sup>R\_P161375</sup> or, alternatively, were not fully purged after mass rearing. This second hypothesis was supported by the discovery of very small amounts of Cb in 8/15 batches of fasted NM1 and GWD2 [Sup\_Tab. 7].

In addition, we recorded the weight of aphid batches after a 4h-fasting period as well as for each

experimental condition following a 17h-feeding period [Fig. 6 and Sup\_Tab. 8]. After feeding on [Védrantais]<sup>R\_P161375</sup>, NM1 and GWD2 recorded marked weight losses (- 19.2 mg and - 9.3 mg on average for NM1 in experiments one and two and - 12.4 mg on average for GWD2) as compared to aphids fed on Védrantais. In contrast, C6 maintained its weight when feeding on [Védrantais]<sup>R\_P161375</sup> compared to Védrantais melon plants (+ 1.5 mg on average). Interestingly, of all the clones used, C6, which is *Vat*-adapted but did not trigger resistance on [Védrantais]<sup>R\_P161375</sup> had the lowest standard deviation for weight (on average 0.5 mg on Védrantais and 1.1 mg on [Védrantais]<sup>R\_P161375</sup>), indicating it had the most stable capacity for nutrient acquisition from its source phloem whether feeding on *Vat* or non-*Vat* melon plants. On the contrary, GWD2, which is *Vat*-adapted and which did trigger resistance in [Védrantais]<sup>R\_P161375</sup>, had a marked weight standard deviation, (on average 4.9 mg on Védrantais and 2.5 mg on [Védrantais]<sup>R\_P161375</sup>), indicating a variable capacity to acquire nutrients from either melon genotype. These results indicate that C6 was the most performant clone in terms of weight gain and weight stability on [Védrantais]<sup>R\_P161375</sup> melon plants.

## Discussion

As a strictly phloemophagous aphid species, *A. gossypii* must deal with an unbalanced diet since phloem sap, containing excess sugars and lacking amino acids, has a high C/N ratio. The cucurbit dual phloem system may offer a possibility for aphids to feed from the EFP which is enriched in nitrogenous compounds and might thereby better fulfil the nutritional needs of aphids than the FP (Fiehn, 2003; Richardson et al., 1982; B. Zhang et al., 2010), despite the presence of potentially toxic and/or repellent metabolites (Gaupels & Ghirardo, 2013; Kanvil et al., 2017).

## RFO accumulation in resistant plants does not affect aphids, but may play a role in aphid adaptation to fasting

In plants, RFO were suggested to be inducible under biotic stresses (Valluru & Van den Ende, 2011). In aphid-infested melon plants, levels of both stachyose and sucrose remained invariant as compared to the control plants, while galactinol, which is a precursor for RFO synthesis, varied significantly between aphid-infested and healthy plants [Fig. 3]. In [Védrantais]<sup>R\_P161375</sup>, infestation by GWD2 and NM1 led to an accumulation of galactinol while infestation by C6 led to a decrease in galactinol. As both NM1 and GWD2 trigger resistance when feeding on [Védrantais]<sup>R\_P161375</sup>, in contrast to C6, it appears that triggering resistance led to an induction of galactinol synthesis. Galactinol biosynthesis has been shown to increase during biotic stress (Kim et al., 2008) and evidence suggested it is involved at least in

systemic resistance induced by fungal and microbial pathogens (Cho et al., 2010; Kim et al., 2008). Also, galactinol exhibits antioxidant properties and, in plants challenged by biotic stress where accumulation of reactive oxygen species typically occurs, it may alter such a cell response (Valluru & Van den Ende, 2011). In *Vat*-melon plants, an oxidative burst is known to occur upon infestation by *A. gossypii* (Villada et al., 2009), but this reaction is limited to cells close to the aphid's probing site. Accumulation of galactinol could explain the restricted oxidative burst when resistance was triggered.

For aphids, the deterrent and negative impact of RFOs on the fecundity of *M. persicae* have been reported in choice and non-choice trials using transgenic RFO-producing *Arabidopsis thaliana* plants and on RFO-supplied artificial media (for example: galactinol, sucrose, raffinose, stachyose). Nevertheless, their toxicity to aphids was suggested to be indirect, potentially contributing to other defence responses in plants, such as the activation of secondary metabolism (T. Cao, 2010; T. Cao et al., 2013). NM1 accumulated more stachyose, sucrose and galactinol when feeding on Védraçais than on [Védraçais]<sup>R\_PI161375</sup> while it multiplied better on Védraçais, showing that these RFOs had no direct effect on *A. gossypii* performance [Fig. 1 and S2].

Moreover, we observed that stachyose, sucrose and galactinol, but not trehalose or raffinose, were purged from aphids following fasting. Trehalose is known to be a common hemolymph sugar storage form in insects and a possible dehydration tolerance agent (S. N. Thompson, 2003), suggesting it might never totally disappear from aphids and that its accumulation might be favoured in response to a fasting period when dehydration occurs. Raffinose is the intermediate form between sucrose and stachyose in the RFO biosynthesis pathway (Sengupta et al., 2015), but while sucrose and stachyose could be purged, raffinose could not [Fig. 4], suggesting its accumulation could confer a potential advantage to the aphid.

## **Cucurbitacin *in planta* is not a real threat to *Aphis gossypii***

LC-MS analysis revealed the presence in leaves of the aglycone cucurbitacin B (Cb) and its glycosylated derivative (Cb-Glc), which has been previously reported in cucurbits (Haq et al., 2019; Rajasree et al., 2016). Both forms were detected in highly variable amounts in melon genotypes [Fig. 2], suggesting that genotypes and growing conditions might influence cucurbitacin synthesis in melon.

In general, the glycosylated derivatives of secondary metabolites are inactive storage forms which are easily mobilized in response to attacks by pests (Gachon et al., 2005; W. Zhang et al., 2022); however, neither Cb nor Cb-Glc levels were modified upon infestation, independently of the triggering of *Vat*-mediated resistance [Fig. 3]. This was unexpected as cucurbitacins have already been shown to be inducible upon herbivory (Gaupels & Ghirardo, 2013). In aphid samples, Cb-Glc was not detected, and

in *A. gossypii* in particular, the description of a mechanism of glycosyltransferase-mediated detoxification of allelochemicals has already suggested that the glycosylated derivatives are non-toxic and easily excreted by aphids (Nagare et al., 2021). In contrast, aglycone allelochemical forms, and notably Cb, are well known for their toxic/repellent effect on insects (Elsayed, 2011; Gaupels & Ghirardo, 2013; Torkey et al., 2009; Yousaf et al., 2018). In *A. gossypii*, such negative impacts of Cb have only been reported *in vitro* or via the application of exogenous cucurbitacin on melon leaves (Yousaf et al., 2018; Zhao et al., 2021) and interestingly Cb was detected in a few batches of *A. gossypii*, suggesting aphids can accumulate certain amounts of Cb during the experiment, even when they fed on *Vat*-melon. Remarkably, Cb levels were very low in the two *Vat* gene donor lines, PI 161375 and PI 482398 [Fig. 2]. Altogether, these results suggest that cucurbitacin-related compounds are not involved in melon resistance to *A. gossypii*.

If not a threat *in vivo*, Cb detoxifying/excreting mechanisms in *A. gossypii* may contribute to its specialization as cucurbit host-race clones. Nevertheless, as Cb was only detected in a few aphid samples and always in low quantities, the hypothesis of an adaptation to cucurbitacin might not be relevant, even for highly adapted clones such as NM1, C6 and GWD2 that could not be multiplied on okra, a Malvaceae species expected to be a generalist host for *A. gossypii* (Carletto et al., 2009).

## ***A. gossypii* clones only feed from cucurbit FP even in resistant melon**

Current knowledge on aphid interaction with the dual phloem system in cucurbits is quite sparse. Using stylectomy, Kanvil et al. (2017) showed *M. persicae* fed exclusively from the FP. However, it is less obvious in the case of *A. gossypii*, as only indirect evidence is available (Kanvil et al., 2017). We showed that all three *A. gossypii* clones exploited the FP as a food source, independently of the plant genotype on which they fed. Remarkably, the two *Vat*-adapted clones, GWD2 and C6, accumulated the same amount of FP markers when fed on Védraçais or [Védraçais]<sup>R-PI161375</sup>, suggesting that both clones accessed the FP, and that when resistance was triggered, *e.g.* by GWD2 on [Védraçais]<sup>R-PI161375</sup>, access to the FP was not impaired. Nevertheless, GWD2 weighed less on [Védraçais]<sup>R-PI161375</sup> than on Védraçais melon plants, whereas C6 only displayed on average a slightly higher weight on [Védraçais]<sup>R-PI161375</sup>. In light of these results, and bearing in mind that C6 does not trigger resistance in [Védraçais]<sup>R-PI161375</sup> whereas GWD2 does, it seems that triggering resistance impaired the nutritional quality of melon plants for the clones which are adapted to *Vat*.

Cb was detected in a few batches of *A. gossypii* fed on melon plants, always at trace levels, suggesting that aphids can probe EFP but only occasionally and unsustainably. The weight loss for NM1 feeding on [Védraçais]<sup>R-PI161375</sup> was thus probably unrelated to EFP feeding, which has been confirmed by

electrical penetration graph studies showing that NM1 on *Vat* plants had difficulty probing, and in the rare cases it reached the phloem, it did not sustainably feed (Boissot, Schoeny, et al., 2016). As for GWD2, a clone adapted to *Vat*, sustained feeding from the FP occurred, as revealed by aphid sugar contents, but reaching the phloem could take longer on [Védrantais]<sup>R\_PI161375</sup> than on Védrantais, potentially explaining the aphid weight loss measured over the short duration of the experiment (17h). Therefore, we have shown that *A. gossypii* cucurbit host-race clones fed from the FP of susceptible melon and that adapted clones on resistant melon also fed from the FP. To feed successfully from the FP, *A. gossypii* has to overcome sieve plate occlusion which was shown in resistant melon when infested by *A. gossypii* (Peng & Walker, 2018). Components of aphid watery saliva have been suggested to play a role in preventing and/or reversing sieve plate occlusion by proteins, callose plugs and sieve element protein-lining (Åhman et al., 2019; Aart J.E. van Bel & Will, 2016). In the melon-*A. gossypii* interaction system, it has been shown that aphid recognition by *Vat* induced callose and lignin deposition on the cell walls along the stylet path (Villada et al., 2009). The decrease in callose deposition has been suggested to involve Ca<sup>2+</sup> sequestration by aphid salivary effectors (Aart J.E. van Bel & Will, 2016). Cucurbit sieve-element proteins PP1 and PP2 were degraded *in vitro* by the saliva of *Acyrtosiphon pisum* and *Macrosiphum euphorbiae* – both species not found to colonize cucurbits (Furch et al., 2015). Therefore, the role of effectors is probably a major part of the ability of aphids to colonize cucurbits, suggesting that evolution of phloem complexity did not play a role in the ability of *A. gossypii* to specialize on cucurbits.

## Materials and methods

### Plant and aphid material

We used two aphid-resistant melon lines: PI 161375 from Korea, and PI 482398 from Zimbabwe, and two aphid-susceptible melon lines: Védrantais and Charentais T from France. The melon lines were obtained from the Centre for Vegetable Germplasm (CRB-Leg) (Salinier et al., 2022). Accessions with their genebank identifier, year of introduction and origin are as follows:

- PI 161375: ME00591 (1967), Korea
- PI 482398: ME00483 (1997), Zimbabwe
- Védrantais: ME01000 (1982) a commercial variety originally supplied by Vilmorin
- Charentais T: ME00695 (1980) part of the French national collection, originally supplied by Vilmorin

Seed exchange complied with all obligations arising from, and does not contravene any specific use restrictions of: the United Nations' Convention on Biological Diversity as entered into force on

December 29, 1993 (CBD); the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity as entered into force on October 12, 2014; the Convention on International Trade in Endangered Species of Wild Fauna and Flora as entered into on March 3, 1973 (CITES); and all their implementing and other relevant national, local or indigenous access and benefit-sharing laws.

We also used a transgenic Védraçais resistant line (built by the team) with an insertion of an 11-kb genomic DNA sequence which included the *Vat* allele under the control of its own promoter isolated from PI 161375 (Catherine Dogimont et al., 2014), named [Védraçais]<sup>R\_P161375</sup>. This line was used according to the agreement DUO 7306 (2020-2025) delivered by the French 'Ministère de l'enseignement supérieur, de la recherche et de l'innovation the 2020/06/22. We also used a near-isogenic line [Charentais T]<sup>R\_P1482398</sup>, (built by the team), derived from a cross between PI 482398 and Charentais T, followed by six backcrosses with the Charentais T parental line. At each generation, plants showing resistance to NM1 were selected. No voucher specimen of this material has been deposited in a publicly available herbarium. Plantlets were grown in an insect-proof greenhouse up to the two-leaf stage and were subsequently used for biological tests.

We used three *A. gossypii* clones of the cucurbit host-race, which produced contrasted patterns of resistance on *Vat* melon plants, i.e. plants carrying a resistant allele in the *Vat* gene cluster. Clones were maintained by synchronous mass rearing on Védraçais melon, or on okra, a plant from the Malvaceae family expected to act as a host for any aphid host race (Carletto et al., 2009) at 24°C/18°C under a 16 h/8 h day/night photoperiod.

## Experimental design and time course for plant/aphid interaction study

### Experimental design

Two independent experiments were conducted. Experiment 1 used NM1 and GWD2 and experiment 2 used NM1 and C6. Both experiments used the Védraçais and [Védraçais]<sup>R\_P161375</sup> melon plants. Plants and aphids were placed in a growth chamber at 24/18°C (day/night), with a photoperiod of 16/8 h (day/night). Védraçais and [Védraçais]<sup>R\_P161375</sup> plants were arranged in staggered rows to reduce positional effects. Plantlets were grown in trays filled with water, each tray containing 6 plantlets with sufficient space between plantlets to avoid leaf overlap. Plastic rings were glued to the base of each of the two plantlet limbs in order to prevent aphids escaping from the leaves [Fig. S3].

### Experimental time course

At time zero (t<sub>0</sub>), batches of 60 mass-reared aphids were transferred to small glass petri dishes using a moist paintbrush, which were then sealed using parafilm. Aphids were then fasted *in vitro* for 4 hours.

After four hours (t=4h), 5 batches of NM1 and 9 batches of GWD2 (experiment 1) and 3 batches of NM1 and 7 batches of C6 (experiment 2; used as control for fasting efficiency), were transferred to 2 mL Eppendorf tubes. The samples were weighed, then immediately frozen and stored at -80°C.

At t=4h, the other batches were gently deposited onto a plantlet (60 aphids per plants *i.e.* 30 aphids per limb) with a paintbrush. For the first experiment, batches of NM1 and GWD2 were deposited onto nine Védraçais and [Védraçais]<sup>R-PI161375</sup> plants, for the second experiment, batches of NM1 and C6 were deposited onto seven plants of each genotype. The same number of control plantlets underwent the same treatments as the aphid-infested ones: with plastic rings glued at the base of their two leafstalks and using a paintbrush to simulate aphid deposition. After 21 hours (t=21h), aphids were removed from the leaflets and transferred to 2 mL Eppendorf tubes. Pairs of limbs from each plantlet were transferred to 15 mL Corning tubes. Samples were immediately frozen and then stored at - 80°C. A total of 54 plants and 50 aphid batches were sampled during the first experiment and 42 plants and 38 aphid batches were sampled during the second experiment.

## Plant and aphid metabolite extraction

All solvents were analytical or LC/MS grade. Pure standards were purchased from Extrasynthèse (Genay, France) and ChemFaces (Wuhan, China).

Leaf samples were ground into a fine powder in liquid nitrogen, approximately 200 mg of powder were transferred into tubes of 2 ml and then 1.6 ml of methanol acidified with formic acid (0.125%; v/v) were added. The tubes were incubated at 70°C for 10 min in a water bath, 0.4 ml of ultra-pure water were added and the samples were incubated under the same conditions. At each step, the samples were homogenized with a vortex. The samples were centrifuged at 5300 rcf for 5 min at 4°C and supernatants were filtered through a 0.2 µm membrane. For LC analysis, the aliquots of filtered extracts were stored at - 20°C. For GC analysis, 10µl of the filtered extracts were dried overnight under vacuum and stored at - 20°C.

Aphid samples were ground directly in the extraction solvent: 0.8 ml of methanol acidified with formic acid (0.125%; v/v) were added and the samples were ground in a mixer mill containing two cooled steel balls for 45 seconds with 30 oscillations per second. The samples were incubated in a water bath at 70°C for 10 min and homogenized twice with a vortex during the incubation period. Then, 0.2 ml of ultrapure water were added and the tubes were placed back in the water bath. The extracts were filtered through a 0.2 µm membrane.

For LC analysis, 700 µl of the filtered extracts were concentrated to dryness in a vacuum concentrator and stored at -20°C. Immediately before injection, the dried residue was solubilized with 175 µl of methanol. For GC analysis, 50 µl of the filtered extracts were dried under vacuum and stored at - 20°C.

## Liquid Chromatography-Mass Spectrometry (LC-MS)

Ultra-Performance Liquid Chromatography (UPLC) was carried out using an Acquity UPLC Class I (Waters, Milford, MA, USA). Metabolite separation was carried out on a reverse phase column, BEH C18 1.7  $\mu\text{m}$  2.1\*100 mm, equipped with a guard column (Waters, USA). The mobile phase, consisting of water with formic acid (0.1%, v/v) (eluent A) and acetonitrile with formic acid (0.1%, v/v) (eluent B), was pumped at 0.4 ml/min. The eluent gradient was: 0-0.5 min: 2% B; 0.5-6.5 min: 2-44.5% B; 6.5-7.5 min: 44.5-100% B; 7.5-9 min: 100% B. Then, the concentration of B was decreased to 2% and the column was equilibrated before the next injection (total run time 10 min). The column temperature was held at 35°C and the samples at 10°C. For leaf samples, 1  $\mu\text{L}$  was injected. For aphid samples, 4  $\mu\text{L}$  were injected. The UPLC system was coupled with a triple quadrupole mass spectrophotometer TQ-XS XEVO (Waters, Milford, MA, USA) with an electrospray source at 150°C in positive ion mode with a 3.5 kV capillary voltage, 550°C desolvation temperature, 1000 L/h desolvation gas flow and 150 L/h scanning cone flow. The system was controlled by Masslynx version 1.2 (Waters, Milford, MA, USA). To optimize the detection of cucurbitacins, we injected authentic standards of cucurbitacin B (Cb) and cucurbitacin B 2-O-beta-D-glucoside (Cb-Glc) [Fig. 7A]. In scan mode (100-1000  $\text{u.m.a}$ ), mass spectra of Cb revealed, at the retention time ( $t_{\text{R}}$ ) of 7.01 min, the ammonium adduct  $[\text{M} + \text{NH}_4]^+$  ( $m/z$  576.3) and a direct fragmentation into ions 499, 481 and 463, which correspond to losses of the methyl ester group, and two water molecules [Fig. 7B]. Mass spectra of Cb-Glc revealed, at a  $t_{\text{R}}$  of 5.3 min, the ammonium adduct  $[\text{M} + \text{NH}_4]^+$  ( $m/z$  738,0) as well as fragments of the 499, 481 and 463 ions. We developed a multiple reaction monitoring (MRM) method using two transitions 576.3 > 499.2 and 499.2 > 481.2, with a cone voltage of 20 V and a collision energy of 60 V. We used the ion abundance measurement of  $m/z$  499.2 for quantification and made an external calibration using a range of cucurbitacin concentrations.

In order to ensure the detection of Cb and Cb-Glc in aphids, we measured the abundance of ions of  $m/z$  499 and 576; we then calculated the ratio of areas measured and compared this with the ratios obtained for the standards [Fig. 7C]. The limit of quantification in aphids was 0.3 nM in extracts and 0.005 nM.mg<sup>-1</sup>.

## Gas Chromatography – Mass Spectrometry (GC-MS)

Sample derivatization was performed online using a multipurpose sampler (Gerstel MPS, CTC Analytics AG, Mülheim an der Ruhr, Switzerland). The dry extracts were incubated in 50  $\mu\text{L}$  of a pyridine solution containing 20 mg/ml of methoxyamide hydrochloride. Incubation was carried out under shaking at 250 rpm for 6 min at 80°C, followed by 12 min at room temperature. Subsequently, 80  $\mu\text{L}$  of BSTFA (N,O-



Bis(trimethylsilyl)trifluoroacetamide) containing a mixture of 9 n-alkanes were added and the total mixture was heated for 30 min with constant stirring at 250 rpm. Derivatization was then continued for 52 hours at room temperature. The separation of the metabolites of the derived samples was carried out with a gas chromatography system (7890B GC, Agilent Technologies, Santa Clara, CA, USA) equipped with a capillary column (ZB-SemiVolatiles, 34.59 m, internal diameter 250  $\mu$ m, film thickness 250  $\mu$ m, Phenomenex, Torrance, USA) coupled to a time-of-flight (TOF) mass spectrometer (Pegasus BT, Leco, Saint Joseph, Benton Harbor, MI, USA). One microlitre of sample was injected in splitless mode at 270°C with helium as the carrier gas at 0.6 ml/min. The oven temperature was held at 80°C for 1 min, then gradually increased to 330°C (18°C/min) and held for 15 min after the set temperature was reached. Source and transfer line temperatures were set at 250°C. The acquisition was run after a delay of 700 seconds. Mass acquisitions were performed in scan mode in the m/z range of 50-630 with a cycle time of 10 scans/s. The multipurpose sampler was controlled by Maestro Version 1.4.40.1. The gas chromatography system and mass spectrometer were controlled by ChromatTOF Version 5.20.38.0.54864 (LECO, Saint Joseph, MI, USA). Data were deconvoluted using the LECO NTD software (LECO, USA), and the annotation was confirmed with the injection of pure standards. The amount of compounds was calculated as the area under the GC-MS peak representing each compound analysed, normalized by the weight of the corresponding sample (mg).

Standard solutions of sucrose, trehalose, raffinose, stachyose and galactinol were injected. Two extracted ion chromatograms (XIC) were selected for signal measurement. We used the sugar-typical XIC of m/z 361.2 for sucrose, trehalose, raffinose and stachyose and used the XIC of m/z 204.2 for galactinol [Fig. 8].

## Chromatogram analysis

Integration of chromatography peaks to infer quantities of metabolites was carried out with TargetLynx (Waters, USA). No quantifiable traces were reported as 'null'. Both Cb and Cb-Glc were quantified using a calibration curve of authentic standards of Cb, Cb-Glc being expressed as equivalent Cb. For metabolites analysed by GC-MS, the areas under the peaks of a characteristic extracted ion chromatogram (EIC) were used for relative quantification. All statistical analyses were conducted with XLSTAT software (AddinSoft, Paris, France). Kruskal-Wallis tests were performed using a Monte Carlo method.

## Declarations

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable.

## Availability of data and materials

All data are available in a supplementary file and in the Data gouv repository accessible at <https://doi.org/10.57745/3LI3FJ>.

## Competing interests

Not applicable

## Funding

This research was funded by Enza Zaden within a CIFRE PhD grant, and with the contribution of Rijk Zwaan.

## Authors' contributions

Biological experiments: PS, MNC and NB; LC-MS: MNC; GS-MS: MNC, PS, RL; Draft manuscript: PS, MNC and NB, Review: RL.

## Acknowledgments

We thank Pascale Mistral, Virginie Chareyron, Karine Leyre and Véronique Chovelon for their excellent technical assistance. We also thank Olivier Chevallier for his significant help in the development of liquid chromatography analysis methods. This work was supported by the University of Avignon through the provision of metabolomic equipment. This manuscript was edited by Rebecca Stevens.

## References

- Åhman, I., Kim, S. Y., & Zhu, L. H. (2019). Plant Genes Benefitting Aphids—Potential for Exploitation in Resistance Breeding. *Frontiers in Plant Science*, *10* (November), 1–14.  
<https://doi.org/10.3389/fpls.2019.01452>.

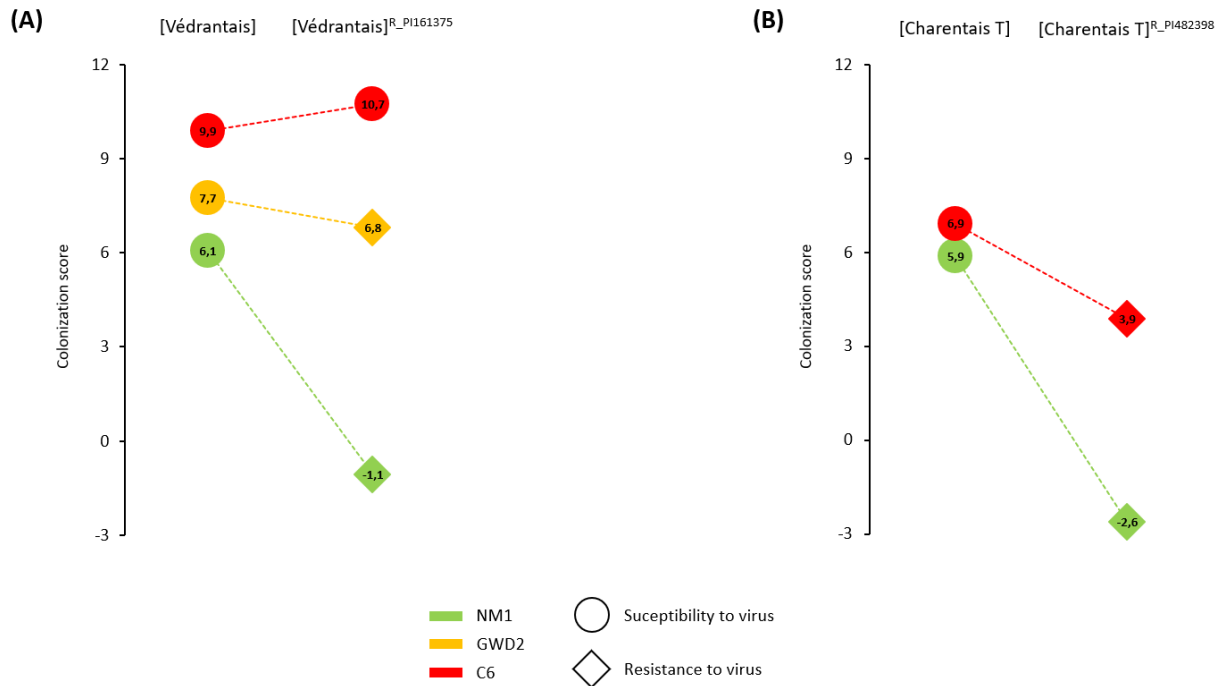
- Boissot, N., Schoeny, A., & Vanlerberghe-Masutti, F. (2016). Vat, an amazing gene conferring resistance to aphids and viruses they carry: From molecular structure to field effects. *Frontiers in Plant Science*, 7 (September 2016), 1–18. <https://doi.org/10.3389/fpls.2016.01420>.
- Boissot, N., Thomas, S., Chovelon, V., & Lecoq, H. (2016). NBS-LRR-mediated resistance triggered by aphids: Viruses do not adapt; aphids adapt via different mechanisms. *BMC Plant Biology*, 16(1), 1–12. <https://doi.org/10.1186/s12870-016-0708-5>.
- Cao, T. (2010). Metabolic engineering of raffinose-family oligosaccharides in the phloem reveals alterations in patterns of carbon partitioning and enhances resistance to green peach aphid. University of North Texas.
- Cao, T., Lahiri, I., Singh, V., Louis, J., Shah, J., & Ayre, B. G. (2013). Metabolic engineering of raffinose-family oligosaccharides in the phloem reveals alterations in carbon partitioning and enhances resistance to green peach aphid. *Frontiers in Plant Science*, 4(JUL), 1–13. <https://doi.org/10.3389/fpls.2013.00263>.
- Carletto, J., Lombaert, E., Chavigny, P., Brévault, T., Lapchin, L., & Vanlerberghe-Masutti, F. (2009). Ecological specialization of the aphid *Aphis gossypii* Glover on cultivated host plants. *Molecular Ecology*, 18(10), 2198–2212. <https://doi.org/10.1111/j.1365-294X.2009.04190.x>.
- Chen, J. Q., Martin, B., Rahbé, Y., & Fereres, A. (1997). Early intracellular punctures by two aphid species on near-isogenic melon lines with and without the virus aphid transmission (Vat) resistance gene. *European Journal of Plant Pathology*, 103(6), 521–536. <https://doi.org/10.1023/A:1008610812437>.
- Cho, S. M., Kang, E. Y., Kim, M. S., Yoo, S. J., Im, Y. J., Kim, Y. C., Yang, K. Y., Kim, K. Y., Kim, K. S., Choi, Y. S., & Cho, B. H. (2010). Jasmonate-dependent expression of a galactinol synthase gene is involved in priming of systemic fungal resistance in *Arabidopsis thaliana*. *Botany*, 88(5), 452–461. <https://doi.org/10.1139/B10-009>.
- Chovelon, V., Feriche-Linares, R., Barreau, G., Chadoeuf, J., Callot, C., Gautier, V., Le Paslier, M.C., Berad, A., Faivre-Rampant, P., Lagnel, J. and Boissot, N. (2021) Building a cluster of NLR genes conferring resistance to pests and pathogens: the story of the Vat gene cluster in cucurbits. *Horticulture Research*, 8, 72. <https://doi.org/10.1038/s41438-021-00507-0>.
- Crafts, A. S. (1932). Phloem anatomy, exudation, and transport of organic nutrients in cucurbits. *Plant Physiology*, 7(2), 1–2.
- Dangl, J. L., Horvath, D. M., & Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science*, 341(6147), 746–751. <https://doi.org/10.1126/science.1236011>.
- Dogimont, C., Chovelon, V., Pauquet, J., Boualem, A., & Bendahmane, A. (2014). The Vat locus encodes for a CC-NBS-LRR protein that confers resistance to *Aphis gossypii* infestation and A.

- gossypii*-mediated virus resistance. *Plant Journal*, 80(6), 993–1004.  
<https://doi.org/10.1111/tpj.12690>.
- Elsayed, G. (2011). Plant secondary substances and insects behaviour. *Archives of Phytopathology and Plant Protection*, 44(16), 1534–1549. <https://doi.org/10.1080/03235408.2010.507957>
- Fiehn, O. (2003). Metabolic networks of *Cucurbita maxima* phloem. *Phytochemistry*, 62(6), 875–886.  
[https://doi.org/10.1016/S0031-9422\(02\)00715-X](https://doi.org/10.1016/S0031-9422(02)00715-X).
- Furch, A. C. U., Van Bel, A. J. E., & Will, T. (2015). Aphid salivary proteases are capable of degrading sieve-tube proteins. *Journal of Experimental Botany*, 66(2), 533–539.  
<https://doi.org/10.1093/jxb/eru487>.
- Gachon, C. M. M., Langlois-Meurinne, M., & Saindrenan, P. (2005). Plant secondary metabolism glycosyltransferases: The emerging functional analysis. *Trends in Plant Science*, 10(11), 542–549.  
<https://doi.org/10.1016/j.tplants.2005.09.007>.
- Garzo, E., Fernández-Pascual, M., Morcillo, C., Fereres, A., Gómez-Guillamón, M. L., & Tjallingii, W. F. (2018). Ultrastructure of compatible and incompatible interactions in phloem sieve elements during the stylet penetration by cotton aphids in melon. *Insect Science*, 25(4), 631–642.  
<https://doi.org/10.1111/1744-7917.12447>.
- Gaupels, F., & Ghirardo, A. (2013). The extrafascicular phloem is made for fighting. *Frontiers in Plant Science*, 4(JUN), 2–5. <https://doi.org/10.3389/fpls.2013.00187>.
- Haq, F. U., Ali, A., Khan, M. N., Shah, S. M. Z., Kandel, R. C., Aziz, N., Adhikari, A., Choudhary, M. I., Ur-Rahman, A., El-Seedi, H. R., & Musharraf, S. G. (2019). Metabolite Profiling and Quantitation of Cucurbitacins in Cucurbitaceae Plants by Liquid Chromatography coupled to Tandem Mass Spectrometry. *Scientific Reports*, 9(1), 1–11. <https://doi.org/10.1038/s41598-019-52404-1>.
- Kanvil, S., Pham, J., Lopez-Cobollo, R., Selby, M., Bennett, M., Beckingham, C., Powell, G., & Turnbull, C. (2017). Cucurbit extrafascicular phloem has strong negative impacts on aphids and is not a preferred feeding site. *Plant Cell and Environment*, 40(11), 2780–2789.  
<https://doi.org/10.1111/pce.13053>.
- Kim, M. S., Cho, S. M., Kang, E. Y., Im, Y. J., Hwangbo, H., Kim, Y. C., Ryu, C. M., Yang, K. Y., Chung, G. C., & Cho, B. H. (2008). Galactinol is a signaling component of the induced systemic resistance caused by *Pseudomonas chlororaphis* O6 root colonization. *Molecular Plant-Microbe Interactions*, 21(12), 1643–1653. <https://doi.org/10.1094/MPMI-21-12-1643>.
- Klingler, J., Powell, G., Thompson, G. A., & Isaacs, R. (1997). Phloem specific aphid resistance in *Cucumis melo* line AR 5: effects on feeding behaviour and performance of *Aphis gossypii*. *Entomologia Experimentalis et Applicata*, 86(1), 79–88. <https://doi.org/10.1046/j.1570-7458.1998.00267.x>.
- Mattson, W. J. (1980). Herbivory in Relation to Plant Nitrogen Content. *Annual Review of Ecology and*

- Systematics*, 11(1), 119–161. <https://doi.org/10.1146/annurev.es.11.110180.001003>.
- Mistral, P., Vanlerberghe-Masutti, F., Elbelt, S., & Boissot, N. (2021). *Aphis gossypii/Aphis frangulae* collected worldwide: Microsatellite markers data and genetic cluster assignment. *Data in Brief*, 36, 106967. <https://doi.org/10.1016/j.dib.2021.106967>.
- Nagare, M., Ayachit, M., Agnihotri, A., Schwab, W., & Joshi, R. (2021). Glycosyltransferases: the multifaceted enzymatic regulator in insects. *Insect Molecular Biology*, 30(2), 123–137. <https://doi.org/10.1111/imb.12686>.
- Peng, H. C., & Walker, G. P. (2018). Sieve element occlusion provides resistance against *Aphis gossypii* in TGR-1551 melons. *Insect Science*, 27(1), 33–48. <https://doi.org/10.1111/1744-7917.12610>.
- Powell, G., Tosh, C. R., & Hardie, J. (2006). HOST PLANT SELECTION BY APHIDS: Behavioral, Evolutionary, and Applied Perspectives. *Annual Review of Entomology*, 51(1), 309–330. <https://doi.org/10.1146/annurev.ento.51.110104.151107>.
- Rajasree, R. S., Sibi, P. I., Francis, F., & William, H. (2016). Phytochemicals of cucurbitaceae family – A review. *International Journal of Pharmacognosy and Phytochemical Research*, 8(1), 113–123.
- Richardson, P. T., Baker, D. A., & Ho, L. C. (1982). The chemical composition of cucurbit vascular exudates. *Journal of Experimental Botany*, 33(6), 1239–1247. <https://doi.org/10.1093/jxb/33.6.1239>.
- Salinier, J., Lefebvre, V., Besombes, D., Burck H., Causse, M., Daunay, M.C., Dogimont, C., Goussopoulos, J., Gros, C., Maisonneuve, B., McLeod, L., Tobal, F., Stevens, R. (2022). The INRAE Centre for Vegetable Germplasm: Geographically and Phenotypically Diverse Collections and Their Use in Genetics and Plant Breeding. *Plants* 2022, 11, 347. <https://doi.org/10.3390/plants11030347>.
- Sengupta, S., Mukherjee, S., Basak, P., & Majumder, A. L. (2015). Significance of galactinol and raffinose family oligosaccharide synthesis in plants. *Frontiers in Plant Science*, 6(AUG), 1–11. <https://doi.org/10.3389/fpls.2015.00656>.
- Thompson, S. N. (2003). Trehalose - The Insect “Blood” Sugar. In *Advances in Insect Physiology* (Vol. 31, Issue 03). [https://doi.org/10.1016/S0065-2806\(03\)31004-5](https://doi.org/10.1016/S0065-2806(03)31004-5).
- Torkey, H. M., Abou-Yousef, H. M., Azeiz, A., & Farid, H. E. A. (2009). Insecticidal effect of Cucurbitacin E Glycoside isolated from *Citrullus colocynthis* against *Aphis craccivora*. *Australian Journal of Basic and Applied Sciences*, 3(4), 4060–4066.
- Turgeon, R. (2016). Phloem Biology of the Cucurbitaceae. *Plant Genetics and Genomics: Crops and Models*, October, 291–311. <https://doi.org/10.1007/7397>.
- Valluru, R., & Van den Ende, W. (2011). Myo-inositol and beyond - Emerging networks under stress. *Plant Science*, 181(4), 387–400. <https://doi.org/10.1016/j.plantsci.2011.07.009>
- van Bel, A. J. E., & Will, T. (2016). Functional evaluation of proteins in watery and gel saliva of aphids. *Frontiers in Plant Science*, 7 (December 2016), 1–19. <https://doi.org/10.3389/fpls.2016.01840>

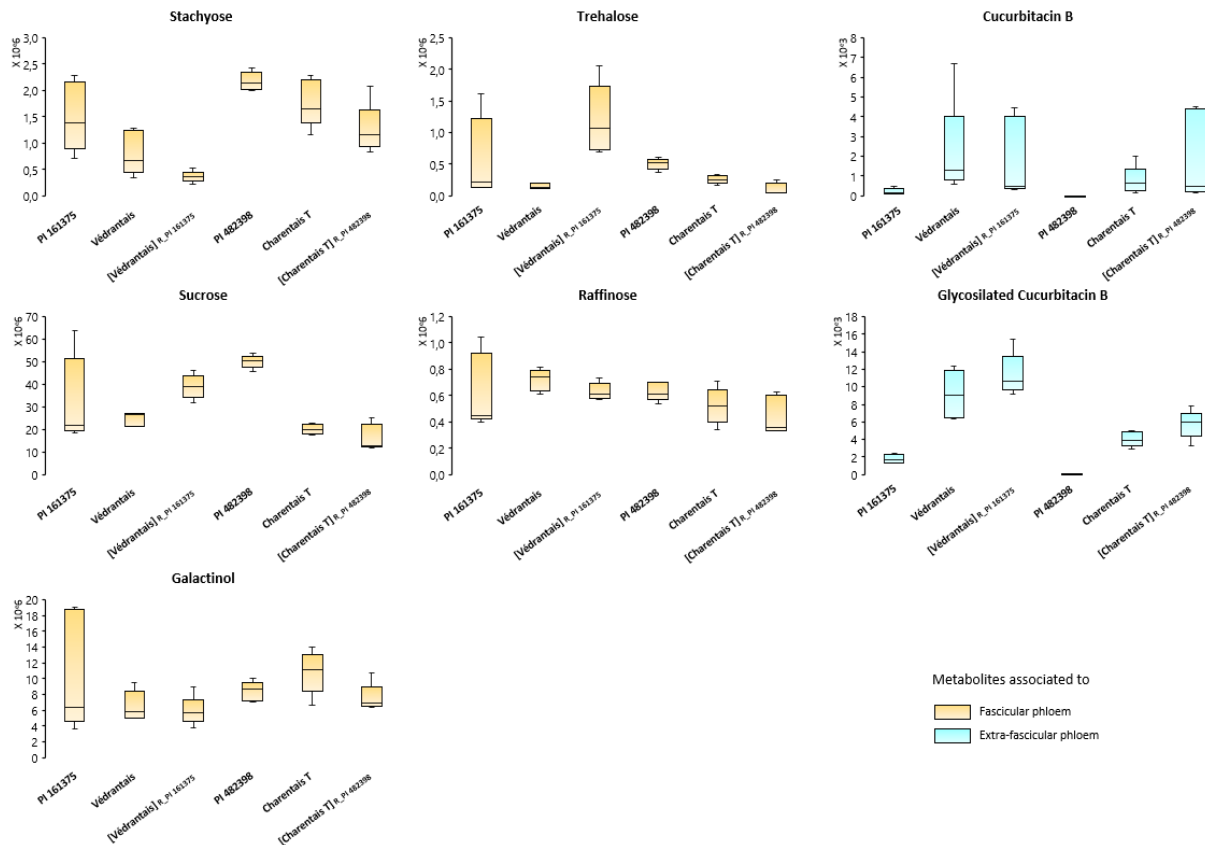
- Villada, E. S., González, E. G., López-Sesé, A. I., Castiel, A. F., & Gómez-Guillamón, M. L. (2009). Hypersensitive response to *Aphis gossypii* Glover in melon genotypes carrying the Vat gene. *Journal of Experimental Botany*, *60*(11), 3269–3277. <https://doi.org/10.1093/jxb/erp163>
- Walker, G. P. (2021). Sieve element occlusion: Interactions with phloem sap-feeding insects. A review. *Journal of Plant Physiology*, *269* (December 2021), 153582. <https://doi.org/10.1016/j.jplph.2021.153582>.
- Yousaf, H. K., Shan, T., Chen, X., Ma, K., Shi, X., Desneux, N., Biondi, A., & Gao, X. (2018). Impact of the secondary plant metabolite Cucurbitacin B on the demographical traits of the melon aphid, *Aphis gossypii*. *Scientific Reports*, *8*(1), 1–10. <https://doi.org/10.1038/s41598-018-34821-w>.
- Zhang, B., Tolstikov, V., Turnbull, C., Hicks, L. M., & Fiehn, O. (2010). Divergent metabolome and proteome suggest functional independence of dual phloem transport systems in cucurbits. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(30), 13532–13537. <https://doi.org/10.1073/pnas.0910558107>.
- Zhang, W., Wang, S., Yang, J., Kang, C., Huang, L., & Guo, L. (2022). Glycosylation of plant secondary metabolites: Regulating from chaos to harmony. *Environmental and Experimental Botany*, *194*(October 2021), 104703. <https://doi.org/10.1016/j.envexpbot.2021.104703>.
- Zhao, C., Ma, C., Luo, J., Niu, L., Hua, H., Zhang, S., & Cui, J. (2021). Potential of cucurbitacin b and epigallocatechin gallate as biopesticides against *Aphis gossypii*. *Insects*, *12*(1), 1–16. <https://doi.org/10.3390/insects12010032>.

## Figures



**Figure 1 Interactions between three aphid clones and two isogenic melon lines Védrantais and [Védrantais]<sup>R\_PI161375</sup>**

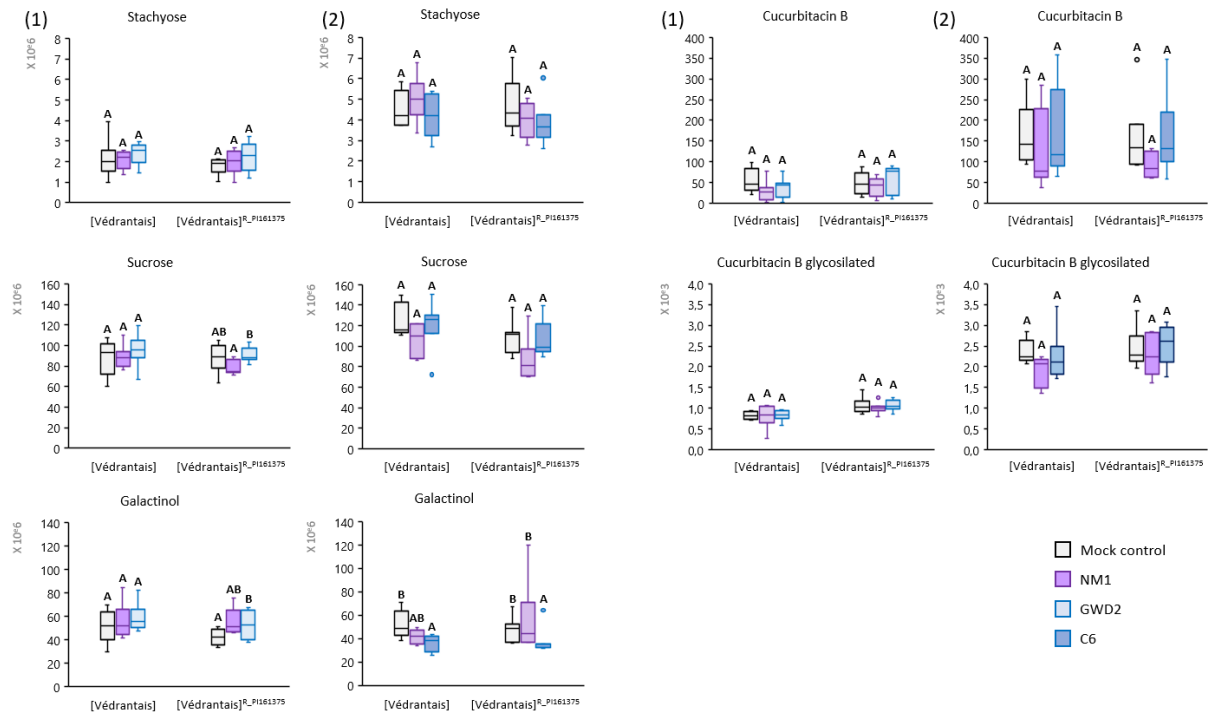
‘Colonization’ by aphids is a quantitative trait inferred from the number of nymphs and adults seven days after a controlled infestation. *Vat*-mediated resistance is triggered when plants are resistant to virus after aphid-inoculation with Cucumber mosaic virus, and NOT triggered when plants are susceptible. Data for the three aphid clones NM1, GWD2 and C6 from (Boissot, Thomas, et al., 2016).



**Figure 2 Quantification of seven phloem-associated metabolites in six melon lines**

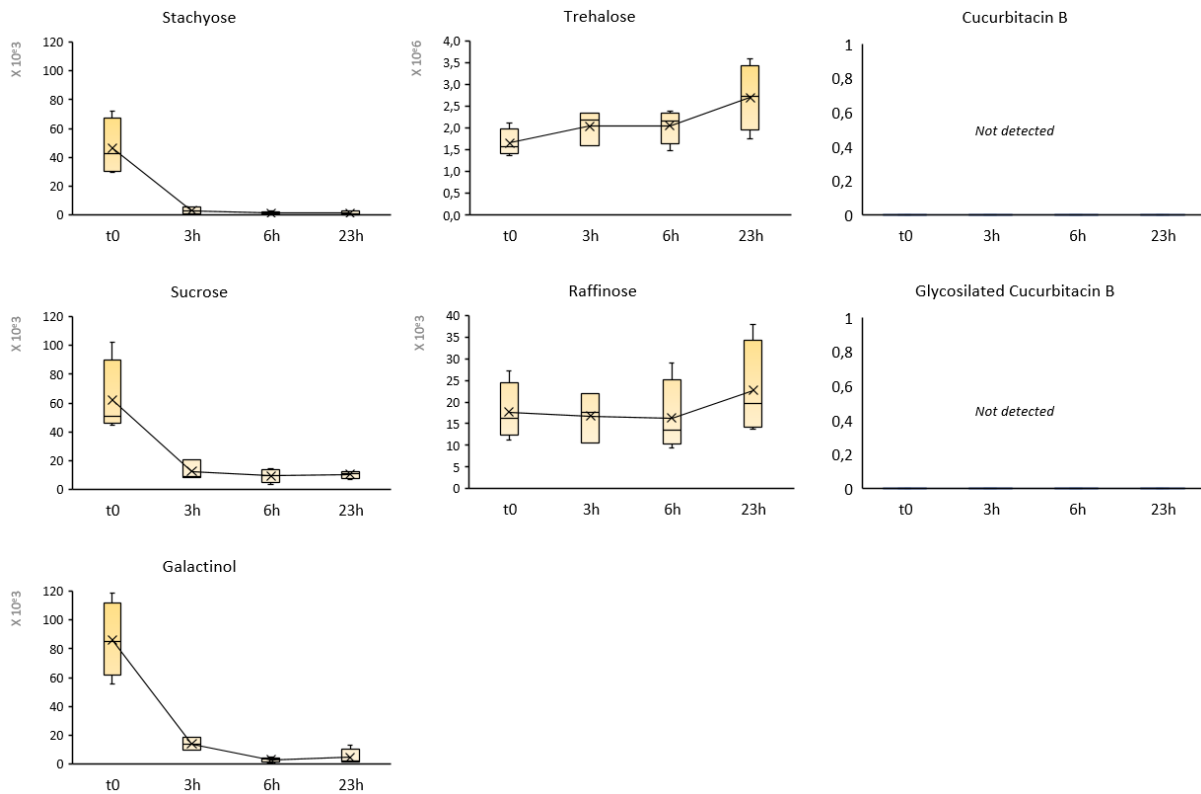
Metabolites associated with the melon fascicular phloem were analysed by GC-MS (displayed in orange) and quantified as the area under the chromatogram peak (no units) corrected by leaf sample weight (g). Metabolites associated with the melon extra-fascicular phloem were analysed by LC-MS (displayed in light blue) and expressed as  $\text{nM} \cdot \text{mg}^{-1}$ . Five melon plants were analysed four each melon genotype. PI 161375 and PI 482398 carry different Vat resistance alleles, Védramtais and Charentais T are susceptible to aphids. Their isogenic and near-isogenic lines with resistance donors are noted [Védramtais]<sup>R</sup><sub>PI 161375</sub> and [Charentais]<sup>R</sup><sub>PI 482398</sub>.





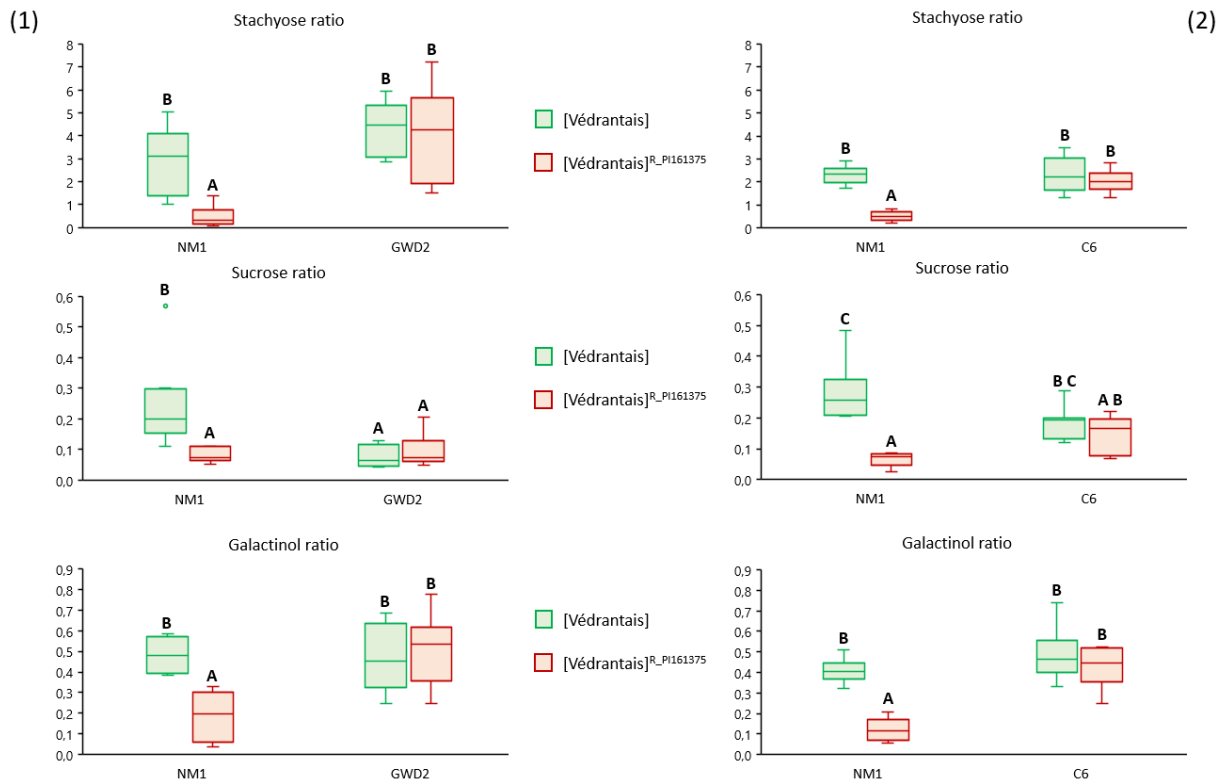
**Figure 3 Quantification of five phloem-associated metabolites in healthy versus aphid-infested melon plants**

Védrantais and its isogenic [Védrantais]<sub>R<sub>Pi161375</sub></sub> plants were infested for 17h by NM1 aphid clone - triggering the Vat-mediated resistance- in both experiments, and by either GWD2 aphids -triggering too the Vat-mediated resistance- in experiment **(1)** or C6 aphids -non-triggering the Vat-mediated resistance in experiment **(2)**. Stachyose, sucrose and galactinol were analysed by GC-MS in melon leaves and expressed as peak area (no units) corrected by leaf sample weight (g). Cucurbitacin B and Glycosylated Cucurbitacin B were analysed by LC-MS and expressed as nMole.mg<sup>-1</sup>. The number of observations per melon genotype was between 6 and 9. Letter codes indicate significant differences between infestation conditions for a melon genotype at alpha=0.05 according to Kruskal-Wallis tests.



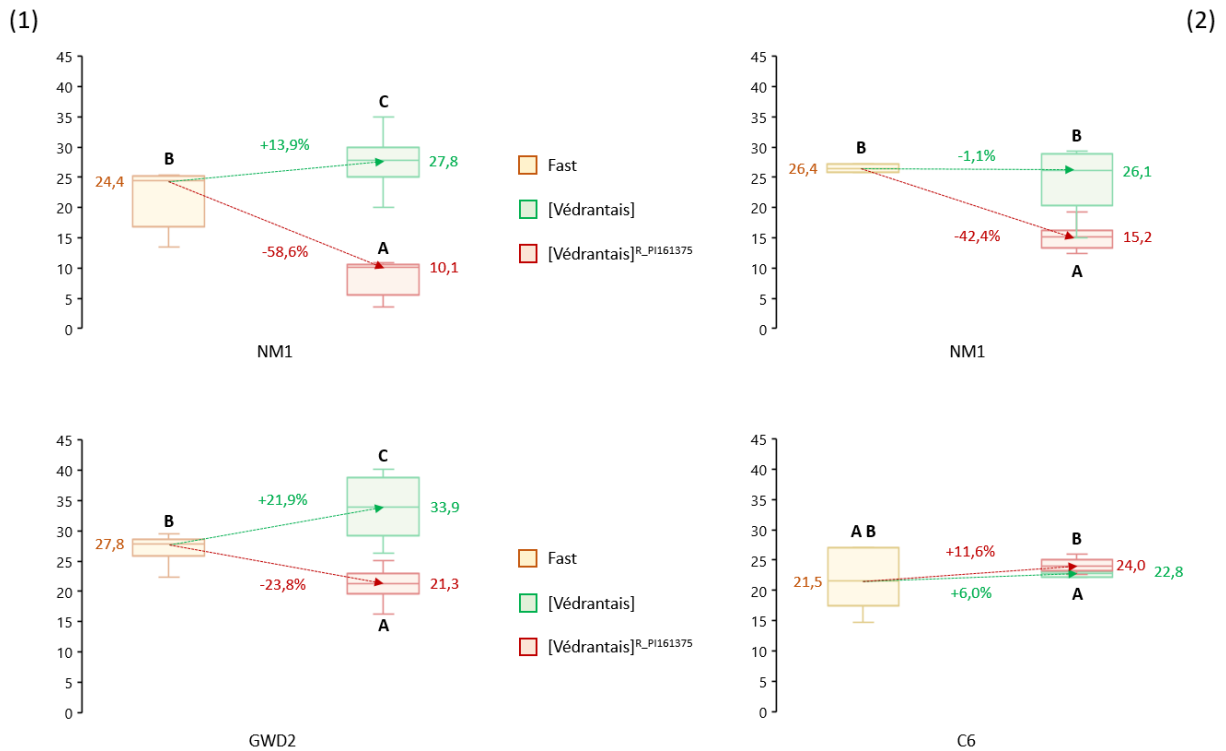
**Figure 4 Seven melon-associated metabolites in aphids during fasting**

C6 adult aphids were reared on Védrantais melon plants, then fasted by batches of 20 to 50 individuals for 3h, 6h and 23h. t0 corresponds to the aphids sampled at rearing (before fasting). For each time point, four aphid batches were sampled, except for the “3h” point where three aphid batches were sampled. Stachyose, sucrose and galactinol were analysed by GC-MS in aphid batches and expressed as peak area (no units) corrected by the aphid batch weight (mg). Cucurbitacin B and Glycosylated Cucurbitacin B were not detected by LC-MS.



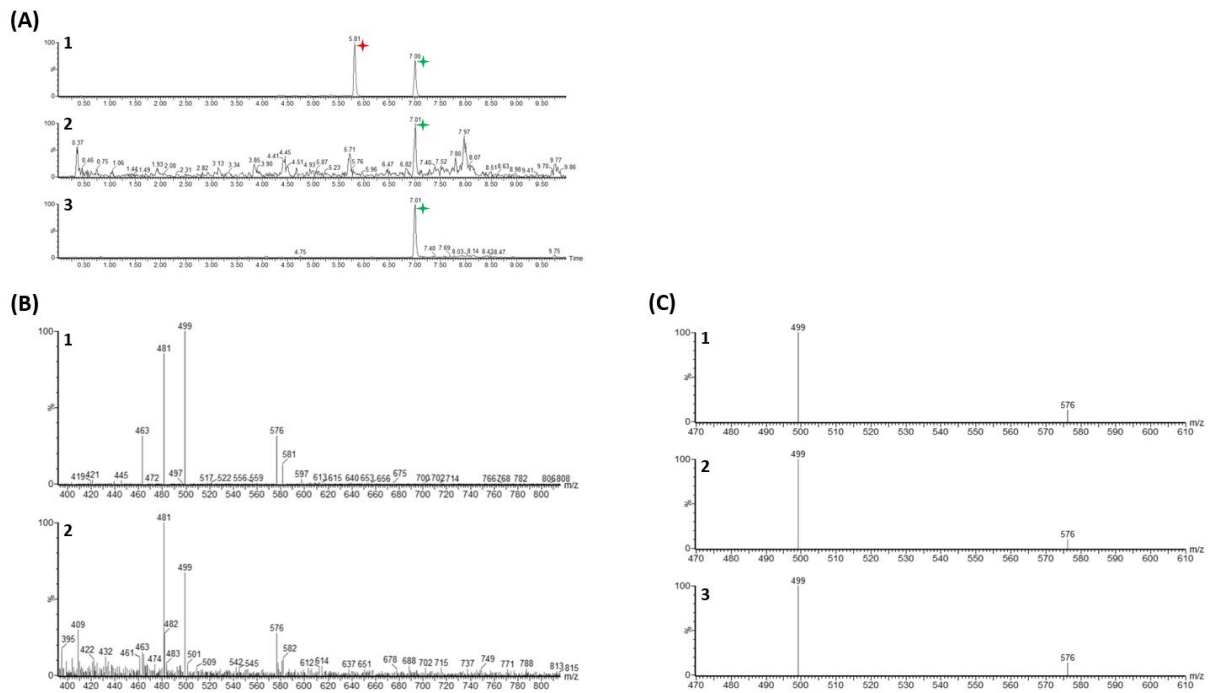
**Figure 5 Stachyose, sucrose and galactinol in three aphid clones fed on *Vat*/*non-Vat* melons**

Aphid clones were fed on Védrantais and its isogenic [Védrantais]<sup>R\_Pi161375</sup> melon lines for 17h. NM1 aphids -triggering the *Vat*-mediated resistance- were used in both experiments, with either GWD2 aphids -triggering the resistance in [Védrantais]<sup>R\_Pi161375</sup>- in experiment (1), or C6 aphids -non-triggering the resistance in [Védrantais]<sup>R\_Pi161375</sup>- in experiment (2). Results are ratios of [metabolite quantity in an aphid batch]/[metabolite quantity in the leaf on which they fed]. The number of observations is between 5 and 9. Letter codes indicate significant differences between all conditions within an experiment at alpha=0.05 according to Kruskal-Wallis tests.



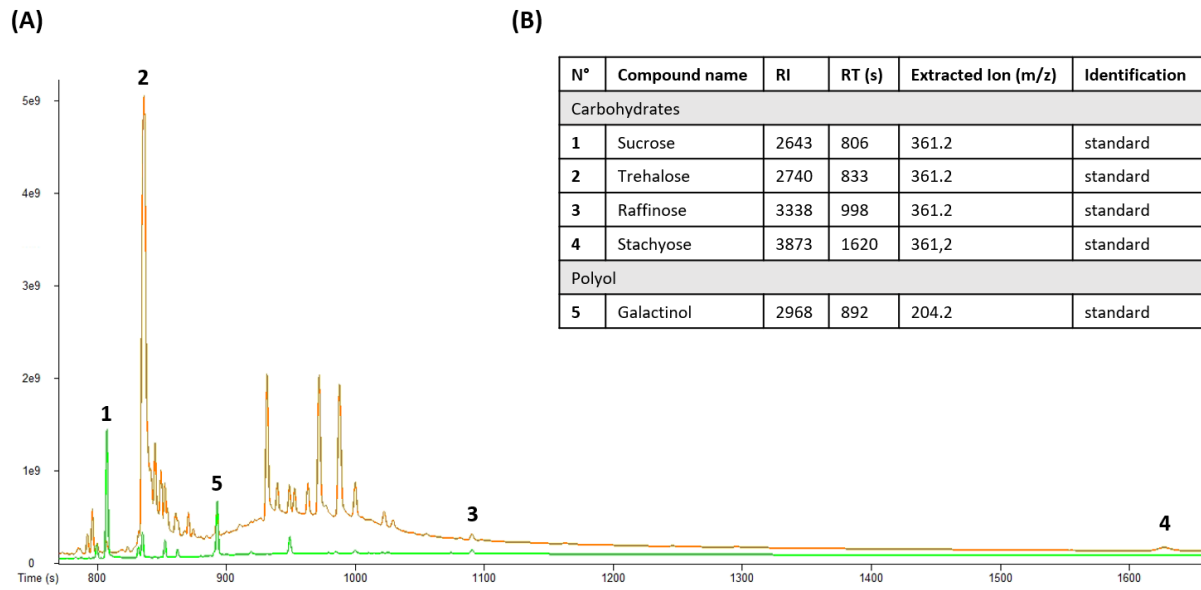
**Figure 6 Weight (mg) of three aphid clones fasted and fed on *Vat*/*non-Vat* melons**

Aphids were fasted for 4h and then fed on *Védrantais* and its isogenic [*Védrantais*]<sup>R\_PI161375</sup> melon lines for 17h. NM1 aphids -triggering the *Vat*-mediated resistance- were used in both experiments, with either GWD2 aphids -triggering the resistance in [*Védrantais*]<sup>R\_PI161375</sup>- in experiment (1), or C6 aphids -non-triggering the resistance in [*Védrantais*]<sup>R\_PI161375</sup>- in experiment (2). Arrows indicate the median loss/gain of weight of fasting aphid batches versus fed aphid batches. Between 3 and 6 aphid batches were analysed after fasting and between 7 and 9 after feeding. Letter codes indicate significant differences between fasted batches and fed batches at alpha=0.05 according to Kruskal-Wallis tests.



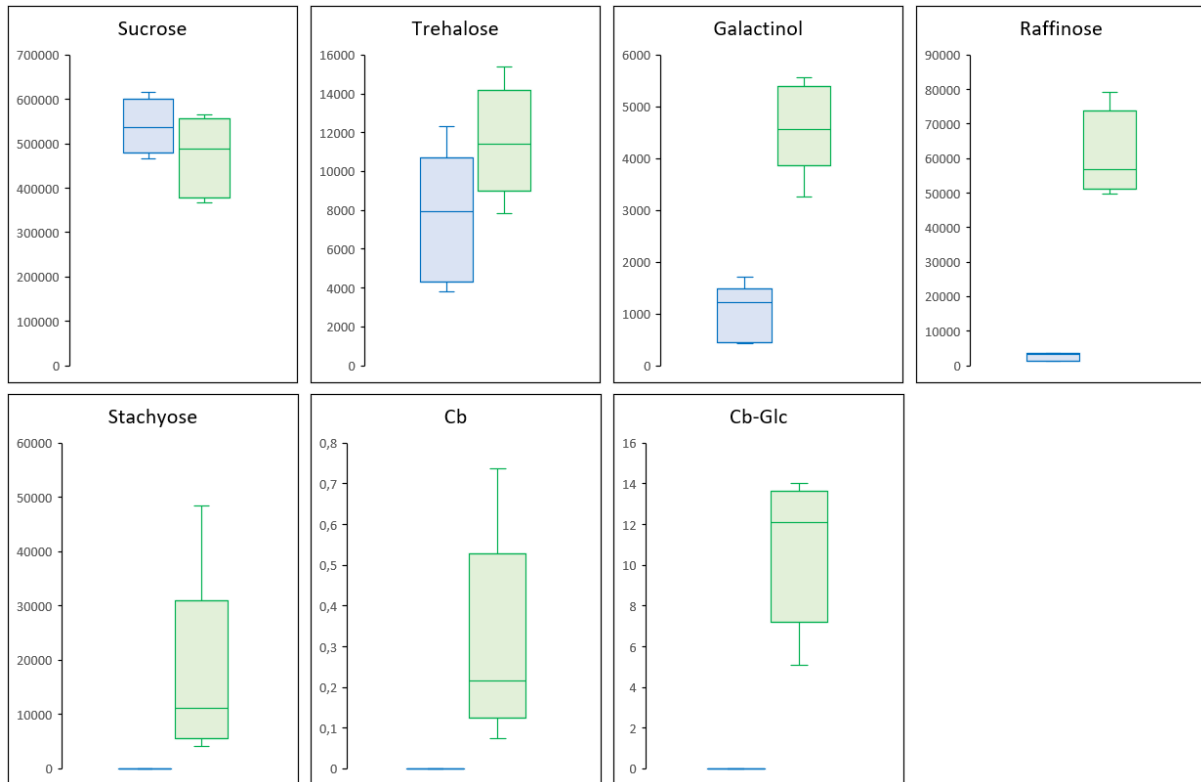
**Figure 7 Identification by LC-MS of cucurbitacin B and glycosylated cucurbitacin B in aphids and melons**

**(A):** LC-MS profiles of cucurbitacins (Extracted Ion Chromatogram of ion 499.2). 1: leaf extract, 2: aphid extract, 3: authentic standard of cucurbitacin B. Peak height was normalised to that of the largest peak. Red mark: glycosylated cucurbitacin B, green mark: cucurbitacin B. **(B):** LC-MS mass spectra of: 1: cucurbitacin B standard, 2: leaf extract at  $t_R = 7.01$  min. The peak heights were normalised to that of the largest peak. **(C):** MRM spectra of the two transitions:  $576.3 > 499.2$  and  $499.2 > 481.2$  at  $t_R 7.01$  min, showing the ratios between the two ions produced, 576 and 499: 1: cucurbitacin B standard (ratio 5.6), 2: aphid extract (ratio 6.7), 3: leaf extract (ratio 6.3).



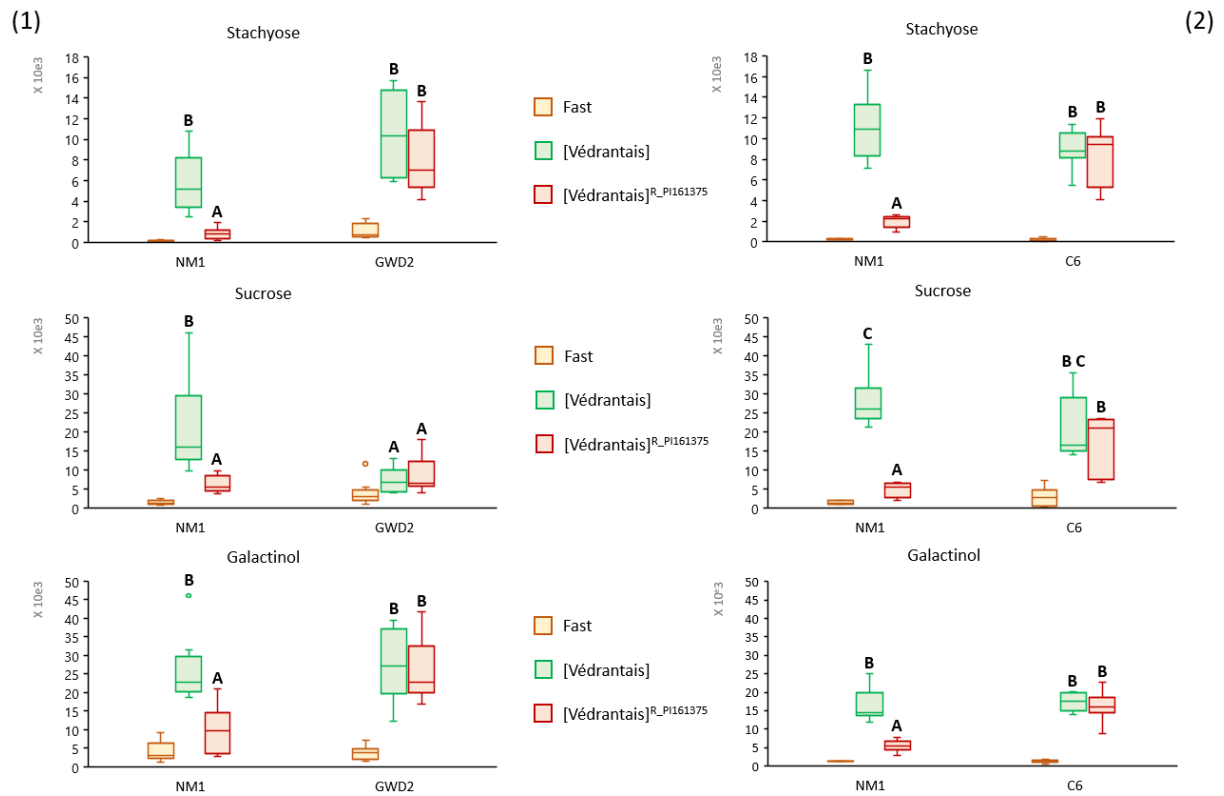
**Figure 8 Identification by GC-MS of four sugars and galactinol in aphids and melons**

**(A):** GC-MS profiles of extracts of leaves (green) and aphids (orange). Total Ion Current chromatogram showing 1: Sucrose, 2: Trehalose, 3: Raffinose, 4: Stachyose, 5: Galactinol. **(B):** list of compounds identified by GC-MS. RI: retention index, RT: retention time.



**Fig. S 1 Quantification of seven metabolites in Okra and Védraçais melon plants**

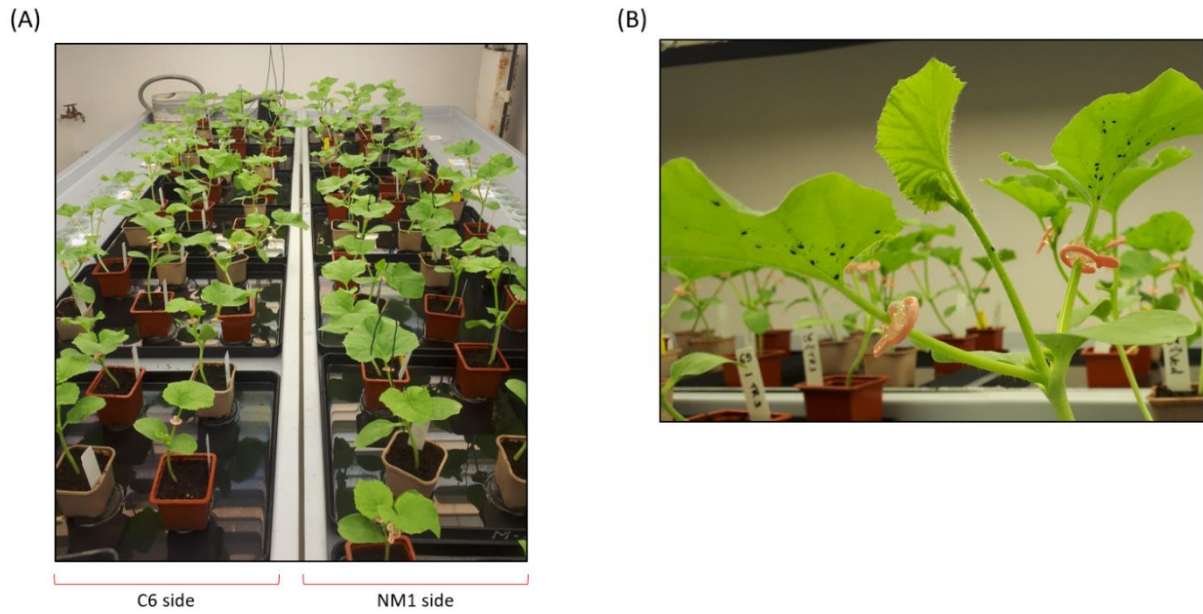
In Okra (blue boxplots) and in the Védraçais melon line (green boxplots). Sucrose, trehalose, galactinol, raffinose and stachyose were analysed by GC-MS and expressed as peak area/mg. Cucurbitacin B (Cb) and Glycosylated Cucurbitacin B (Cb-Glc) were analysed by LC-MS and expressed as nM/g. The number of observations per plant genotype was 5.



**Fig. S 2 Quantification of stachyose, sucrose and galactinol in three aphid clones fasted or fed on *Vat*/*non-Vat* melons**

Aphid clones were fasted for 4h and then fed on *Védrantais* and its isogenic [*Védrantais*]<sup>R\_PI161375</sup> melon lines for 17h. NM1 aphids -triggering the *Vat*-mediated resistance- were used in both experiments, with either GWD2 aphids -triggering the resistance in [*Védrantais*]<sup>R\_PI161375</sup>- in experiment (1), or C6 aphids -non-triggering the resistance in [*Védrantais*]<sup>R\_PI161375</sup>- in experiment (2). Stachyose, sucrose and galactinol were analysed by GC-MS in aphid batches and expressed as peak area (no units) corrected by the batch weight (mg). The number of observations is between 3 and 9. Letter codes indicate significant differences between fed aphids at level alpha=0.05 by Kruskal-Wallis analysis.





**Fig. S 3 – Growth chamber showing the experimental design for aphid feeding on melon plants**

**(A):** Global design. Each side was used for a different clone. Beige coloured plant pots corresponded to Védraçais and brown plant pots to [Védraçais]<sup>R<sub>PI161375</sub></sup>. The picture was taken one day before aphid infestation, during plant acclimation to the growth chamber. **(B):** 15-day old melon plantlets infested with a batch of aphids (C6). Red plastic rings were glued at the base of each of the infested limbs to prevent aphids escaping from leaves.



## **Chapitre 2 – Recherche de l'effecteur aphidien interagissant avec la protéine NLR VAT-1 du melon**

## 1. Contexte de l'étude

Dans le chapitre précédent, nous avons montré que la race d'hôte d'*Aphis gossypii* spécialisée sur les cucurbitacées s'alimente principalement dans le phloème fasciculaire (FP). Comme les agents pathogènes, les pucerons injectent dans la plante, (dans l'apoplaste ainsi que dans les cellules sondées lors du trajet vers le phloème) **des effecteurs** via leur **salive**. Ils établissent une relation compatible avec la plante lorsque ces pièces buccales peuvent atteindre le phloème dans lequel ils s'alimentent. Les pucerons doivent **manipuler les défenses de la plante pour empêcher le colmatage du phloème** (Will et al., 2007), la piqûre des pucerons étant perçue par la plante comme une blessure. Ils intègrent le **paradigme de l'évolution pour les interactions plante-pathogène** qui repose sur l'évolution de motifs moléculaires (effecteurs) des agents pathogènes reconnus par la plante (Dangl et al., 2013) : ces agents pathogènes s'adaptent à une réaction immunitaire des plantes, soit une réaction à l'échelle de l'espèce pour s'adapter à une résistance non hôte, soit une réaction à l'échelle d'une accession spécifique pour s'adapter à une **résistance de type NLR** (dite résistance hôte héritable).







**Le gène NLR *Vat-1*** du melon a la particularité de conférer une **résistance à la colonisation** par *A. gossypii* et une **résistance aux virus transmis** selon le mode non persistant, résistance élicitée par *A. gossypii* (Catherine Dogimont et al., 2014). La résistance *Vat-1*, présente dans la lignée PI 161375, est **déclenchée par un éliciteur aphidien et révélée par le phénotype de résistance au virus** inoculé par le puceron en même temps que l'éliciteur aphidien (Boissot, Thomas, et al., 2016; Boualem et al., 2016). Le marqueur du déclenchement de la résistance *Vat-1* est une **réponse d'hypersensibilité microscopique** située au niveau de la pique infligée par les stylets des pucerons (Villada et al., 2009). Cette réponse immunitaire est rapide et caractéristique de la résistance hôte héritable, au niveau macroscopique elle se traduit par une résistance aux virus.

La protéine VAT-1 est prédite **cytoplasmique** et comprend un domaine CC-NBS-LRR (coiled-coil, nucleotide-binding, leucine-rich-repeat) (Catherine Dogimont et al., 2014). Cette catégorie de protéines est associée aux mécanismes de défenses contre des agents pathogènes ainsi qu'aux hémiptères (Kaloshian & Walling, 2005, 2016b; Nicolis & Venter, 2018; Rafiqi et al., 2009; Tamura et al., 2014). En accord avec le cadre conceptuel établi pour les gènes codant pour des protéines NLR, le fonctionnement de *Vat* implique un **découplage de la phase de reconnaissance de l'invasion et celle de la réponse immunitaire** consécutive. Lors de la phase de reconnaissance, la protéine cytoplasmique VAT-1 doit reconnaître un éliciteur aphidien qui a été injecté via les stylets d'*A. gossypii* à l'intérieur des cellules de l'épiderme ou du mésophylle. La réponse immunitaire est déclenchée au niveau de la zone blessée et bloque la réplication et le mouvement des virus transmis (Boualem et al., 2016).

Nous définissons l'**avirulence** comme la **capacité des pucerons à éliciter la résistance au CMV** (virus de la mosaïque du concombre). Ces clones sont désignés comme 'avirulents' (abréviation **AVR**). Inversement, nous définissons la **virulence** comme la **capacité des pucerons à ne pas éliciter la résistance au CMV**. Ces clones sont désignés comme 'virulents' (abréviation **VIR**). Comme la résistance *Vat-1* doit être déclenchée par un éliciteur aphidien (reconnaissance) et peut être révélée par le phénotype de résistance au virus inoculé par le puceron en même temps que l'éliciteur aphidien, je ferai par la suite référence à l'avirulence comme à la **reconnaissance au puceron par *Vat-1***.

**Ce chapitre de ma thèse est consacré à la recherche de l'effecteur de puceron qui interagit avec la protéine VAT-1 du melon PI 161375.** Mes travaux ont consisté à faire des **analyses comparatives de données d'expression** de clones AVR et VIR pour établir des listes de protéines candidates à l'AVR.

L'avirulence/virulence est définie par la réponse observé sur le melon PI 161375 ou la lignée Védrantais transgénique contenant l'allèle *Vat-1* isolé à partir de PI 161375. Douze clones d'*A. gossypii* appartenant à trois races d'hôtes (Cucurbitacée, Malvacée et Solanacée) ont été utilisés au cours de ma thèse [Figure 8]. D'après les données disponibles au laboratoire, sept clones sont avirulents (*i.e.* déclenchent la résistance au CMV) et cinq sont virulents (*i.e.* ne déclenchent pas la résistance au CMV). Dans ma démarche de recherche, je me suis concentré sur les trois clones spécialisés sur les cucurbitacées C4 (AVR), C9 (AVR) et C6 (VIR). Ces trois clones sont génétiquement proches mais ils appartiennent à des lignées clonales indépendantes (génotypage à 8 marqueurs microsatellites) (Mistral et al., 2021).

				
<b>Races d'hôtes</b>		<b>Avirulents</b>	<b>Virulents</b>	
<b>Cucurbitacée</b>		C4 C9 NM1 CUC1 GWD2	C6	
<b>Malvacée</b>		Hib22 Hib23	Burk	
<b>Solanacée</b>			Aub1 Psp4 Pot1	

**Figure 8** – Caractéristiques des 12 clones d'*A. gossypii* utilisés pour la recherche de candidats à l'AVR

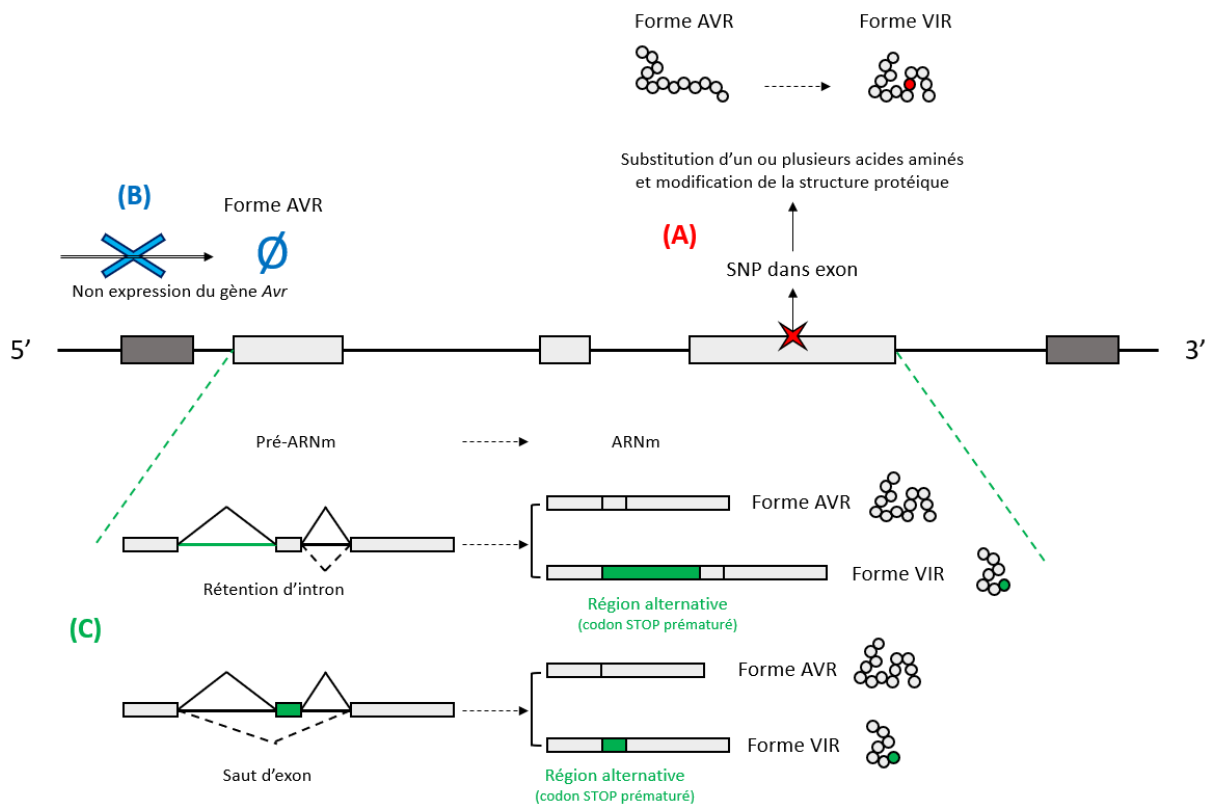
Les clones sont caractérisés par leur génotypes multilocus (Mistral *et al.*, 2021). Les données biologiques sont disponibles au laboratoire : famille de plante hôte colonisé, et capacité à éliciter la résistance au CMV (*i.e.* définition de l'avirulence).

## 2. Hypothèses de travail

Les connaissances sur les effecteurs aphidiens sont limitées à quelques entreprises de caractérisations fonctionnelles incomplètes relativement récentes (Chaudhary et al., 2019; N. Cui et al., 2019; Mutti et al., 2008; Naessens et al., 2015; Wang, Dai, et al., 2015; Wang, Luo, et al., 2015; Y. Zhang, Liu, Francis, et al., 2022; Y. Zhang, Liu, Fu, et al., 2022) et le plus souvent à des prédictions fonctionnelles par recherche d'homologie de séquence (Boulain et al., 2018, 2019). Je retiens de l'ensemble des connaissances produites à ce sujet qu'il n'est ni possible, du fait la variété des fonctions proposées et du manque de connaissances, ni pertinent, dans la mesure où seule une fraction peptidique de l'effecteur d'AVR est sensée être reconnue par une protéine NLR et donc ne permet pas d'affirmer que cet effecteur est fonctionnel, de poser l'hypothèse d'un lien entre la fonction d'un effecteur et le fait qu'il soit reconnu par la protéine VAT.

Pour ces raisons, j'ai choisi de **définir le profil de l'effecteur d'AVR** uniquement sur sa nature **salivaire** (Bos et al., 2010) et du **point de vue de la reconnaissance** par la protéine NLR VAT-1 (Boissot, Thomas, et al., 2016; Boualem et al., 2016; Dangl et al., 2013) sur la base de **cinq hypothèses** : **(i)** la reconnaissance, directe ou indirecte, du facteur aphidien d'avirulence par le gène *Vat* du melon est une **relation de type peptide-peptide** qui n'implique qu'une fraction du facteur moléculaire du puceron (Dangl et al., 2013), **(ii)** la régulation de l'expression du gène aphidien responsable de la reconnaissance par le gène *Vat-1* du melon n'est pas conditionnée par la présence de *Vat-1*, **(iii)** le gène codant pour le peptide d'AVR est **exprimé dans les cellules des glandes salivaires** du puceron et ce peptide est sécrété **par les glandes salivaires** dans la lumière du canal salivaire par un mécanisme d'adressage à la sécrétion, puis injecté dans les cellules végétales (Bos et al., 2010), **(iv)** le peptide d'AVR et le virus transmis par le puceron sont **co-injectés dans le cytoplasme des cellules végétales** où se trouve également la protéine VAT-1 (Boissot, Thomas, et al., 2016; Boualem et al., 2016) et **(v)**, chez un clone VIR, les peptides 'alléliques' produits par le gène d'AVR sont l'expression d'une **combinaison allélique spécifique ne déclenchant pas la résistance *Vat-1***.

Concernant la nature de la transition AVR-VIR, j'émetts l'hypothèse que la perte de reconnaissance de la protéine AVR par VAT-1 résulte d'un **changement à l'échelle de la protéine**, ou de la **délétion du gène** codant pour cette dernière. Au cours de ma thèse, j'ai envisagé **trois mécanismes pour expliquer la transition AVR-VIR**: (i) le **polymorphisme** (SNP) dans la séquence peptidique (Allen et al., 2004; Fudal et al., 2009; Houterman et al., 2009; Luderer et al., 2002; Westerink et al., 2004), (ii) l'**absence** de la protéine d'AVR (L. Cui et al., 2012; Qutob et al., 2009; Tyler & Gijzen, 2014) et (iii) l'**épissage alternatif** (Betz et al., 2016; Göhre et al., 2013; Rao et al., 2019; Ste-Croix et al., 2021). Ces trois cas sont illustrés dans la **figure 9** et documentés plus en détails ci-dessous.



**Figure 9** – Trois hypothèses pour la transition AVR-VIR

**Au centre en gris** : représentation de la structure attendue pour le gène d'AVR autour de la zone codante. Les sections en gris clair représentent les exons, celles en gris foncé représentent les UTR en 5' et en 3'. **(A)** Un polymorphisme ponctuel (SNP) dans la zone codante du gène d'Avr entraîne une substitution d'un ou plusieurs acides aminés et altère la structure de la protéine AVR, perdant ainsi la fraction peptidique impliquée dans la reconnaissance VAT-1. **(B)** La non expression du gène d'Avr entraîne l'absence de la protéine AVR et donc l'absence de la fraction peptidique impliquée dans la reconnaissance par VAT-1. **(C)** l'épissage alternatif dans le gène d'aviorulence génère des isoformes protéiques exempts de la fraction peptidique impliquée dans la reconnaissance par VAT-1.



**Hypothèse 1 : SNP dans la zone codante du gène d'AVR.** L'effecteur aphidien contient au moins un SNP dans la zone codante. Ce SNP est commun aux clones AVR C4 et C9 [Figure 9.A]. De nombreux cas connus dans lesquelles des mutations ponctuelles entraînent des **modifications d'un seul ou plusieurs acides aminés** assurent la transition AVR-VIR (Allen et al., 2004; Fudal et al., 2009; Houterman et al., 2009; Luderer et al., 2002; Westerink et al., 2004).

**Hypothèse 2 : non expression du gène d'AVR.** L'effecteur aphidien n'est **pas exprimé chez les clones VIR** [Figure 9.B] ou le gène est absent dans le génome du clone VIR (L. Cui et al., 2012; Qutob et al., 2009; Tyler & Gijzen, 2014).

**Hypothèse 3 : épissage alternatif (EA) dans le gène d'avirulence.** L'EA est un processus qui intervient lors de l'expression d'un gène **pour créer une diversité protéique** (les protéines issues de l'épissage alternatif d'un gène sont dites **isoformes**). La majorité des cas connus résultent de l'utilisation d'un ou plusieurs des **quatre évènements d'épissage alternatif (EEA) de base** suivant : **(i)** choix du site d'EA 5', **(ii)** choix du site EA 3', **(iii)** saut de cassette d'exon et **(iv)**, la rétention d'intron (Nilsen & Graveley, 2010). Dans la suite de mon manuscrit, je me suis concentré sur les EEA **(iii)** et **(iv)**. Dans le cadre des interactions plante-pathogène, **ce processus pourrait représenter un mécanisme de diversification pour des gènes codant pour des effecteurs au niveau transcriptionnel**. La très grande majorité des cas connus ont été identifiés chez des champignons ou des oomycètes. Toutefois et notamment, chez *Nilaparvata lugens* (Hemiptera: Delphacidae), la famille NI40 de gènes d'effecteurs donne lieu à de nombreuses isoformes (Rao et al., 2019). En revanche, aucun cas d'EA n'a été identifié chez les pucerons. Toutefois, l'effecteur MpC002, identifié chez *M. persicae*, contient cinq répétitions en tandem parfaites constituées de sept acides aminés et des effecteurs contenant des répétitions semblent être sujet à de l'EA (Betz et al., 2016; Bos et al., 2010; Mesarich et al., 2015). **Dans mes travaux**, j'ai étudié la possibilité que l'effecteur aphidien AVR reconnu par la protéine cytoplasmique VAT-1 soit un variant d'épissage spécifique des clones AVR [Figure 9.C]. Un isoforme est candidat à l'avirulence est si : **(i)** le gène qui lui correspond est exprimé chez les 3 clones C4, C6 et C9 et **(ii)**, si cet isoforme est présent chez les deux clones avirulents C4 et C9 et absent chez le clone avirulent C6.

## 3. Matériels et Méthodes

### 3.1 Les clones de pucerons

Les clones spécialisés sur Cucurbitacées sont élevés et maintenus sur des plants de melon Védrantais, les clones spécialisés sur solanacées sont maintenus sur des plants d'aubergine et les clones spécialisées sur malvacées sont maintenus sur des plantes de Gombo. Ces maintiens se font dans une chambre climatique contrôlée à 24:18°C jour/nuit, avec une photopériode de 16:8 h jour/nuit. *A. gossypii* a un mode de reproduction vivipare dans ces conditions, conduisant à des populations clonales.

### 3.2 Construction de la base donnée salivaire

**Récupération des séquences salivaires d'hémiptères.** Les banques de séquences d'EST et de protéines salivaires ont été récupérées sur le site du NCBI (<https://www.ncbi.nlm.nih.gov/>). Au total, 19 sources de données ont été exploitées et concernent 6 hémiptères : *Acyrtosiphon pisum*, *Myzus persicae*, *Macrosiphum euphorbiae*, *Diuraphis noxia*, *Schizaphis graminum* et *Nilaparvata lugens*. Les séquences d'EST et les séquences protéiques strictement redondantes ont été éliminées avec le **script 1** (cf. annexe).

**Prédiction et filtrage des cadres de lecture.** Pour les séquences d'EST, les différents cadres de lecture ont été prédits à l'aide de Transdecoder v5.5.0 et tous les cadres de lecture codants pour des protéines d'au moins 100 acides aminés ont été conservés. Pour chaque EST, le cadre de lecture le plus long a été récupéré avec le **script 2** (cf. annexe). Enfin, ces cadres de lecture ont été traduits avec Biopython v1.70.

### 3.3 Données transcriptomiques/génomiques mobilisées pour ma thèse

**Production de transcriptomes à partir de données RNAseq.** Préalablement à mon arrivée en thèse au GAFL, **des ARN de têtes** de pucerons (300 têtes par clone) ont été extraits par l'équipe Melon en 2013 (Fricaux-Dutartre et al, non publié). Ces ARN ont été séquencés sur une plateforme HiSeq2000, produisant 134, 160, et 179 millions de reads courts pour C4, C6 et C9 respectivement [Table 1]. Après le retrait des adaptateurs et le contrôle de la qualité des données, les librairies contiennent 118, 132 et 146 millions de reads pour C4, C6 et C9 respectivement [Table 1].

**Table 1** - Résultat du séquençage HiSeq2000 des ARN de têtes de pucerons

Clones	AVR/VIR	NGS HiSeq2000	Post filtres de qualité
C4	AVR	146 M reads	121 M reads
C9	AVR	179 M reads	146 M reads
C6	VIR	160 M reads	132 M reads

A partir de ces données, j'ai produit les assemblages *de novo* des transcriptomes de chacun des trois clones. Les assemblages ont été réalisés à l'aide de Trinity v2.6.6 (Grabherr et al., 2013) avec les paramètres par défaut. Pour chacun des trois assemblages produits par Trinity, le cadre de lecture le plus long a été prédit pour chaque transcrit et récupéré à l'aide de getORF (Rice et al., 2000).

**Production des transcriptomes Isoseq.** Les mêmes **ARN de têtes** de pucerons ont été séquencés suivant la **technologie Isoform Sequencing (Isoseq)** développée par Pacific Biosciences (An et al., 2018; Hong et al., 2020) reposant sur l'analyse de reads de grande taille et visant à reconstituer des transcriptomes pleine-longueur. Les données ont été acquises (2020) par la Plateforme de service en génotypage et séquençage à haut débit GENTYANE (Clermont-Ferrand). Une SMRTcell (*i.e.* unité de séquençage) a été utilisée pour séquencer en multiplex les ARN des trois clones de pucerons (C4, C6 et C9) et 691Go de données brutes ont été obtenues. A partir de ces données, j'ai produit, en collaboration avec Vincent Pailler (INRAE Clermont-Ferrand, GDEC-UCA), les transcriptomes pleine-longueur pour chacun des trois clones.

Le pipeline standard Iso-Seq3 (<https://github.com/PacificBiosciences/IsoSeq>) et le software SMRTlink V9.0 ont été utilisés pour produire ces transcriptomes, par des étapes successives de clustering de séquences [Table 2].

**Table 2** – Résumé du traitement des données de séquençage SMRT des ARN de têtes de pucerons

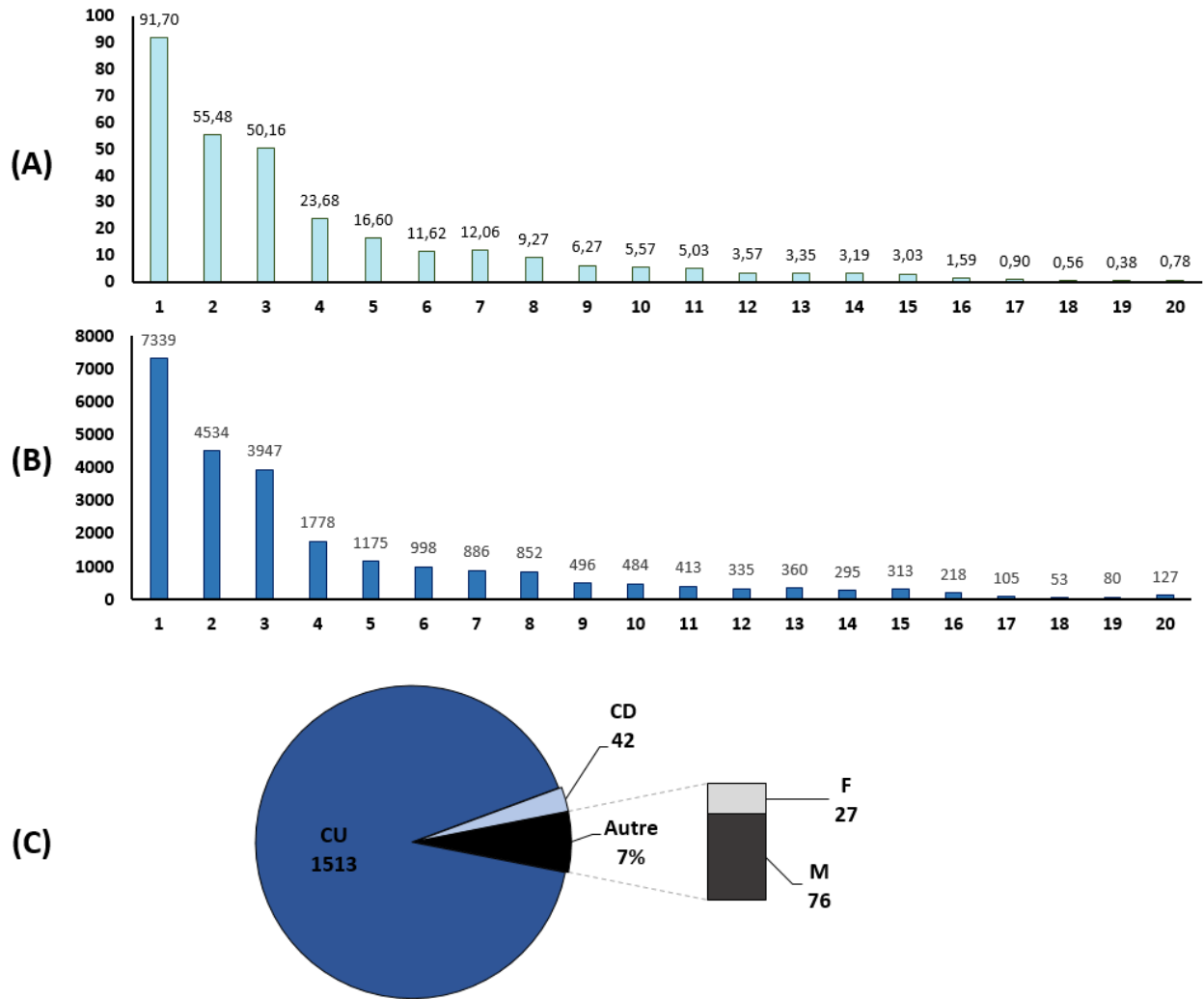
	C4 (AVR)	C9 (AVR)	C6 (VIR)
Séquences consensus circulaires	1 137 379	765 436	1 154 374
Sub-reads complets non chimériques	1 134 769	760 472	1 151 825
Dont avec polyA	1 134 476	760 252	1 151 494
Transcrits pleine-longueur	95 097	70 330	96 533

Au départ, des séquences consensus circulaires (CCS) ont été générées à partir des sub-reads avec la ligne de commande : `ccs --min-passes 3 et --min-rq 0.99`. De cette façon, nous avons obtenu entre 765 000 et 1 154 000 CCS pour les trois clones. Les CCS ont été classés en catégories pleine longueur et non pleine longueur en fonction de la présence des adaptateurs aux extrémités 5'

et 3', en plus des queues poly(A) en 3'. Les sub-reads contenant les adaptateurs et 3' et les queues poly(A) ont été considérés comme des sub-reads complets. Les sub-reads complets non chimériques appelés « complete non-chimeric sequences » (FLNC) ont été générés avec la ligne de commande : `isoseq3 refine --require-polya`. De cette façon, nous avons obtenu entre 760 000 et 1 150 000 et FLNC complets pour C4, C6 et C9 respectivement. Un clustering itératif pour la correction d'erreurs (iterative clustering for error correction ou ICE) a été effectué pour regrouper les FLNC sur la base d'un alignement par paires et d'une affectation réitérative à l'aide de l'algorithme hiérarchique :  $n \cdot \log(n)$ . Finalement, entre 70 000 et 96 000 transcrits pleine-longueur ont été reconstruits pour les trois clones, ils constituent les transcriptomes Isoseq. Les différents cadres de lecture ont été prédits à l'aide de Transdecoder v5.5.0 et tous les cadres de lecture codants pour des protéines putatives d'au moins 100 acides aminés ont été conservés. Les cadres de lecture ont été filtrés sur la base des annotations fournies par Transdecoder pour **(i)** garder ceux qui ont une orientation (+) et **(ii)** qui sont complets, ou **(iii)** partiels en 5'. Ensuite, pour chaque transcrit, le cadre de lecture le plus long a été récupéré avec le **script 2** (cf. annexe). Enfin, ces cadres de lecture ont été traduits avec Biopython v1.70. Pour chaque clone, la redondance stricte des séquences protéiques a été éliminée avec le **script 1** (cf. annexe).

Les transcrits Isoseq 100% identiques sur au moins 90% de la taille du transcrit le plus long ont été regroupées en clusters avec CD-HIT v4.8.1 (paramètres : `-c 1 -s 0.9 -d 0`). Finalement, les transcrits le plus long de chaque cluster ont été conservés. Ils constituent les transcriptomes Isoseq clusterisés.

**Récupération du génome de référence pour le clone C9.** J'ai utilisé un génome du clone C9 produit en 2021 par l'équipe I2B du GAFL (Lagnel et al., 2020) [Figure 10]. L'assemblage finalisé est une séquence génomique de 304,8 Mb scindée en 20 scaffolds [Figure 10.A], avec une taille moyenne de 15,2 Mb, une taille N50 de 50,2 Mb et un total de 24788 gènes codant pour des protéines annotées [Figure 10.B]. Egalement, 95,4 % des gènes attendus chez les insectes ont été découverts dans le génome de C9 [Figure 10.C]. Ensemble, ces métriques soulignent la qualité du génome de référence, à la fois en termes de résolution d'assemblage (*i.e.* à l'échelle pseudo-moléculaire,  $2n=8$ ) et de contenu génétique attendu, que j'ai utilisé dans mes recherches.



**Figure 10** - Principales caractéristiques du génome de C9

X = Les scaffolds 1-20. **(A)** Taille des scaffolds (Mb). **(B)** Nombre de gènes par scaffold codant pour des protéines annotées. **(C)** Évaluations d'exhaustivité BUSCO réalisée avec la version 3.0.2 de BUSCO utilisant la base de données insecta\_odb9 (42 espèces, 1658 domaines BUSCO).

**Description statistique des assemblages.** L'évaluation de la qualité des assemblages en terme de taille des séquences reconstituées a été réalisée avec le script N25\_50\_75\_90\_assembling.pl, mis à disposition sur le serveur de calcul du GAFL par Jacques Lagnel (équipe I2B du GAFL). Les métriques ainsi calculées comprennent : la longueur totale de l'assemblage, le N50, ainsi que le nombre de séquences (gènes/CDS/ transcrits/contigs) dans les assemblages et les indicateurs de dispersion de la taille de ces séquences.

**Analyse de la complétion des assemblages.** L'estimation de l'exhaustivité assemblages en termes de contenu génétique attendu a été réalisée avec le programme **BUSCO** (Benchmarking

Universal Single-Copy Orthologs) (Simão et al., 2015). Un ensemble de 1658 protéines orthologues conservées chez les insectes (lineage database = insecta\_odb9) a été utilisé dans BUSCO v3.0.2 pour évaluer l'exhaustivité du contenu conservé du génome de C9 d'une part (mode génome), et des transcriptomes reconstruits *de novo* à partir de données RNAseq ou de données Isoseq d'autre part (mode transcriptome).

## 3.4 Production de listes de candidats d'AVR in silico

### 3.4.1 Utilisation de données RNAseq

#### Recherche des candidats de non expression

**Mappings et filtrage qualité.** Les bibliothèques paired-end du clone VIR C6 ont été mappées sur les transcriptomes de CDS des clones AVR C4 et C9 à l'aide de Bowtie2 en utilisant le paramétrage par défaut. Les reads ont été filtrés pour une qualité de mapping minimale de 30 avec Samtools v1.8 (commande : `view -h -b -q 30`).

Les tables de comptage pour les deux mappings ont été générées avec Samtools v1.8 (commande : `idxstats`) puis filtrés avec Awk pour identifier les CDS des clones AVR non ou très faiblement couverts par les reads du clone VIR (*i.e.* si le rapport entre le nombre de reads mappés sur un CDS et la taille de ce CDS est au plus égale à 0,001). Les CDS communs aux deux clones AVR C4 et C9 ont été déterminés par des alignements réciproques à l'aide de Blastn v2.6.0+ (paramètres : `perc_id=95` et `perc_qcov=95`).

#### Recherche des candidats de polymorphisme

**Mappings et filtrage qualité.** Les bibliothèques paired-end des clones avirulents (C4 et C9) ont été mappées sur le transcriptome de CDS du clone VIR C6 à l'aide de Bowtie2 en utilisant le paramétrage par défaut. De la même manière que pour la recherche des candidats non exprimés chez C6, les reads ont été filtrés pour une qualité de mapping minimale de 30 avec Samtools v1.8 (commande : `view -h -b -q 30`).

**Appel de SNP et filtrage.** Les deux mappings ont été fusionnés (Samtools `merge @20`) et la probabilité de génotype à chaque position a été calculée avec mpileup. Ensuite, l'appel de SNP a été effectué avec VarScan (Koboldt et al., 2014) en utilisant le paramétrage par défaut pour un appel de variants conservatif, puis les SNP ont été filtrés avec VCFTools (Danecek et al., 2018) pour les paramètres de `DP>5` et `GQ>5`. La prédiction de l'effet des SNP sur la modifications d'acides aminés a

été réalisée avec le SnpEff (Cingolani et al., 2012). Cependant, en l'absence d'un génome de référence, un fichier d'annotation a été généré à partir du transcriptome du clone C6. Les SNP non synonymes détectés ont ensuite été filtrés à partir du fichier VCF tel qu'aucune forme allélique AVR n'existe chez le clone VIR et que les deux clones AVR présentent le même profil allélique.

**Réduction des sets de candidats.** Les sets de candidats totaux ont été réduits à un ensemble défini sur (i) la découverte de signaux de sécrétion à l'aide de programmes de prédiction et (ii) l'absence de signaux associés à une rétention transmembranaire de la protéine. Les peptides signaux ont été prédits avec SignalP v4.1 (Nielsen, 2017) (paramètres : -t euk -u 0.34 -f summary). La prédiction de la localisation subcellulaire des protéines a été réalisée avec Targetp v1.1 (paramètres par défaut) (Emanuelsson et al., 2007) et les domaines insérés dans la membrane ont été prédits à l'aide de TMHMM v2.0 (Emanuelsson et al., 2007).

### 3.4.2 Utilisation de données Isoseq

**Note explicative.** Les transcriptomes Isoseq des trois clones ont été utilisés sous l'hypothèse suivante : pour un gène exprimé, s'il est homozygote on attend un seul transcrit, s'il est hétérozygote on attend deux transcrits.

**Prédiction des cadres de lecture et élimination de la redondance.** Les différents cadres de lecture ont été prédits à l'aide de Transdecoder v5.5.0 et, pour chaque transcrit, le cadre de lecture le plus long a été récupéré avec le **script 2** (cf. annexe). Enfin, ces cadres de lecture ont été traduits avec Biopython v1.70. Les séquences protéiques strictement redondantes ont été éliminées avec le **script 1** (cf. annexe).

**Identification des protéines communes aux clones AVR.** Un alignement BLASTP des séquences protéiques de C9 contre les séquences protéiques de C4 a été réalisé avec noblast (paramètres : -e 0.001 -b 20 -v 20 -m 19) (Lagnel et al., 2009). Les résultats de BLASTP ont été filtrés avec le **script 3** (cf. annexe) : (paramétrage : 1 100 100 100 0) pour récupérer les protéines de C9 dont les séquences sont strictement identiques à une protéine de C4. A cette étape les candidats conservés des clones AVR sont : soit homozygotes dans les deux clones, soit hétérozygotes mais identiques dans les deux clones.

**Identification des protéines spécifiques aux clones AVR.** Les protéines communes aux deux clones AVR ont été alignées contre les séquences protéiques de C6 (VIR) et les résultats de BLASTP

ont été filtrés avec le **script 3** (*cf.* annexe) pour : le meilleur alignement par hit, un alignement réciproque minimum de 90% entre la query et la subject, une identité supérieure ou égale à 90% entre acides aminés du point de vue de leurs propriétés sur la zone alignées, et avec au maximum 1% gaps sur la zone alignée (paramètres d'entrée : [INPUT] 1 90 0 90 1). Selon les résultats d'alignement, deux cas ont été identifiés : **(i)** les protéines de C9 qui n'ont pas d'homologue chez le clone VIR C6 ont été traitées par la suite comme des candidats de type 'non expression' (*i.e.* exprimés chez C4 et C9 et pas chez C6), et **(ii)** les protéines de C9 ayant une forte similarité avec au moins une protéine du clone VIR ont été traitées par la suite comme des candidats 'de type polymorphisme'.

Les protéines de C9 identifiées en **(ii)** constituent l'entrée du traitement ci-après. Les mapping des RNA seq sur une référence soit des Isoseq consensus AVR soit des Isoseq consensus VIR avec Bowtie2 v2.2.3.4.1 (Langmead & Salzberg, 2012) lors de deux mappings indépendants, en utilisant le paramétrage par défaut. Ensuite, les reads ont été filtrées pour une qualité de mapping minimale de 20 avec Samtools v1.9 (commande : `view -h -b -q 20`). L'appel de variants a été effectué à l'aide de Bcftools v1.11 (commande : `call -mv -Ov`). Les variants ont ensuite été filtrés avec awk tel que : aucun des variants présents chez les clones AVR n'existe chez le clone VIR et que l'allèle AVR soit le même chez les deux clones AVR.

### 3.4.3 Intégration du génome de référence pour le clone C9

**Code.** La partie 'Mappings' et 'Appel de SNP' a été réalisée à l'aide du **script 4** (*cf.* annexe). Ce script, développés par Jacques Lagnel et Sonia Elbelt (INRAE GAFL, équipe I2B), est disponible sur le git de l'Unité GAFL : [https://forgemia.inra.fr/gafl/pipelines\\_snakemake\\_rnaseq\\_de\\_snps](https://forgemia.inra.fr/gafl/pipelines_snakemake_rnaseq_de_snps).

## Recherche des candidats de non expression

**Mappings et production d'une table de comptage.** Les librairies paired-end des trois clones C4, C6 et C9 ont été mappées sur le génome de référence de C9 avec STAR v2.7.3 (Dobin et al., 2016). Les tables de comptage pour le mapping ont été générées avec DiCoExpress (Lambert et al., 2020) puis filtrées pour identifier les gènes exprimés par les deux clones AVR C4 et C9 et non exprimés par le clone VIR C6. La table de comptage comprend quatre colonnes (colonne 1 : identifiant des gènes, colonnes 2-4 : comptages de reads pour les clones C4, C6 et C9 respectivement).

**Filtrage de la table de comptage pour le profil des candidats.** Les gènes sélectionnés sont ceux pour lesquels : **(i)** le comptage des reads des clones AVR est supérieur ou égal à 30 et **(ii)** pour lesquels le comptage des reads du clone VIR est inférieur ou égal à 10 et **(iii)** pour lesquels le



comptage des reads du clone VIR est inférieur ou égal à 5% de celui des reads de chaque clone AVR.

### **Recherche des candidats de polymorphisme**

**Appel de SNP et filtrage des SNP bruts.** L'appel de SNP a été effectué avec GATK4 v4.1.4.1 (Mckenna et al., 2010; Van der Auwera et al., 2013). Dans un premier temps, les SNP ont été filtrés en dur comme indiqué dans les documents GATK (<https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants>). Les filtres appliqués pour un appel conservatif sont les suivants : (commande : `gatk VariantFiltration`) `QUAL < 40.0, QD < 4.0, FS > 55.0, MQ < 55.0, SOR > 3.0, MQRankSum < -10.5` et `ReadPosRankSum < -5.0`. Dans un deuxième temps, les SNP ont été filtrés pour `DP>8` et `GQ>20`.

**Sélection du profil des SNP bialléliques pour les candidats.** Seuls les SNP bialléliques ont été conservés (commande : `gatk SelectVariants --restrict-alleles-to BIALLELIC`). Les SNP devenus non variants suite à l'application des filtres précédents ont été retirés (commande : `gatk SelectVariants --exclude-non-variants`). Aussi, seuls les SNP génotypés pour chacun des trois clones ont été conservés par la suite (commande : `gatk SelectVariants --max-ncall-number 0`). Puis, ces SNP ont été filtrés avec le **script 5** (cf. annexe), selon les profils déterminés par la table de vérité logique matérialisant la transition AVR-VIR sous l'hypothèse d'une mutation non synonyme **[Figure 11]**.

**Annotation des SNP.** Les SNP ont ensuite été annotés avec le pipeline Snakemake d'extraction d'allèles AllMine (<https://github.com/tbersez/Allmine>) développé au GAFL par Jacques Lagnel et Thomas Bersez, en utilisant le génome de référence de C9 annotée **[Figure 10]**. Enfin, ces annotations ont été filtrées pour sélectionner les SNP situés dans des zones codantes qui sont et non synonymes.

Composition allélique de C4 (AVR)

	0/0	0/1	1/1
Composition allélique de C9 (AVR)	0/0	1/1	1/1
	0/1	1/1	ou 1/1 0/0
	1/1	-	0/0

Composition allélique de C6 (VIR)  
lorsque la reconnaissance ne peut pas avoir lieu

**Figure 11** – Table de vérité logique déterminant le profil des SNP bialléliques pour les candidats d’AVR

Le génotype est codé, en accord avec le format VCF, sous forme de valeurs d’allèles séparées par "/".

Seuls les SNP bialléliques sont considérés : les valeurs d’allèle sont 0 pour l’allèle de ‘référence’ et 1 pour l’allèle ‘alternatif’ (*i.e.* non présent dans la séquence consensus du génome de référence pour C9).

Trois génotypes sont possibles pour le SNP assurant la transition AVR-VIR avec :

0/0 : homozygote ‘référence’,

0/1 : hétérozygote et

1/1 : homozygote ‘alternatif’.

Les cases vertes donnent la composition allélique des clones AVR : C4 (première ligne) et C9 (première colonne).

Les cases blanches correspondent à la composition allélique du clone VIR C6 dans les cas où la reconnaissance ne peut pas voir lieu.

### **Recherche des candidats d’épissage alternatif**

**Principe et objectif de la démarche.** L’objectif est : **(i)** d’identifier dans le génome de C9 les gènes mappés par au moins un transcrit Isoseq des trois clones, **(ii)** identifier les coordonnées des exons ces gènes et **(iii)** implémenter la table obtenue pour les caractéristiques ‘candidats’ ou ‘non candidats’. Sont déclarés candidats les gènes pour lesquels une combinaison d’exons existe chez C4 et C9 et n’existe pas chez C6 ou dit autrement un exon surnuméraire existe chez C4 et C9 et n’existe pas chez C6 et l’inverse. La recherche des candidats a été réalisée avec le **script 6** (*cf.* annexe) que j’ai édité durant ma thèse. Ce script est disponible sur sur le GitLab communautaire d’INRAE : <https://forgemia.inra.fr/>.

**Mappings et production des coordonnées des régions génomiques.** Les transcriptomes Isoseq clusterisés de chaque clone ont été mappés indépendamment sur le génome de référence à l'aide de Minimap2 (H. Li, 2018) (options : `ax -splice`). Les sorties brutes des mappings ont été converties au format bam (Binary Alignment Map) puis triées par chromosome et par position avec Samtools v1.9. Ces fichiers ont été filtrés pour une qualité de mapping minimale de 20 (*i.e.* chaque transcrit a au moins 99% de chance d'être aligné de manière unique sur le génome) avec Samtools v1.9 (commande : `view -h -b -q 30`). Ensuite, ces fichiers ont été convertis au format BED (Browser Extensible Data), stockant ainsi les coordonnées de début à fin de chaque alignement (*i.e.* une seule entrée par alignement) à l'aide de Bedtools v2.29.2 (Quinlan & Hall, 2010) (commande : `bamtobed`).

**Sélection des gènes mappés par des transcrits de tête.** Pour chaque clone séparément, et après avoir été triés par chromosome puis par position de départ (commande bash : `sort -k1,1 -k2,2n`), les entrées BED (*i.e.* les plages de coordonnées pour chaque alignement) partageant au moins une position ont été fusionnées (*i.e.* réunions d'ensembles) avec Bedtools v2.29.2 (commande : `merge`, options : `-s -c 4,6 -o collapse,distinct`), stockant ainsi les coordonnées de début à fin de la réunion des entrées BED fusionnées, les identifiants des transcrits concernés et l'orientation des alignements sur le génome dans de nouveaux fichiers BED. Ensuite et de la même manière, les fichiers BED fusionnés par clone ont été fusionnés entre eux puis filtrés sur les préfixes des identifiants des transcrits des clones de pucerons, permettant de récupérer les coordonnées des CDS sur le génome de C9 pour lesquelles au moins un transcrit Isoseq de chacun des trois clones est mappé.

**Annotation des exons mappés.** Les fichiers BAM triés et filtrés ont été utilisés de nouveau pour produire les coordonnées des transcrits mappés sur le génome de référence. Cette fois-ci, les fichiers bam ont été convertis au format BED en "mode fractionné" (commande de Bamtools : `bamtobed`, option `-split`), signalant ainsi les coordonnées des exons pour chaque transcrit en tant qu'entrées BED séparées (une entrée BED correspondant à un seul exon). Des champs d'annotations personnalisés ont été ajoutés à chacune de ces entrées BED, renseignant pour chacune d'entre elles : le positionnement relatif de l'exon dans le CDS auquel il correspond, la taille de l'exon et la taille du transcrit auquel il correspond.

**Recherche d'exons déterminants pour l'identification de candidats.** Les fichiers BED produits en "mode fractionné" ont été utilisés pour rechercher des intersections ou des disjonctions entre les plages de coordonnées des exons des différents clones. Dans un premier temps, les exons

communs aux clones AVR ont été identifiés en calculant les intersections entre les plages de coordonnées des exons pour le clone C9 et celles pour le clone C4. Un overlap réciproque minimum de 90% a été fixé pour déclarer une intersection entre deux exons avec la commande : `Bedtools intersect` et les options : `-wa -r 0.9 -a` (nb : le fichier BED du clone C9 doit être donné en premier paramètre d'entrée et celui pour le clone C4 doit être donné en second). En parallèle, les exons communs aux clones AVR et absents chez le clone VIR C6 ont été identifiés en calculant les disjonctions entre les plages de coordonnées des exons communs aux clones AVR précédemment générées et celles pour le clone C6. Une disjonction réciproque minimum de 30% a été fixée pour déclarer que deux exons sont disjoints (commande : `Bedtools intersect` et options : `-v -wa -r 0.3 -a`) (nb : le fichier BED avec les exons communs aux clones AVR doit être donné en premier paramètre d'entrée et celui pour le clone C6 doit être donné en second). De la même façon que dans l'étape précédente, les exons du clone VIR C6 qui ne sont pas communs aux clones AVR ont été identifiés.

**Production de tables de caractéristiques pour C9 et C6.** Les fichiers BED ont été croisés pour récupérer les transcrits avec des exons surnuméraires non bordants. Ils ont été utilisés pour produire deux tables de caractéristiques clés pour l'identification de transcrits candidats avec, soit au moins un exon supplémentaire commun aux clones AVR, soit au moins un exon surnuméraire chez le clone VIR.

**Filtrage des tables de caractéristiques pour le profil des candidats.** Chacune de ces deux tables de caractéristiques ont été filtrées pour extraire les transcrits Isoseq clusterisés mappant sur une zone commune aux trois clones, comportant au moins un exon spécifique aux clones AVR vs VIR et non bordant. Les séquences FASTA de ces transcrits candidats ont été récupérées à partir de leur identifiant avec la commande : `awk -F'>' 'NR==FNR{ids[$0]; next} NF>1{f=($2 in ids)} f'` admettant en premier paramètre d'entrée la liste des identifiants de ces transcrits et en second paramètre les transcriptomes Isoseq clusterisés de C9 vs C6.

**Prédiction des cadres de lecture et recherche de candidats salivaires.** Les différents cadres de lecture ont été prédits à l'aide de `Transdecoder v5.5.0` et, pour chaque transcrit, le cadre de lecture le plus long a été récupéré avec le **script 2** (cf. annexe). Enfin, ces cadres de lecture ont été traduits avec `Biopython v1.70`. Pour identifier les transcrits candidats à l'avorulence putativement salivaires, un alignement BLASTP avec `nblast` (Lagnel et al., 2009) a été réalisé contre la base de données salivaire et les résultats de BLAST ont été filtrés avec le **script 3** (cf. annexe) (paramétrage : 1 90 0 90 1).

### 3.5 Confirmation des gènes candidats en laboratoire

**Extractions d'ARN et d'ADN de pucerons.** Les extractions d'ARN et d'ADN ont été réalisées pour les 12 clones de la collection du GAFL [Figure 8]. Par extraction, 30 à 50 individus adultes prélevés vivants et congelés dans l'azote liquide ont été utilisés. L'ARN total a été extrait à l'aide d'un Kit RNeasy Plant mini kit (Qiagen ref 74904) conformément aux instructions du fabricant. L'ADN génomique à l'aide d'un Kit DNeasy Plant mini kit (Qiagen ref 69104). L'ARN et l'ADN extraits ont été quantifiés en mesurant l'absorbance à 260 nm sur un spectrophotomètre Nanodrop (Thermo Fisher Scientific, Inc.).

**Rétrotranscription.** La RT a été réalisée sur 1.5 µg d'ARN purifié avec 10 µM d'oligoT, 4 mM de chaque dNTP et 10 U d'AMV Reverse Transcriptase, selon les instructions du fabricant (Promega). Les amorces ont été définies avec Primer3 (Untergasser et al., 2012). Les séquences des amorces, amplifiant les candidats à l'avirulence identifiés au cours de ma thèse sont disponibles sur le serveur du GAFL. L'amplification des amplicons à partir des cDNA a été réalisée par une réaction de polymérisation en chaîne (PCR) dans un volume final de 25 µL dans un thermocycleur (Mastercycler, Eppendorf). La composition du mix PCR utilisé est la suivante : 2.5 µL de TAKARA dNTPmix, 2.5 µL de buffer 10X TAKARA, 1 unités d'ADN polymérase, 5 µM de chaque amorce, 2 µL de cDNA, qsp 25 µL avec H<sub>2</sub>O RNase free.

**PCR.** La PCR a été réalisée selon un protocole standard avec les amorces à une concentration de 0.2 nM, une dénaturation initiale de 4 min à 94°C suivi de 35 cycles comprenant une dénaturation pendant 30s à 94°C, une hybridation pendant 30 s à 50°C<T<sub>m</sub>< 60°C (*i.e.* selon les caractéristiques des amorces), et une extension de 1min/kb à 72°C, puis d'une extension finale pendant 10 min à 72°C. Les produits PCR ont été caractérisés par électrophorèse sur gel d'agarose à 1 ou 2% selon leurs tailles prévues. Pour les candidats attendus en polymorphisme de type SNP, les produits de PCR ont été conservés à -20°C jusqu'au séquençage (méthode Sanger) par la société GenoScreen. La lecture des chromatogrammes a été réalisée par Vincent Rittener (GAFL).

## 4. Résultats

### 4.1. Construction d'une base de données de protéines salivaires d'hémiptères

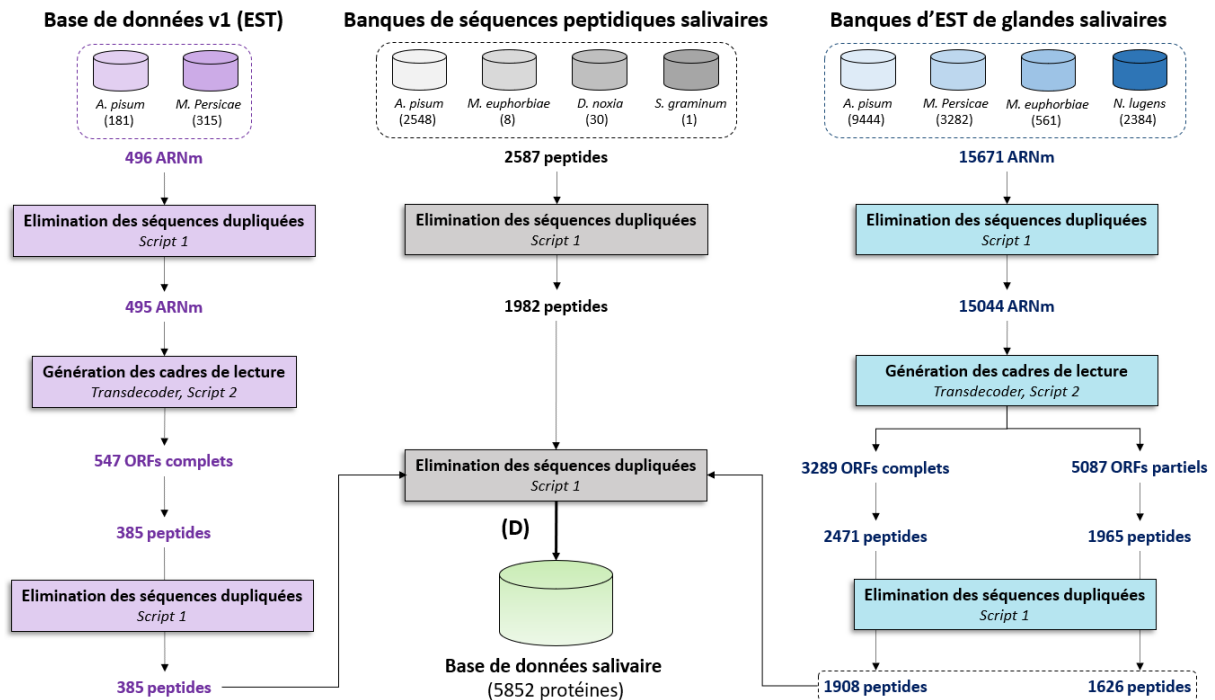
**A l'origine de la base de données salivaires.** De nombreuses bases de données ont été construites sur la base de **protéines identifiées dans la salive d'hémiptères** ou d'**ARNm identifiés dans leurs glandes salivaires**. J'ai créé une base de données regroupant ces deux types de données **pour 6 hémiptères**. La **première version** de cette base de données a été construite **en 2014** par Leslie Dutarte-Fiscaux (postdoctorante INRAE) et comportait **385 protéines non redondantes** exprimées dans les glandes salivaires chez *Myzus persicae* (315 protéines) (Bos et al., 2010; Harmel et al., 2008; John S. Ramsey et al., 2007) et *Acyrtosiphon pisum* (181 protéines) (Bos et al., 2010; Carolan et al., 2009, 2011). **J'ai implémenté cette première version au cours de ma thèse.**

**Implémentation de la base de données salivaires.** J'ai exploité les bases de données publiques de **marqueurs de séquence exprimés (EST) des glandes salivaires** de pucerons et de **protéines issues de la salive** chez *A. pisum* (Boulain et al., 2018; Mutti et al., 2008; Sabater-Muñoz et al., 2006; Wang, Dai, et al., 2015; Wang, Luo, et al., 2015), *M. persicae*, le puceron vert du pêcher (Naessens et al., 2015; Rodriguez et al., 2014, 2017), *M. euphorbiae*, le puceron vert et rose de la pomme de terre (Atamian et al., 2013), *D. noxia*, le puceron russe du blé (Nicholson et al., 2015), *S. graminum*, le puceron vert des graminées (Y. Zhang et al., 2015) et la cicadelle brune *N. lugens* (Ji et al., 2017; Noda et al., 2008). La **version implémentée** de la base de données salivaire, comme détaillée dans la **figure 12**, contient un total de **5852 peptides non redondants**, avec une taille moyenne de 304 AA et une taille médiane de 208 AA. J'ai contribué à un **enrichissement important** (x15,2) en termes de nombre de séquences peptidiques ainsi qu'en espèces représentées avec **l'ajout de trois espèces de pucerons** (*i.e.* *Macrosiphum euphorbiae*, *Diuraphis noxia* et *Schizaphis graminum*) et avec **l'élargissement à une autre famille d'hémiptères**, les *Delphacidae*, avec *Nilaparvata lugens*. Quarante-huit protéines de la première version (*i.e.* 12,5% de la première version) ont été retrouvées dans les données ajoutées et 5 467 nouvelles protéines ont été ajoutées. Notamment, la première version de la base de données salivaire incluait deux effecteurs aphidiens **Mp1** chez *M. persicae* (Rodriguez et al., 2014, 2017) et **Armet** chez *A. pisum* (Wang, Dai, et al., 2015) et son enrichissement a permis d'intégrer **six effecteurs additionnels bien étudiés** : **MIF1** chez *A. pisum* et *M. persicae* (Naessens et al., 2015), **Mp10** chez *M. persicae* (Bos et al., 2010), **Me47**, identifié chez *M. persicae* et *M. euphorbiae* (Kettles & Kaloshian, 2016), **NIEG1** chez *N. lugens* (Ji et al., 2017), **Me10** chez *M.*

*euphorbiae* (Atamian et al., 2013) et **SgC002** chez *S. graminum* (Y. Zhang et al., 2015). J'ai utilisé cette base de données pour réduire un set de gènes candidats établi sur la base d'une analyse comparative de données d'expression de clones VIR et AVR.

**Accessibilité.** La base de données est actuellement disponible sur le serveur du GAFL :

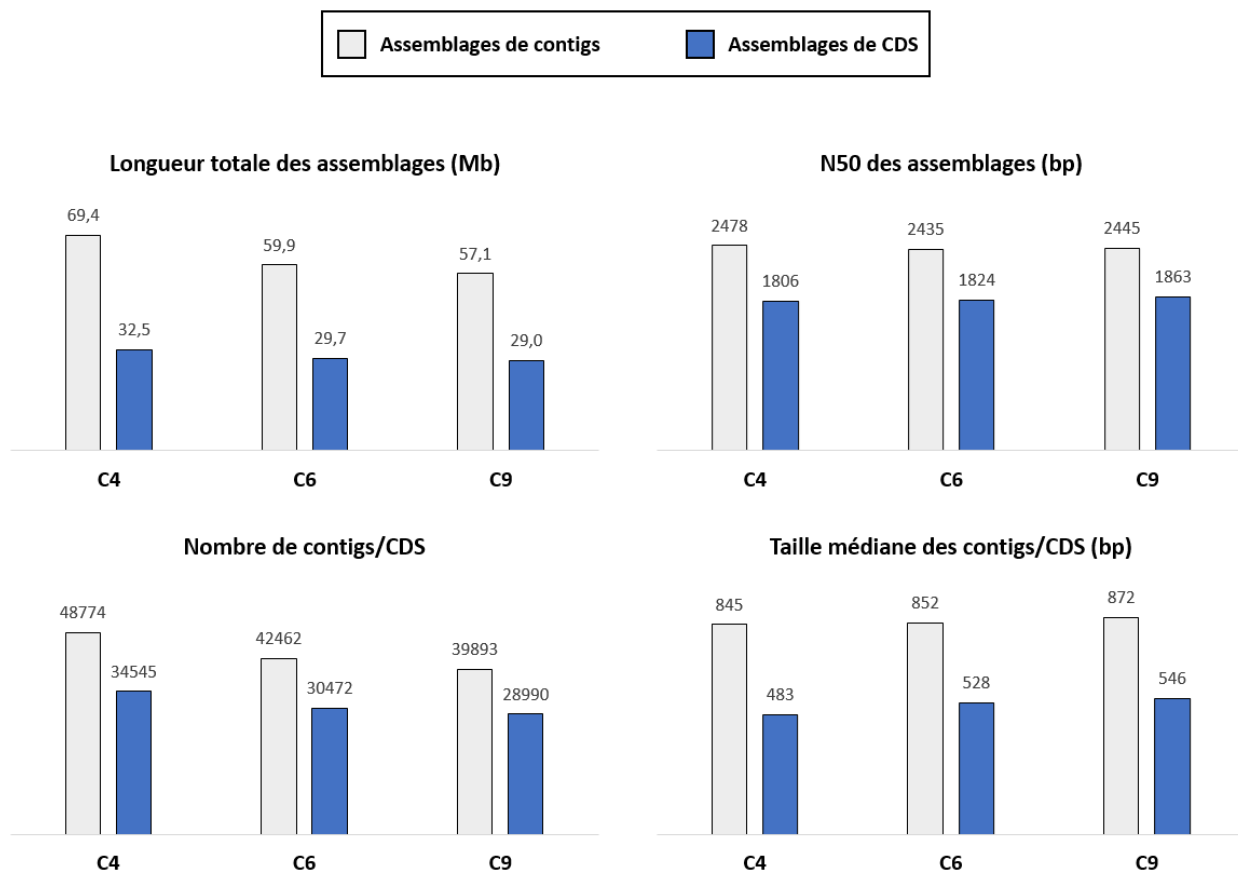
L:\Especes\Melon\APHIS\Aphis general\3-Avirulence gossypii.



**Figure 12** – Construction de la base de données de protéines salivaires d'hémiptères

## 4.2 Génération de transcriptomes *de novo* de têtes de pucerons

**Transcriptomes de têtes assemblés à partir de données RNAseq.** J'ai reconstruit des transcriptomes **de têtes** de pucerons pour les clones C4, C6 et C9 à partir des données de séquençage haut débit (NGS Hiseq2000) acquises [Table 1] entre 2013 et 2014 au GAFL (Fricaux-dutartre et al., 2014). Les caractéristiques relatives à la taille des séquences qui composent les transcriptomes sont présentées dans la **figure 13**.

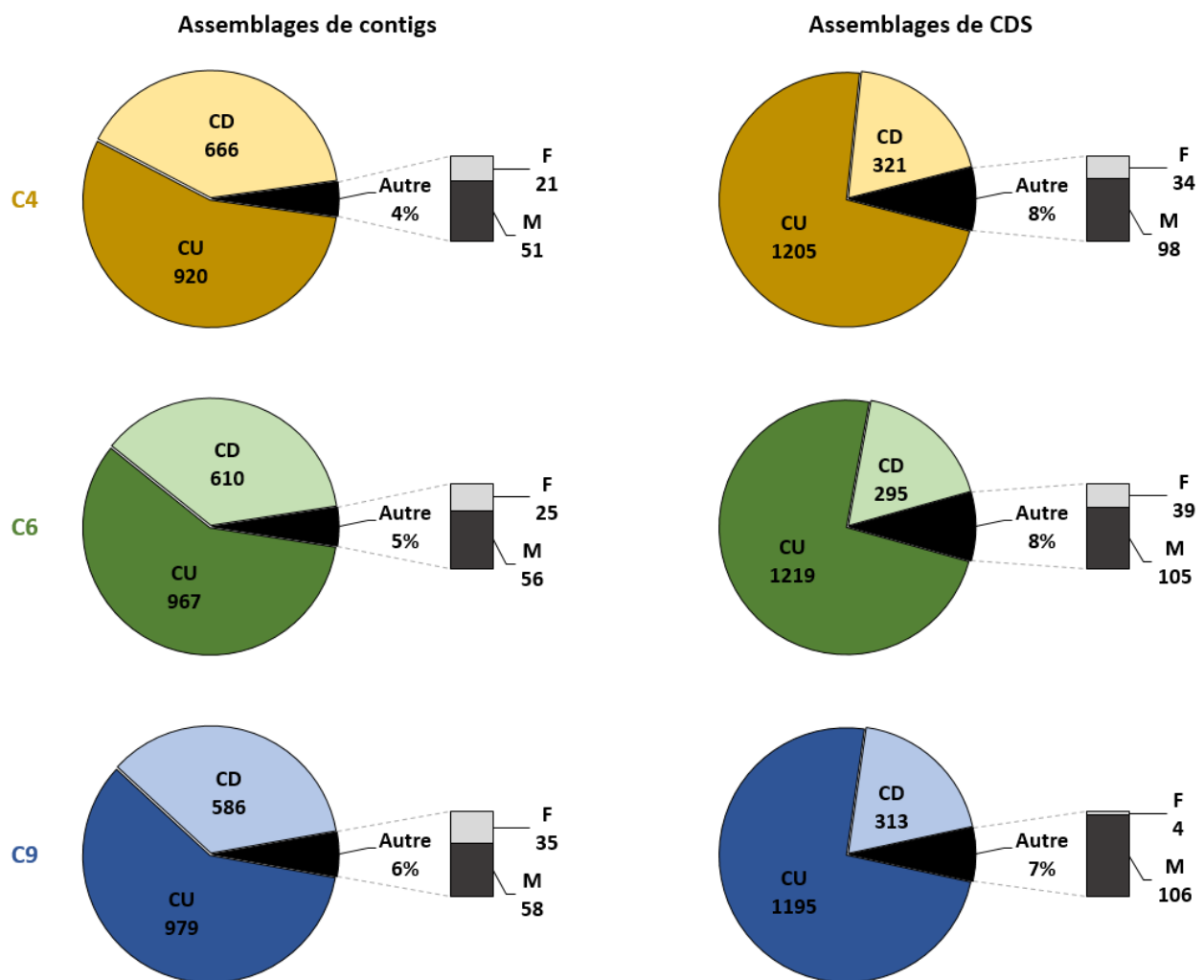


**Figure 13** - Principales caractéristiques des transcriptomes de têtes de pucerons assemblés à partir de données RNA-seq

En moyenne par clone, une séquence totale de 62 Mb a été reconstruite [Figure 13] à partir de 130 M de reads. Autour de 43000 contigs ont été reconstruits pour chacun des trois clones, avec une longueur médiane de 850 bp [Figure 13]. Tous les contigs de longueur supérieure à 100 bp ont ensuite été utilisés pour la production de CDS. Autour de 31000 CDS ont été prédits pour chaque clone, avec une longueur médiane de 520 bp et un N50 de 1831 [Figure 13]. Parmi ces CDS prédits, environ 65 (0,2%) sont potentiellement des artéfacts d'assemblage, avec une longueur supérieure à 10 Kb.



J'ai effectué une analyse BUSCO (*i.e.* Benchmarking Universal Single- Copy Orthologs, BUSCO est un outil de référence pour évaluer le niveau de complétion de l'assemblage d'un génome ou d'un transcriptome en recherchant des orthologues à copie unique attendues hautement conservés parmi des espèces étroitement apparentées) de l'ensemble de gènes d'insectes conservés à la fois pour les 43000 contigs ainsi que les 31000 CDS. L'analyse de la complétude des transcriptomes avec BUSCO est présentée dans la **figure 14**.



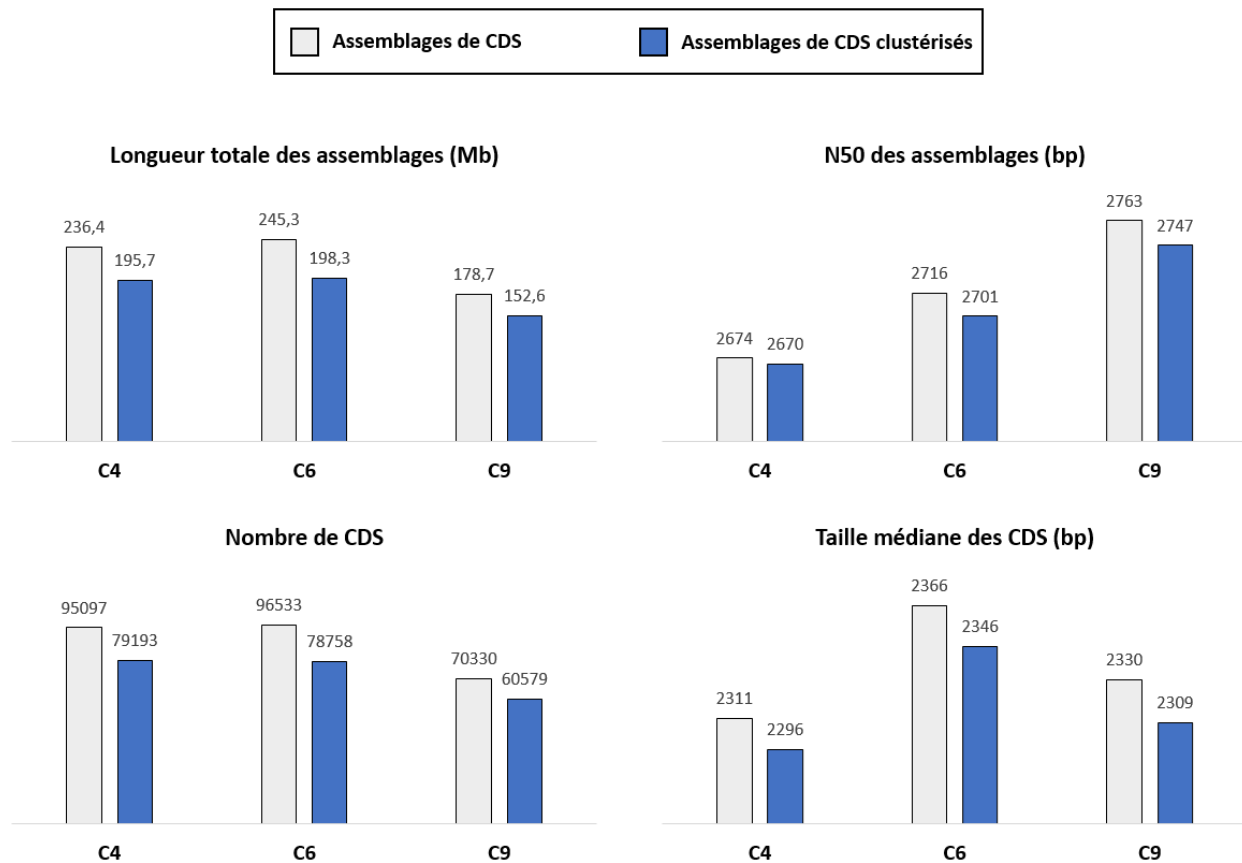
**Figure 14** - Analyse BUSCO des transcriptomes de têtes de pucerons assemblés à partir de données RNA-seq

**Au gauche** : assemblages bruts. **A droite** : assemblages de CDS. En **orange** : transcriptomes de C4, En **vert** : transcriptomes de C6, En **bleu** : transcriptomes de C9. **CU** : domaines BUSCO Complètes et Uniques, **CD** : domaines BUSCO Complètes et Dupliqués, **F** : domaines BUSCO Fragmentés, **M** : domaines BUSCO Manquants. Evaluation d'exhaustivité réalisée avec la version 3.0.2 de BUSCO utilisant la base de données insecta\_odb9 (42 espèces, 1658 domaines BUSCO attendus).

Sur les 43000 contigs prédits [Figure 14, à gauche], en moyenne, BUSCO a identifié 955 gènes complets à copie unique (58% des 1658 gènes d'insectes conservés attendus) et 621 doublons (37%). 27 gènes supplémentaires ont été notés comme fragmentés (1,6%) et 55 étaient manquants (3,3%). L'observation de ces gènes manquants est probablement liée au fait que nous avons travaillé sur des transcriptomes de tête de pucerons.

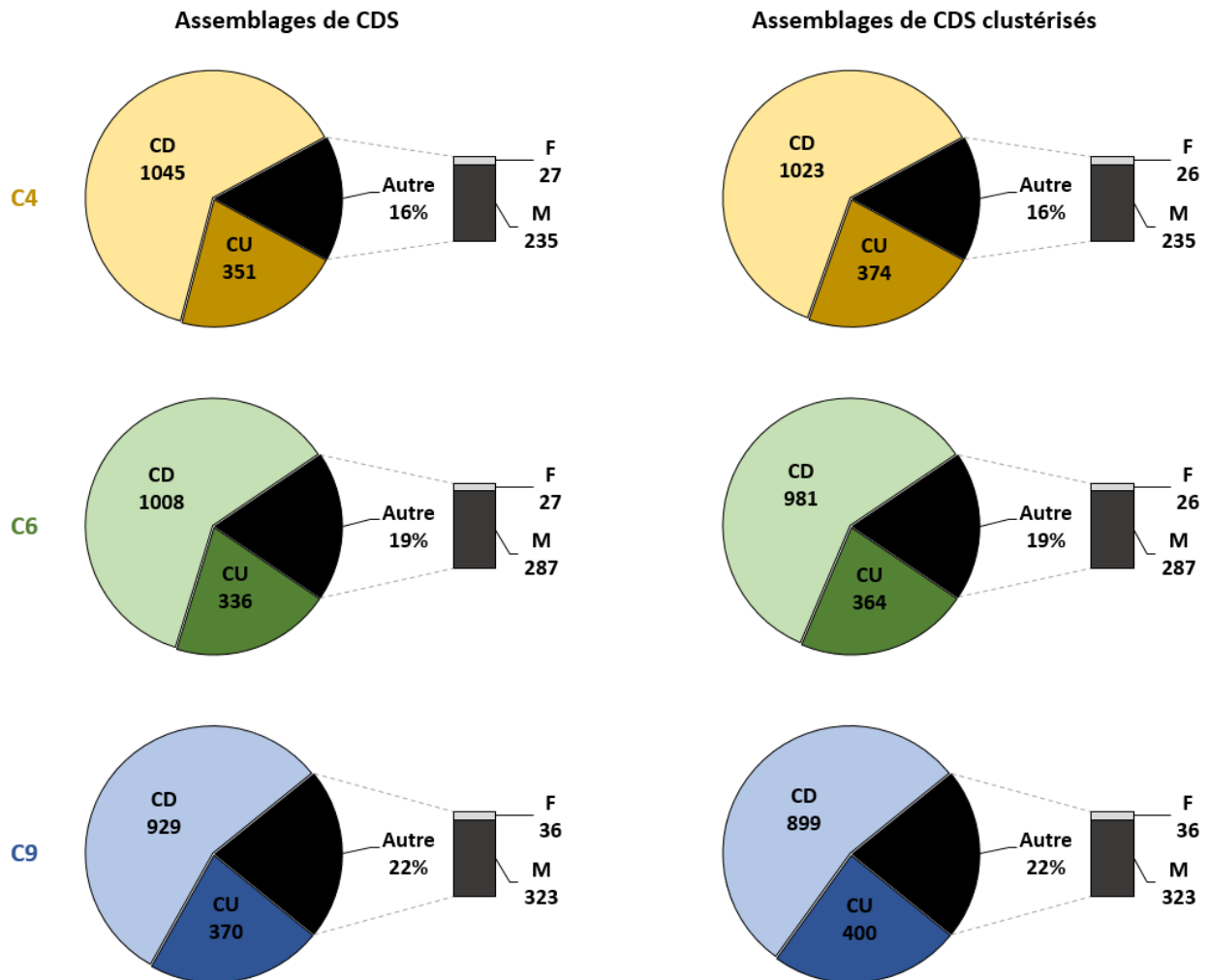
Sur les 31000 CDS prédits [Figure 14, à droite], BUSCO a identifié 1206 gènes complets à copie unique (73% des 1658 gènes d'insectes conservés attendus) et 310 doublons (19%). 26 gènes supplémentaires ont été notés comme fragmentés (1,5%) et 103 étaient manquants (6,2%). En résumé, l'analyse BUSCO indique un niveau de complétude global élevé des transcriptomes de têtes de pucerons. Cependant, la part de doublons est importante et représente en moyenne 19% dans les transcriptomes de CDS alors qu'une faible proportion de cas de duplications est attendue, les domaines BUSCO évoluant sous le contrôle d'une seule copie (Simão et al., 2015; Waterhouse et al., 2011). En conséquence, un nombre important de cas de duplications suggère l'existence d'erreurs dans les assemblages au niveau des haplotypes. Le manque de résolution des assemblages au niveau des haplotypes pourrait être une contrainte dans la recherche de candidats porteur de SNP [Figure 14].

**Transcriptomes de têtes assemblés à partir de données Isoseq.** J'ai assemblé des transcriptomes pleine-longueur de têtes pour les clones C4, C6 et C9, en collaboration avec Vincent Pailler (INRAE GDEC-UCA, Clermont-Ferrand), à partir des transcrits pleine-longueur [Table 2] produits avec la technologie de séquençage Isoseq (Isoform Sequencing), développée par Pacific Biosciences et reposant sur l'analyse de reads de grande taille. Ces données ont été acquises et traitées en 2020 par la Plateforme de service en génotypage et séquençage à haut débit GENTYANE (Clermont-Ferrand) à partir des mêmes **ARN de têtes** que ceux séquencés en short reads. Les caractéristiques relatives à la taille des séquences qui composent ces transcriptomes sont présentées dans la **figure 15**.



**Figure 15** - Principales caractéristiques des transcriptomes de têtes de pucerons assemblés à partir de données Isoseq

En moyenne par clone, une séquence totale de 200 Mb a été reconstruite [Figure 15] à partir de 1 M de séquences consensus circulaires (CCS, *i.e.* le produit du séquençage avec la technologie PacBio Isoseq) [Table 2]. Avec le pipeline Isoseq3 (<https://github.com/PacificBiosciences/IsoSeq>), autour de 87000 transcrits ont été reconstruits pour chacun des trois clones, avec une longueur médiane de 2335 bp [Figure 15]. Autour de 72800 transcrits non redondants (-16% par rapport aux transcriptomes issus du pipeline Isoseq3) ont été produits pour chaque clone avec longueur médiane et un N50 faiblement impactés par l'élimination de la redondance des séquences, respectivement de 2,3 Kb de 2,7 Kb [Figure 15]. J'ai effectué une analyse BUSCO de l'ensemble de gènes d'insectes conservés à la fois pour les 87000 transcrits issus du pipeline Isoseq3, ainsi que les 72800 transcrits non redondants. L'analyse de la complétude des transcriptomes avec BUSCO est présentée dans la figure 16.



**Figure 16** - Analyse BUSCO des transcriptomes de têtes de pucerons assemblés à partir de données Isoseq

**Au gauche** : assemblages bruts. **A droite** : assemblages clusterisés. En **orange** : transcriptomes de C4, En **vert** : transcriptomes de C6, En **bleu** : transcriptomes de C9. **CU** : domaines BUSCO Complets et Uniques, **CD** : domaines BUSCO Complets et Dupliqués, **F** : domaines BUSCO Fragmentés, **M** : domaines BUSCO Manquants. Evaluation d'exhaustivité réalisée avec la version 3.0.2 de BUSCO utilisant la base de données insecta\_odb9 (42 espèces, 1658 domaines BUSCO).

Sur les 87K transcrits issus du pipeline Isoseq3 [Figure 16, à gauche], en moyenne, BUSCO a identifié 352 gènes complets à copie unique (21% des 1658 gènes d'insectes conservés attendus) et 994 doublons (60%). 30 gènes supplémentaires ont été notés comme fragmentés (2%) et 282 étaient manquants (17%). Sur les 72,8 K transcrits non redondants [Figure 16, à droite], BUSCO a identifié 379 gènes complets à copie unique (23% des 1658 gènes d'insectes conservés attendus) et 968 doublons (58%). 29 gènes supplémentaires ont été notés comme fragmentés (2%) et 282 étaient manquants (17%). En résumé, plus de 80% des domaines BUSCO attendus ont été retrouvés dans les transcriptomes assemblés avec les données Isoseq, indiquant un niveau de complétude global élevé

des transcriptomes de têtes de pucerons. Cependant, la part de doublons est importante et représente en moyenne 59% dans les transcriptomes de CDS alors qu'une faible proportion de cas de duplications est attendue, les domaines BUSCO évoluant sous le contrôle d'une seule copie (Simão et al., 2015; Waterhouse et al., 2011). En conséquence, un nombre important de cas de duplications suggère l'existence d'erreurs dans les assemblages au niveau des haplotypes.

### 4.3 Transcriptomes vs base de données salivaires

Quelques effecteurs aphidiens notables sont relativement bien décrits à ce jour et ont fait l'objet d'une caractérisation fonctionnelle: **C002** chez *A. pisum* sur *V. faba* (Mutti et al., 2008), **Armet** chez *A. pisum* sur *V. faba* (Wang, Dai, et al., 2015), **ACE1/ACE2** chez *A. pisum* sur *V. faba* (Wang, Luo, et al., 2015), **MIF1** chez *A. pisum* et *M. persicae* (Naessens et al., 2015) et, **Me10** chez *M. euphorbiae* (Atamian et al., 2013), **Me47** chez *M. persicae* et *M. euphorbiae*, respectivement sur *N. benthamiana* et *A. thaliana*, et *S. lycopersicum* (Kettles & Kaloshian, 2016), **Mp1** chez *M. persicae* (Rodriguez et al., 2014, 2017) et **Mp10** chez *M. persicae* sur *N. benthamiana* (Deshoux et al., 2022; Rodriguez et al., 2014). Ces effecteurs sont présentés de manière détaillée dans la **partie 3.3 de l'introduction générale**. J'ai recherché ces huit effecteurs aphidiens ainsi que **NIEG1** chez *N. lugens* (Ji et al., 2017) par BLASTP, à partir des jeux de données d'expression, que j'ai utilisés pour rechercher des candidats d'AVR : **(i)** les protéines issues des CDS prédits à partir des transcriptomes de novo RNAseq et **(ii)**, dans les transcriptomes Isoseq clustérisés mais aussi **(iii)**, dans les listes de candidats produites. Ainsi, l'ensemble de ces huit effecteurs a été retrouvé dans les deux jeux de données de base, et pour les trois clones C4, C9 et C6. Ceci nous permet de confirmer que nos jeux de données comprennent bien **des ARN produits dans les glandes salivaires**.

## 4.4 Production de listes de candidats d'AVR *in silico*

Dans cette partie, j'ai mis en place trois stratégies en fonction de la date de mise à disposition des données [Figure 17].

Hypothèses pour la transition Avirulence-Virulence			
	Absence du transcrit codant pour la protéine AVR	Polymorphisme non synonyme dans le gène AVR	Evènement d'épissage alternatif spécifique dans le gène AVR
Stratégies : utilisation de données de séquençage	<b>RNAseq</b> Données d'entrée : 28-38 M de CDS Identification : 1514 candidats Dont testés : 73 Confirmation : 0/47 Figure 18	Données d'entrée : 28-38 M de CDS Identification : 40 candidats Dont testés : 3 Confirmation : 0/3 Figure 19	Non traitée
	<b>RNAseq Isoseq</b> Données d'entrée : 70-96 M de CDS Identification : 47 candidats Dont testés : 47 Confirmation : 0/47 Figure 20	Données d'entrée : 70-96 M de CDS Identification : 2 candidats Dont testés : 2 Confirmation : 0/2 Figures 20-21	Non traitée
	<b>RNAseq Isoseq Génome</b> Données d'entrée : 24000 gènes Identification : 10 candidats Dont testés : 10 Confirmation : 0/10 Figure 23	Données d'entrée : 24000 gènes Identification : 7 candidats Dont testés : 7 Confirmation : 0/7 Figure 23	Données d'entrée : 24000 gènes Identification : 324 candidats Dont testés : 29 Confirmation : 0/29 Figure 25

Figure 17 - Récapitulatif des stratégies de production de listes de candidats d'AVR

Trois hypothèses biologiques pour la transition AVR-VIR sont envisagées (de gauche à droite) : le gène d'AVR est uniquement exprimé chez les clones AVR, un SNP spécifique des clones AVR est responsable de la reconnaissance un évènement d'épissage alternatif spécifique des clones AVR est responsable de la reconnaissance. Les références des figures détaillant les différentes stratégies présentées ici sont données dans le coin inférieur droit de chaque case du tableau.

**Première stratégie.** Cette stratégie [Figure 17, première ligne] repose exclusivement sur l'utilisation des données de RNAseq de têtes de pucerons [Table 1] disponibles dès le début de ma thèse (2018).

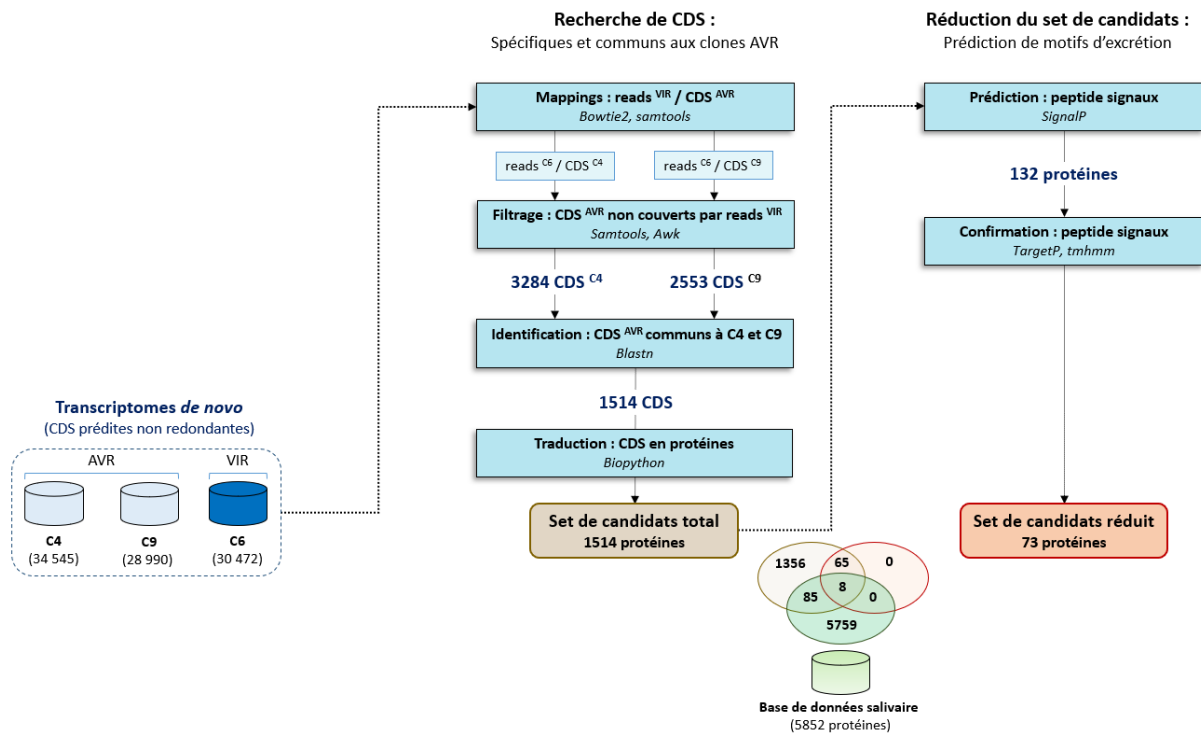
**Seconde stratégie.** Cette stratégie [Figure 17, deuxième ligne] combine l'utilisation des données de RNAseq de têtes de pucerons [Table 1] et des données d'Isoseq produites à partir des même ARN de têtes [Table 2], disponibles en 2020. J'ai recherché des transcrits candidats à l'AVR sous l'hypothèse que la transition AVR-VIR est gouvernée par : (i) l'absence du transcrit codant pour la protéine d'AVR [Figure 9A] ou (ii), par un polymorphisme dans le gène d'AVR induisant un changement à l'échelle de la protéine et concerne un allèle d'AVR commun aux deux clones AVR [Figure 9B].

**Troisième stratégie.** Cette stratégie [Figure 17, troisième ligne] intègre, en plus des données évoquées ci-dessus, le génome de référence annoté pour le clone C9 [Figure 10], disponibles en 2022. L'ensemble des matériels disponibles m'a permis de raisonner au niveau des gènes (là où les données disponibles lors de la mise en place des deux premières stratégies ne permettaient uniquement de raisonner à l'échelle des transcrits). J'ai ici considéré de nouveau les deux hypothèses pour transition AVR-VIR données plus haut ainsi qu'en figure 9A et B. Egalement, j'ai recherché des gènes candidats à l'AVR sous une troisième et nouvelle hypothèse, où la transition est gouvernée par un évènement d'épissage alternatif dans le gène d'AVR [Figure 9C].

**Désignation des types de candidats.** Par la suite, je ferai référence aux candidats d'AVR recherchés sous l'hypothèse que la transition AVR-VIR est gouvernée par (i), l'absence du transcrit codant pour la protéine d'AVR comme des candidats 'de non expression' ou non exprimés, (ii) un polymorphisme dans le gène d'AVR induisant un changement à l'échelle de la protéine comme des candidats 'de polymorphisme' ou porteurs de SNP et (iii), un évènement d'épissage alternatif dans le gène d'AVR comme des candidats 'd'épissage alternatif'.

#### 4.4.1 Stratégie 1 : utilisation des données RNA-seq

**Recherche de transcrits non exprimés du clone VIR.** J'ai recherché des transcrits candidats à l'AVR sous l'hypothèse que la transition AVR-VIR est gouvernée par l'absence du transcrit codant pour la protéine d'AVR. Cette stratégie, présentée en figure 18, est basée sur le mapping des reads du clone VIR C6 sur les CDS prédites des transcriptomes pour chacun des deux clones AVR C4 et C9, indépendamment, puis les CDS non mappées sont recherchés.



**Figure 18** – Recherche de candidats ‘de non expression’ partir de données RNAseq de têtes

Environ 75% des reads de C6 sont mappés sur les CDS des transcriptomes des deux clones AVR. Les deux mappings ont été filtrés pour sélectionner les CDS des clones AVR non couverts par les lectures de C6. De cette façon, 3284 CDS de C4 et 2553 CDS de C9 ont été identifiées et la recherche d’homologie de séquence par BLAST a révélé 1514 CDS communs et spécifiques aux deux clones AVR [Figure 18]. Ces CDS, spécifiques et communes aux deux clones AVR, ont été converties dans le cadre de lecture +1 (*i.e.* lecture des triplets débutant au premier nucléotide) en 1514 protéines candidates (*i.e.* appelé « set total » dans la figure 18) dont, 85 (5,6% du set total) ont été identifiés dans la version de la base de données salivaires que j’ai implémentée (5852 protéines), disponible en 2020 [Figure 12] avec une recherche d’homologie stringente avec le script 3 (*cf.* annexe) (le meilleur alignement par hit, un alignement réciproque minimum de 90% entre la query et la subject, une identité supérieure ou égale à 90% entre acides aminés du point de vue de leurs propriétés sur la zone alignée, et avec au maximum 1% gaps sur la zone alignée).

En 2018, j’ai utilisé des programmes de prédiction de motifs de sécrétion courants afin de réduire le set de candidats total à un ensemble prioritaire défini sur (i) la découverte de signaux de sécrétion attendus dans les bases de données associées aux programmes de prédiction et (ii) l’absence de signaux associés à une rétention transmembranaire de la protéine. De cette manière, j’ai identifié 132 protéines candidates parmi les 1514 pour lesquelles des peptides signaux ont été prédits, dont un peu plus de la moitié (73 protéines, 55% du set total) étaient également prédites dans des réseaux

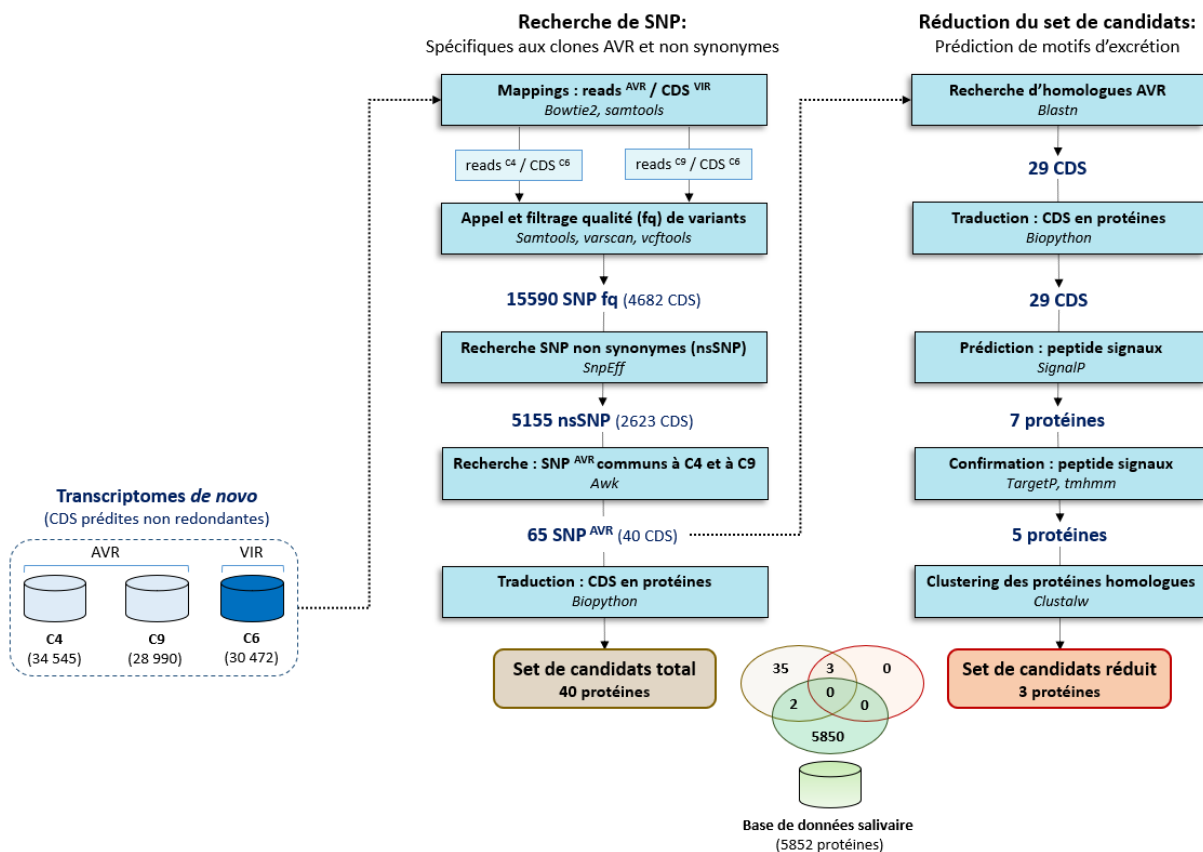


métaboliques liés à des mécanismes de sécrétion et pour lesquelles aucun motif de rétention transmembranaire n'a été identifié.

**Amplification des candidats 'de non expression' sur la gamme de 12 clones.** Pour chacun des **73 transcrits candidats AVR attendus en non expression**, des amplicons de 300-700 bp inclus dans leurs CDS respectifs, ont été amplifiés par **RT-PCR à partir d'extractions d'ARN des 12 clones** de la collection du GAFL **[Figure 8]** pour lesquels le phénotype AVR/VIR a clairement été identifié. L'examen des produits de PCR (*i.e.* une absence d'amplification est attendue chez les clones VIR alors qu'une amplification est attendue chez les clones AVR), ainsi que leurs tailles, par **électrophorèse sur gel d'agarose n'a permis de confirmer le profil AVR/VIR attendu d'aucun de ces 47 candidats.**

**Recherche candidats porteur de SNP non synonymes spécifiques des clones AVR.** J'ai établi une stratégie de recherche des transcrits candidats à l'AVR **sous l'hypothèse que la transition AVR-VIR est gouvernée par un polymorphisme** dans le gène d'AVR induisant un changement dans la protéine et que l'allèle d'AVR est commun aux deux clones AVR.

Cette stratégie, présentée en **figure 19**, est basée sur le mapping des reads des clones AVR sur le transcriptome du clone VIR C6 pour identifier les SNP communs aux deux clones AVR, C4 et C9.



**Figure 19** - Recherche de candidats 'de polymorphisme' partir de données RNAseq de têtes

Environ 75% des reads de C4 et des reads de C9 sont mappés sur les CDS du transcriptome du clone VIR C6. Les deux mappings ont ensuite été fusionnés. L'appel de SNP a mis en évidence un total de 15590 positions sur les CDS du transcriptome de C6, pour lesquelles un variant existait chez au moins l'un des deux clones AVR, concernant 4682 CDS (15% des CDS du transcriptome du clone C6) [Figure 19]. Parmi ces 15590 SNP, un tiers 5155 SNP sont non synonymes et correspondent à 2623 CDS (soit environ 9% du transcriptome de CDS de C6) [figure 19].

Soixante-cinq SNP (0,4% des SNP filtrés) étaient communs aux deux clones AVR, correspondant à 40 CDS de C6 [Figure 19]. La recherche d'homologie réciproque de ces 40 CDS par BLAST avec des CDS des deux clones AVR a révélé que seulement 29 CDS étaient communs [Figure 19]. Ces 29 CDS ont été traduites en 29 protéines candidates (*i.e.* appelé « set total » dans la figure 19), dont deux (5% du set total) ont été identifiés dans la base de données salivaire (5852 protéines), disponible en 2020 [Figure 8] avec une recherche d'homologie stringente avec le script 3 (*cf.* annexe) (le meilleur alignement par hit, un alignement réciproque minimum de 90% entre la query et la subject, une identité supérieure ou égale à 90% entre acides aminés du point de vue de leurs propriétés sur la zone alignées, et avec au maximum 1% gaps sur la zone alignée). En 2018, j'ai identifié 7 protéines candidates pour lesquelles des peptides signaux ont été prédits, dont 5 étaient également prédites dans des réseaux métaboliques liés à des mécanismes de sécrétion et pour lesquelles aucun motif de rétention transmembranaire n'a été identifié. Trois protéines ont été retenues pour l'étape de confirmation sur la base de l'alignement de séquences multiples des 5 protéines précédemment validées par les programmes de prédictions de sécrétion.

**Amplification et reséquençage des candidats 'de polymorphisme' sur la gamme de 12 clones.** Pour chacun des 3 transcrits candidats AVR attendus en polymorphisme, des amplicons de 300-700 bp encadrant leur SNP respectifs attendus responsables de la transition AVR-VIR, ont été amplifiés par RT-PCR à partir d'extractions d'ARN des 12 clones de la collection du GAFL [Figure 8] pour lesquels le phénotype AVR vs VIR a clairement été identifié. Le reséquençage Sanger de ces amplicons n'a pas permis de confirmer le profil AVR/VIR attendu de ces 3 candidats.

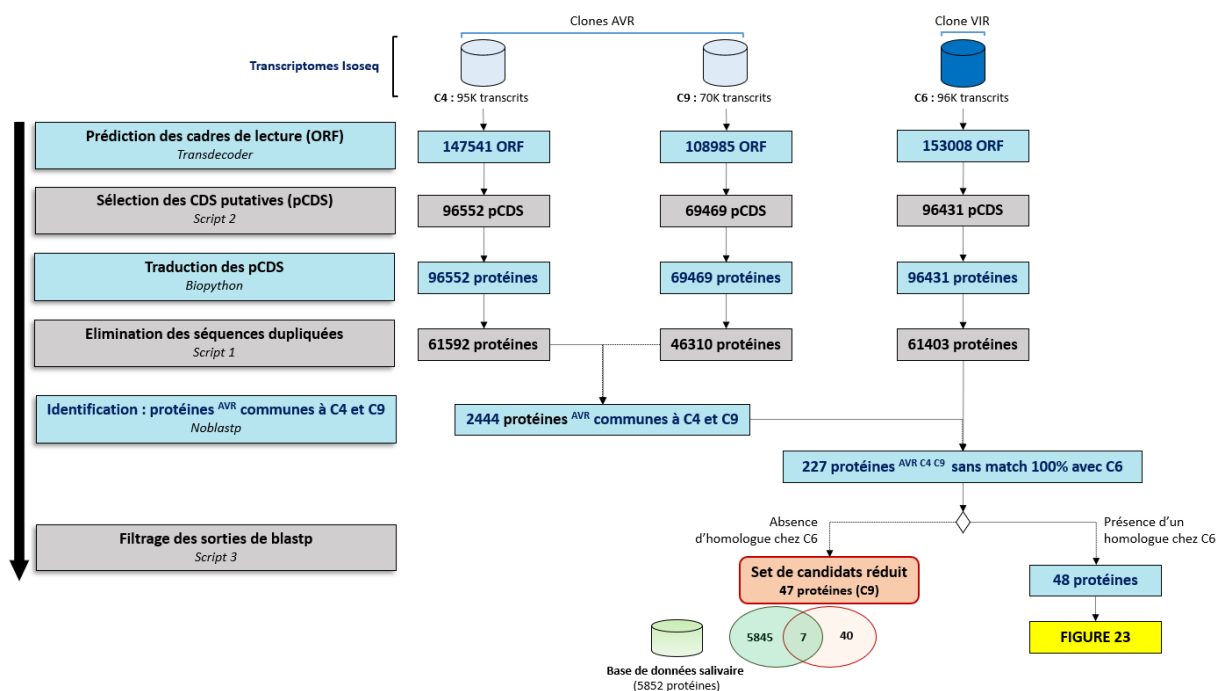
**Conclusion.** Les approches *in silico* basée sur la seule utilisation des données RNAseq produites à partir d'ARN de têtes de pucerons m'ont permis d'identifier : (i) un total de 73 transcrits candidats d'AVR non exprimés chez le clone VIR C6, dont 8 (11%) ont été identifiés dans la base de données salivaire (5852 protéines), (ii) un total de 3 transcrits candidats avec un SNP non synonyme (aucune correspondance dans la base de données salivaire).

Les profils AVR-VIR obtenus par amplification/reséquençage de ces candidats chez la gamme de douze clones a montré une discordance générale avec les profils AVR/VIR attendus, ayant par conséquent conduit à l'élimination de l'ensemble des candidats : (i) de 'non expression' et (ii) 'de polymorphisme'.

### 4.4.2 Stratégie 2 : combinaison de données RNAseq et Isoseq

**Recherche des candidats sur les données Isoseq *sensu stricto*.** J'ai établi une stratégie de recherche de transcrits candidats à l'avirulence sous l'hypothèse que la transition AVR-VIR est gouvernée par (i) par l'absence du transcrit codant pour la protéine d'AVR et que l'allèle d'avirulence est commun aux deux clones AVR ou (ii) un polymorphisme dans le gène d'AVR induisant un changement à l'échelle de la protéine.

**Note explicative sur la stratégie implémentée.** Dans cette stratégie, j'ai utilisé : (i) les transcriptomes pleine longueur reconstruits à partir des données Isoseq [Figures 15 et 16] produites par séquençage longs fragments des ARN de têtes de pucerons et (ii) les données RNAseq produites par séquençage en lecture courte des mêmes ARN de têtes de pucerons dont les caractéristiques sont présentées dans les figures 13 et 14. Cette stratégie, présentée en figure 20, consiste dans un premier temps à produire, à partir des transcriptomes Isoseq, un set de peptides uniques pour chacun des trois clones puis à éliminer ceux qui sont identiques à la base près.



**Figure 20 - Recherche de candidats 'de non expression' et 'de polymorphisme' avec les données Isoseq de têtes de pucerons**

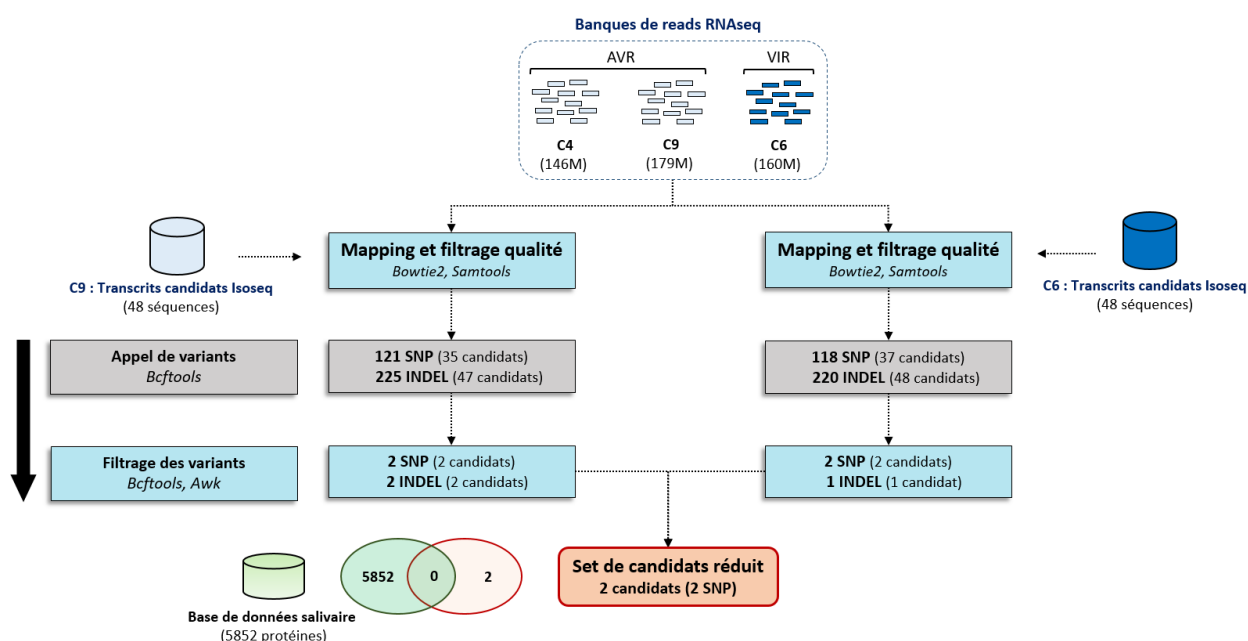
Seuls les transcrits de longueur supérieure à 100 bp (plus de 99,9% des transcriptomes de têtes pleine longueur pour chacun des clones) ont été utilisés. Après conversion dans les 6 cadres de lecture (étape 2), le cadre le plus long pour chaque transcrit a été conservé et traduit à l'aide du **script 2** (cf. annexe) (étape 3), aboutissant à environ 70000 à 96000 protéines putatives. Autour de 35% des protéines étaient des duplicats et ont été retirées des jeux de données pour la suite à l'aide du **script 1** (cf. annexe) (étape 4), aboutissant à une moyenne de 56000 protéines uniques par clone [Figure 20]. Ensuite, un total de 2444 protéines communes aux deux clones AVR C4 et C9 ont été identifiées par alignement BLASTP (étape 5) à l'aide du **script 3** (cf. annexe). Aussi, 227 protéines AVR (9% de 2444) sont sans correspondance exacte (alignement BLASTP) avec les protéines VIR et 25 (11%) ont été identifiés dans la base de données salivaire (**script 3**).

**Définition des candidats 'de non expression' vs 'de polymorphisme'**. Parmi ces 227 protéines, 47 (21%) n'avaient pas d'homologue chez le clone VIR et ont été traitées comme des candidats de non expression (*i.e.* exprimés chez C4 et C9 et pas chez C6). Parmi ces 227 protéines, 48 (21%) étaient hautement similaires (avec un alignement réciproque minimum de 85% entre la séquence d'entrée et une séquence de la base interrogée, une identité supérieure ou égale à 95% entre acides aminés du point de vue de leurs propriétés sur la zone alignées) avec au moins une protéine du clone VIR et ont été traitées comme des candidats de polymorphisme. Les 131 protéines restantes (58%) n'ont été attribuées à aucune des deux hypothèses [Figure 20] et n'ont par conséquent pas été traitées par la suite.

**Amplification des candidats 'de non expression' sur la gamme de 12 clones.** Pour chacun des **47 transcrits candidats AVR attendus en non expression**, des amplicons de 300-700 bp inclus dans leurs CDS respectifs, ont été amplifiés par RT-PCR à partir d'extractions d'ARN des **12 clones** de la collection du GAFL [Figure 8] pour lesquels le phénotype AVR/VIR a clairement été identifié. L'examen des produits de PCR (*i.e.* une absence d'amplification est attendue chez les clones VIR alors qu'une amplification est attendue chez les clones AVR), ainsi que leurs tailles, par **électrophorèse sur gel d'agarose n'a permis de confirmer le profil AVR/VIR attendu d'aucun de ces 47 candidats**.

**Identification des candidats 'de polymorphisme'**. Pour les 48 candidats, les données Isoseq étant des consensus, j'ai utilisé les reads de RNAseq [Tables 1 et 2] pour confirmer un polymorphisme de type SNP/INDEL. Cette stratégie, présentée en **figure 21**, est basée sur la détection de SNP communs aux deux clones AVR, C4 et C9, et absents chez le clone VIR C6. En amont, les reads RNAseq du clone VIR ont été mappées sur les 48 transcrits correspondant à l'allèle AVR et vice et versa.

Dans les deux mappings, l'ensemble des transcrits étaient couverts. Les appels de variants ont été effectués sur des critères de qualité conservatifs. Le mapping des reads AVR et VIR sur les 48 transcrits candidats de C9 (AVR) a révélé un total de 121 SNP sur 35 transcrits et 225 INDEL sur 48 transcrits [Figure 21]. Le mapping utilisant les 48 transcrits de C6 (VIR) comme référence a mis en évidence un total de 118 SNP (sur 37 candidats) et 48 INDEL (sur 48 candidats) [Figure 21]. Au final, 2 des 48 transcrits candidats d'avirulence attendus avec un polymorphisme de type SNP/INDEL, précédemment identifiés par comparaison des transcriptomes pleine longueur AVR/VIR, ont été redécouverts avec l'appel de variants [Figure 21]. Ces deux candidats portent chacun un SNP non synonyme.



**Figure 21** – Recherche de candidats ‘de polymorphisme’ en combinant des données Isoseq et RNAseq de têtes de pucerons

**Amplification et reséquençage des candidats ‘de polymorphisme’ sur la gamme de 12 clones.** Pour chacun des 2 transcrits candidats AVR attendus en polymorphisme, des amplicons de 300-700 bp encadrant leur SNP respectifs attendus responsables de la transition AVR-VIR, ont été amplifiés par RT-PCR à partir d’extractions d’ARN des 12 clones de la collection du GAFL pour lesquels le phénotype AVR vs VIR a clairement été identifié. Le reséquençage Sanger de ces amplicons n’a pas permis de confirmer le profil AVR/VIR attendu de ces deux candidats sur l’ensemble de la collection de clones. Cependant, pour l’un d’eux, les données de reséquençage ont permis de confirmer le profil AVR vs VIR attendu chez les trois clones C4, C6 et C9, pour lesquels les données RNAseq et Isoseq sont disponibles (*nb* : les ARN utilisées pour les séquençages et ceux utilisées en biologie moléculaire proviennent de deux extractions indépendantes). Par la suite, je désignerai ce transcrit Isoseq par son identifiant FASTA unique : ‘55596’.

**A propos du transcrit Isoseq candidat 55596.** En accord avec les analyses *in silico* mobilisant les données RNAseq et Isoseq de têtes de pucerons, un **SNP** est attendu **en position 275** avec un profil **homozygote A/A dans l'allèle AVR** (clone C4 et C9), et un profil **homozygote G/G chez l'allèle VIR** (clone C6). Les séquences d'acides aminés prédites pour les deux versions alléliques de ce transcrit candidat révèlent la **substitution d'une alanine (forme AVR) par une thréonine (forme VIR)** [Figure 22]. La séquence nucléotidique (551 bp) du **transcrit Isoseq candidat 55596**, a été recherchée par BLAST dans la base de données NR (non redondante) du NCBI (version 2020). Le top 15 des correspondances contre la base NR du NCBI a fait unanimement ressortir l'annotation fonctionnelle '**Cytochrome P450 Monooxygenase**' chez **7 espèces de pucerons** différentes. La famille des Cytochrome P450 est notamment connue pour son rôle dans la détoxification des métabolites secondaires des végétaux.



**Figure 22 - Confirmation du candidat '55596'**

Portion de la séquence nucléotidique du transcrit Isoseq candidat de C9 55596 (551 bp). Les séquences d'acides aminés ont été prédites avec Emboss Transeq. Les ARN à partir desquels les transcrits Isoseq candidats (en haut) et les amplicons Sanger (en bas) ont été obtenus sont issus de deux extractions indépendantes.

**Conclusion.** Les approches *in silico* combinant l'utilisation **données d'expression** (RNAseq) et de **structure** (Isoseq) produites à partir d'**ARN de têtes** de pucerons m'ont permis d'identifier : **(i)** un total de **47 transcrits candidats d'AVR non exprimés chez le clone VIR C6**, dont 7 (15%) ont été identifiés dans la base de données salivaire (5852 protéines), **(ii)** un total de **2 transcrits candidats avec un SNP non synonyme** (aucune correspondance dans la base de données salivaire).

Les profils AVR-VIR obtenus par **amplification/reséquençage** de ces candidats chez la gamme de **douze clones** a montré une **discordance générale avec les profils AVR/VIR attendus**, ayant par conséquent conduit à **l'élimination de l'ensemble des candidats** : **(i)** de 'non expression' et **(ii)** 'de polymorphisme'. Pour le candidat 'de polymorphisme' **55596** cependant, les profils AVR/VIR observés et attendus sont **100% concordants au sein de la race d'hôte des cucurbitacées** (*i.e.* une concordance partielle sur l'ensemble de la gamme de clones utilisée), représentant la moitié de la gamme d'hôte testée.

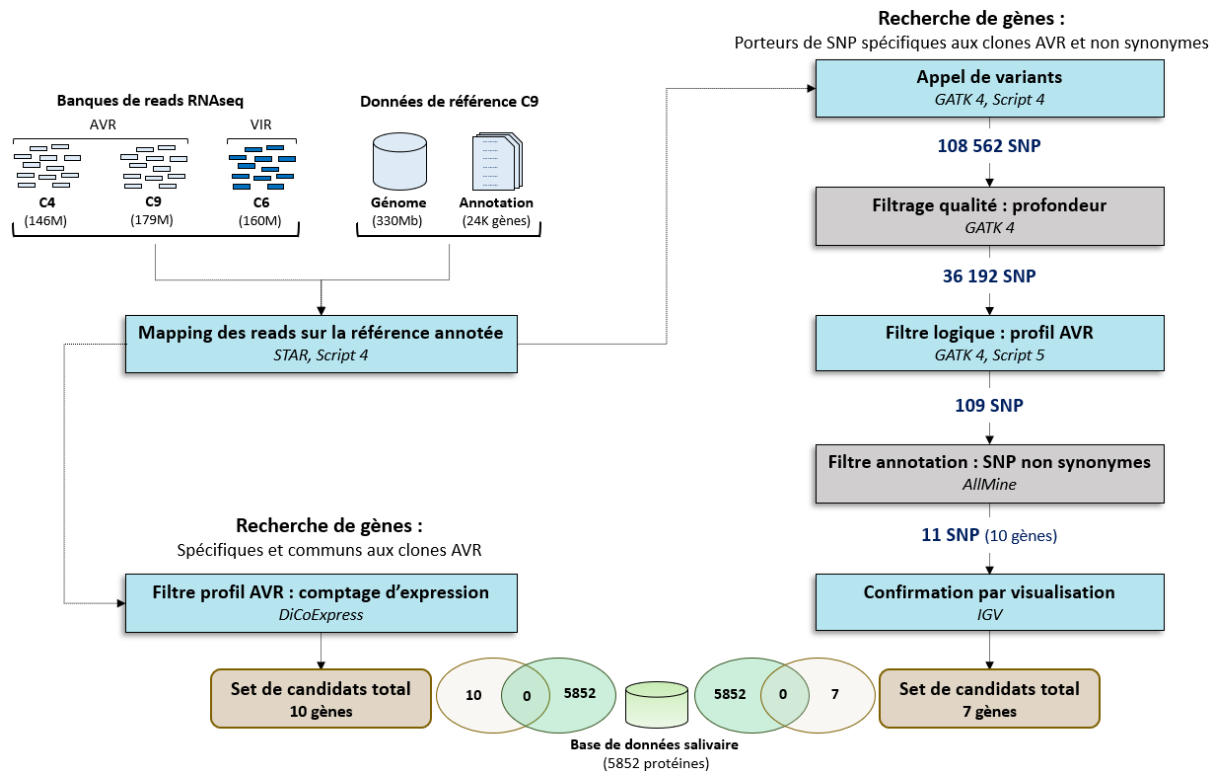
#### 4.4.3 Stratégie 3 : combinaison de données RNA-seq, PacBio Isoseq et génome de référence

**Note explicative.** Le génome du clone C9, disponible en 2022, m'a permis de développer une **stratégie 'conventionnelle'** pour la recherche de candidats d'AVR. Pour la première fois, la disponibilité du génome annoté de C9 m'a permis d'étudier une troisième hypothèse où la transition AVR-VIR est gouvernée par un évènement d'épissage alternatif.

##### Recherche de gènes non exprimé du clone VIR et de gènes portant des SNP d'AVR.

J'ai établi une stratégie de recherche de transcrits candidats à l'AVR sous l'hypothèse que la transition AVR-VIR est gouvernée par l'absence du transcrit codant, de la présence d'un polymorphisme pour la protéine d'AVR ou d'un évènement d'épissage alternatif. Dans cette stratégie, j'ai utilisé : **(i)** le génome de référence du clone AVR C9 dont les caractéristiques sont présentées dans la **figure 10**, **(ii)** son annotation (24K gènes) établie par Raphaël Feriche-Linares et Jacques Lagnel (INRAE Avignon, UR GAFL, équipe I2B) qui combine des données Isoseq **[Table 2]** produites par la Plateforme de service en génotypage et séquençage à haut débit GENTYANE (INRAE Clermont-Ferrand, GDEC-UCA), la base de données salivaire que j'ai implémentée **[Figure 12]** et **(iii)**, les données d'expression issues du séquençage RNAseq des têtes de pucerons **[Table 1]**. Cette stratégie, présentée en **figure 23**, est basée sur le mapping des reads RNAseq des trois clones (C6 étant le clone VIR, et C4 et C9 étant les deux clones AVR) et sur les gènes prédits du génome de référence pour le clone C9 **[Figure 23 en haut à gauche]**.





**Figure 23** - Recherche de candidats 'de non expression' et 'de polymorphisme' en combinant des données RNAseq de têtes de pucerons et le génome de référence annoté pour le clone C9.

La partie 'Mappings' et 'Appel de SNP' a été réalisée à l'aide du **script 4** (cf. annexe). Environ 95% des reads de chaque clone sont mappés sur le génome du clone C9. A partir de ce mapping, la chaîne de traitement pour l'identification de gènes candidats se scinde en deux parties : l'une destinée à l'identification de gènes spécifiquement exprimés chez les deux clones AVR C4 et C9 **[Figure 23 en bas à gauche]**, et l'autre destinée à l'identification de gènes porteurs de SNP (i) communs et spécifiques aux deux clones AVR et (ii) induisant un changement à l'échelle de la protéine codée par le gène d'avirulence **[Figure 23 à droite]**.

**Recherche de candidats de non expression.** Les tables de comptage pour le mapping précédemment produit ont été générées puis filtrées pour identifier des gènes candidats de non expression. De cette manière, au total, 10 gènes candidats ont été identifiés comme exprimés (plus de 30 comptages) par les deux clones AVR C4 et C9, et pas ou peu exprimés (nombre de comptages inférieur ou égal à 5% de celui des reads de chaque clone AVR et n'excédant pas 10) par le clone VIR C6 ont été identifiés.

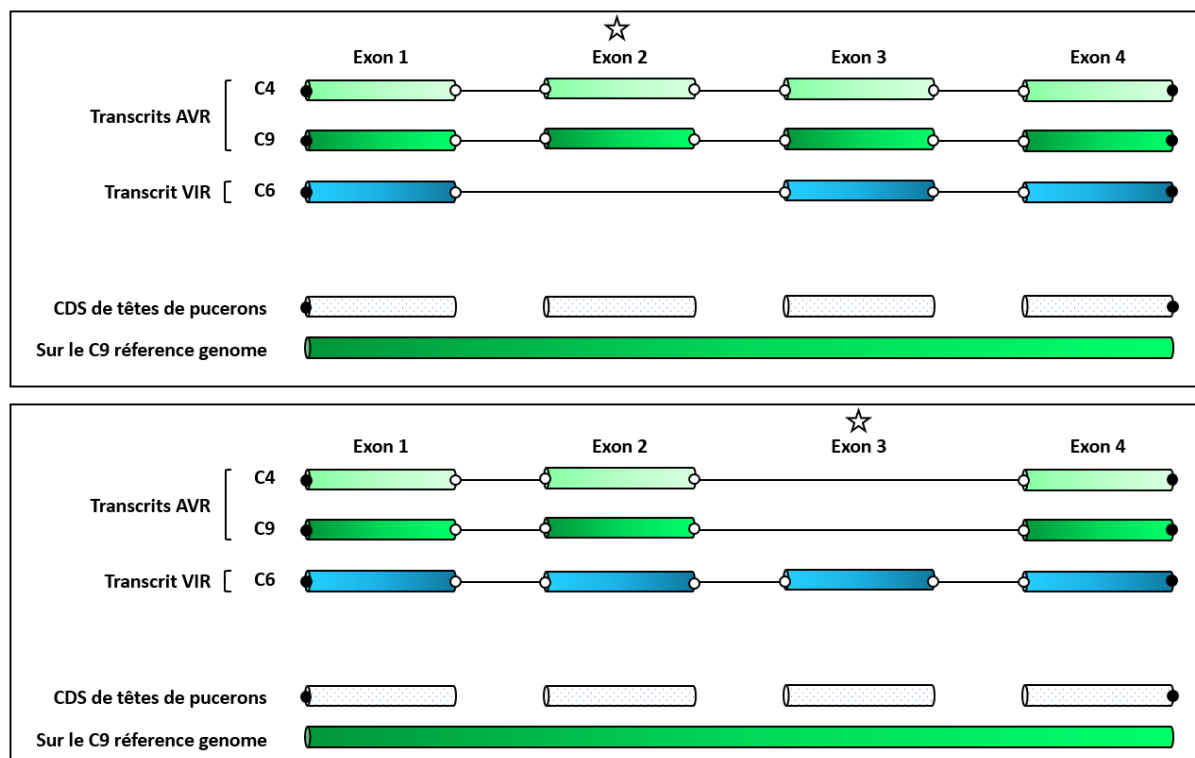
**Amplification des candidats ‘de non expression’ sur la gamme de 12 clones.** Pour chacun des **10 transcrits candidats AVR attendus en non expression**, des amplicons de 300-700 bp inclus dans leurs CDS respectifs, ont été amplifiés par **RT-PCR à partir d’extractions d’ARN des 12 clones** de la collection du GAFL [Figure 8] pour lesquels le phénotype AVR/VIR a clairement été identifié. L’examen des produits de PCR (*i.e.* une absence d’amplification est attendue chez les clones VIR alors qu’une amplification est attendue chez les clones AVR), ainsi que leurs tailles, par **électrophorèse sur gel d’agarose n’a permis de confirmer le profil AVR/VIR attendu d’aucun de ces 10 candidats.**

**Appel et filtrage des SNP.** L’appel de SNP, effectué à partir du même mapping que celui utilisé pour la recherche de candidats ‘de non expression’ a mis en évidence un total de 108562 positions sur le génome de référence de C9 pour lesquelles au moins un variant existait entre les trois clones. Seuls les SNP bialléliques ont été considérés dans la suite. Un tiers des SNP bruts (36192 SNP) ont validé les critères de qualité donnés dans **la partie Matériels et Méthode [page 103]** et sont bialléliques. Puis, ces SNP ont été filtrés avec le **script 5** (*cf.* annexe) pour le profil des SNP bialléliques définissant les candidats d’AVR [Figure 11].

**Identification des SNP candidats.** L’annotation des SNP avec Allmine, utilisant le génome de référence et son annotation, ont révélé 11 SNP dans des zones codantes et entraînant un changement au niveau protéique (*i.e.* non synonymes), correspondant à 10 gènes. Les CDS issus de l’expression de ces gènes ont été traduits en protéines, puis recherchés dans la base de données salivaire disponible en 2020 [Figure 12] avec une recherche d’homologie stringente avec le **script 3** (*cf.* annexe) (le meilleur alignement par hit, un alignement réciproque minimum de 90% entre la query et la subject, une identité supérieure ou égale à 90% entre acides aminés du point de vue de leurs propriétés sur la zone alignée, et avec au maximum 1% gaps sur la zone alignée). Aucune de ces protéines n’a été identifiée dans la base de données salivaires.

**Amplification et reséquençage des candidats ‘de polymorphisme’ sur la gamme de 12 clones.** Pour chacun des **gènes candidats AVR attendus en polymorphisme**, des amplicons de 300-700 bp encadrant leur SNP respectifs attendus responsables de la transition AVR-VIR et chevauchant une jonction exon-exon, ont été amplifiés par **RT-PCR à partir d’extractions d’ARN des 12 clones** de la collection du GAFL [Figure 8] et également à partir d’extractions d’ADN pour C4, C6 et C9. Le **reséquençage Sanger** de ces amplicons n’a **pas permis de confirmer le profil AVR/VIR** attendu de ces 7 candidats **sur l’ensemble de la collection de clones.**

**Recherche de candidats en épissage alternatif.** J'ai établi une stratégie de recherche de gènes candidats à l'AVR sous l'hypothèse que la transition AVR-VIR est gouvernée par un évènement d'épissage alternatif dans le gène d'avirulence. Le profil attendu des candidats est présenté en **figure 24**.



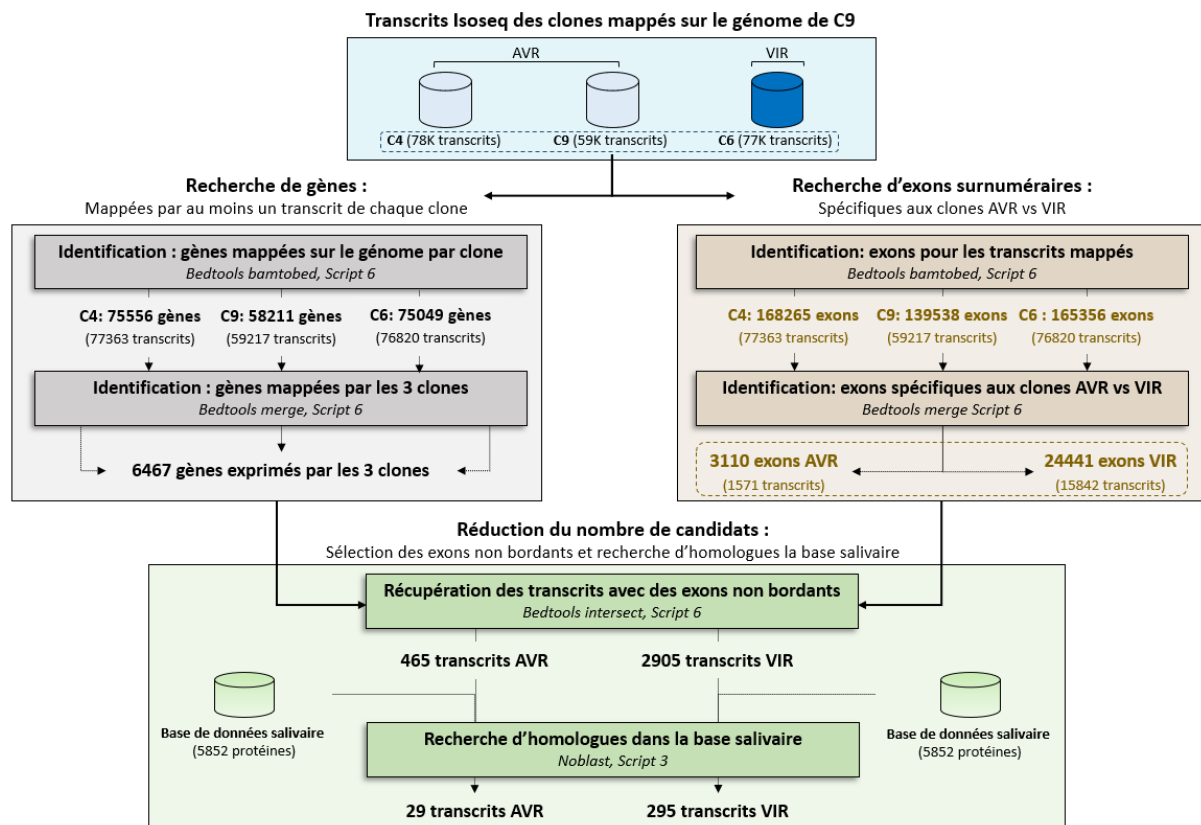
**Figure 24** – Profil attendu d'un gène candidat à l'AVR de type 'épissage alternatif'

Dans cette illustration, un CDS de tête de pucerons (cylindres blancs) est mappé sur le génome de référence du clone C9 (cylindre vert foncé en bas). Chacun des 3 clones (AVR en vert et VIR en bleu) expriment une seule isoforme (une ligne) et le CDS pleine longueur comporte 4 exons distincts (tronçons cylindriques). Les lignes noires et fines représentent les introns. Les points noirs représentent les bornes des CDS et les points blancs représentent les bornes des exons. Les étoiles blanches désignent les exons porteurs de l'AVR.

Deux évènements d'épissage alternatifs sont pris en compte : le saut d'exon, illustré dans l'encart supérieur (*i.e.* l'exon surnuméraire est dans l'isoforme VIR), et la rétention d'intron, illustrée dans l'encart inférieur (*i.e.* l'exon surnuméraire est dans l'isoforme AVR).

**Encart supérieur** : le gène codant pour les CDS représentés est un candidat à l'avirulence car l'exon 2 est présent dans les deux isoformes AVR, et absent de l'isoforme VIR. **Encart inférieur** : le gène codant pour le CDS représenté est un candidat à l'avirulence car l'exon 3 est présent dans les deux isoformes AVR, et absent de l'isoforme VIR.

Cette stratégie, présentée en **figure 25**, est basée sur la recherche d'exon surnuméraire, soit dans les clones AVR vs le clone VIR, soit dans le clone VIR vs AVR. Cette recherche se base sur l'analyse des coordonnées des exons des clones transcrits dans les gènes annotés du génome de référence du C9. Dans la mesure où les régions UTR peuvent englober tout ou partie des exons flanquants, et que les transcrits Isoseq ne sont pas systématiquement reconstitué dans leur intégralité, **la stratégie que j'ai développée se concentre sur la recherche d'exons candidats qui ne flanquant pas les régions UTR** (*i.e.* un gène candidat doit posséder au moins trois exons).



**Figure 25** - Recherche de candidats avec des évènements d'épissage alternatifs en combinant des données Isoseq et RNAseq de têtes de pucerons et le génome de référence annoté pour le clone C9.

J'ai identifié des gènes qui sont exprimés chez les 3 clones C4, C6 et C9, et qui présentent un évènement d'épissage alternatif commun aux deux clones AVR, C4 et C9, et absent chez le clone VIR C6. A la différence des deux stratégies précédentes, l'identification des gènes candidats à l'AVR repose sur **l'analyse et la manipulation des coordonnées des séquences** au lieu des séquences nucléotidiques. Actuellement, des solutions existent pour détecter des évènements d'épissage alternatifs chez un individu donné, mais il n'en existe pas pour détecter et comparer des évènements d'épissage alternatifs entre individus. Par conséquent, j'ai implémenté une stratégie à partir de zéro pour identifier des candidats avec des évènements d'épissage alternatifs avec le **script 6** (*cf.* annexe).

Dans cette stratégie, **les exons sont les éléments clés** pour déterminer si un gène sur le génome de référence C9 est candidat à l'AVR [Figure 24]. Dans la mesure où j'étudie la transition AVR-VIR, je considère que la forme AVR est l'évènement d'épissage **constitutif** des ARN pré-matures exprimés dans les têtes des pucerons. Ici, **je considère deux évènements d'épissage alternatifs** : le saut d'exon [Figure 25, encart supérieur], où l'exon surnuméraire est dans l'isoforme **VIR**) et la rétention d'intron [Figure 25, encart inférieur] (*i.e.* l'exon surnuméraires est dans l'isoforme **AVR**).

Dans un premier temps, j'ai produit les mappings des transcrits Isoseq clusterisés [Figure 15 et 16] sur le génome de référence pour le clone C9 [Figure 10]. Environ 99% des transcrits Isoseq clusterisés de chaque clone ont été mappés sur le génome de C9 avec 99% de chance d'être aligné de manière unique [Figure 25, encart bleu].

Pour chacun des trois mappings, j'ai généré les coordonnées des régions génomiques mappées et sélectionné les positions (*i.e.* plages de coordonnées) sur le génome de référence pour C9 qui sont mappées par au moins un transcrit Isoseq de chaque clone [Figure 25, encart gris]. Au total, entre 58000 et 75500 positions sur le génome de C9 ont été couvertes par 59000 à 77000 transcrits Isoseq. Les coordonnées des régions génomiques mappées par les transcrits de chaque clone ont été comparées, révélant ainsi 6471 régions génomiques (entre 8,5 et 11,1% de positions uniques pour chaque clone) sur le génome de C9 couvertes par au moins un transcrit Isoseq de chacun des trois clones [Figure 25, encart gris]. Ces 6471 régions génomiques seront utilisées par la suite.

En parallèle, j'ai recherché des exons, soit spécifiques et communs aux deux clones AVR [Figure 25, encart du haut], soit spécifiques au clone VIR C6 [Figure 25, encart du bas]. L'analyse des coordonnées des transcrits mappés sur le génome de référence a permis de détecter un total de 140000 à 170000 exons, correspondant à 59000 à 77000 transcrits (soit une moyenne globale de 2,3 exons par CDS). Seulement 3110 exons étaient spécifiquement présents chez les deux clones AVR (2,0% du total des exons AVR détectés) et correspondaient à 1571 transcrits candidats (2,3% des transcriptomes Isoseq clusterisés des clones AVR en moyenne). Un total de 24441 exons étaient spécifiquement présents chez le clone VIR C6 (15% du total des exons VIR détectés), correspondant à 15842 transcrits candidats (20% du transcriptomes Isoseq clusterisés du clone C6).

J'ai ensuite croisé les coordonnées des 6471 régions génomiques couvertes par au moins un transcrit de chaque clone [Figure 25, encart gris] avec les coordonnées transcrites ayant au moins un exon spécifique à la forme AVR vs VIR [Figure 25, encart doré]. Au total, 30% des transcrits possédant au moins un exon AVR surnuméraire (465 transcrits) et 18% des transcrits possédant au moins un exon

VIR (2905 transcrits) surnuméraire codent pour des gènes avec au moins 3 exons et donc, ont le profil attendu.

Ces transcrits candidats ont été convertis dans les différents cadre de lecture et le cadre le plus long a été conservé puis traduit, aboutissant à 465 protéines candidates côté AVR et 2905 côtés VIR. Enfin, 29 protéines candidates côté AVR (6% de 465) et 295 protéines candidates côté VIR (10% de 296) ont été identifiés dans la base de données salivaire (5852 protéines), disponible en 2020 [Figure 12] avec une recherche d'homologie stringente avec le **script 3** (cf. annexe) (le meilleur alignement par hit, un alignement réciproque minimum de 90% entre la query et la subject, une identité supérieure ou égale à 90% entre acides aminés du point de vue de leurs propriétés sur la zone alignées, et avec au maximum 1% gaps sur la zone alignée).

**Amplification des candidats 'd'épissage alternatif' sur la gamme de 12 clones.** Pour chacun des 29 gènes candidats AVR attendus avec un évènement d'épissage alternatif '**exon supplémentaire chez les clones AVR**', des amplicons de 300-700 bp inclus dans leurs CDS respectifs et chevauchant une jonction exon-exon, ont été amplifiés par **RT-PCR à partir d'extractions d'ARN des 12 clones** de la collection du GAFL [Figure 8] et également à partir d'extractions d'ADN pour C4, C6 et C9. L'examen des produits de PCR (*i.e.* une absence d'amplification est attendue chez les clones VIR alors qu'une amplification est attendue chez les clones AVR), ainsi que leurs tailles, par **électrophorèse sur gel d'agarose n'a permis de confirmer le profil AVR/VIR attendu d'aucun de ces 29 candidats.**

**Conclusion.** Les approches *in silico* combinant l'utilisation **données d'expression** (RNAseq) et de **structure** (Isoseq) produites à partir d'**ARN de têtes** de pucerons, ainsi que du **génomme de référence annoté** pour le clone AVR C9 m'ont permis d'identifier : **(i)** un total de **10 gènes candidats d'AVR non exprimés chez le clone VIR C6** (aucune correspondance dans la base de données salivaire), **(ii)** un total de **7 gènes candidats avec un SNP non synonyme** (aucune correspondance dans la base de données salivaire) et **(iii)**, un total de **324 gènes candidats avec un évènement d'épissage alternatif**, dont 29 avec un exon surnuméraire chez les clones AVR et 295 avec un exon surnuméraire chez le clone VIR, (100% ont été identifiés dans la base de données salivaire) avec **un effecteur connu**, Me47.

Les profils AVR-VIR obtenus par **amplification/reséquençage** de ces candidats chez les douze clones a montré une **discordance générale avec les profils AVR/VIR attendus**, ayant par conséquent conduit à **l'élimination de l'ensemble des candidats** : **(i)** de 'non expression', **(ii)** 'de polymorphisme' et **(iii)** d'épissage alternatif pour les cas où l'exon surnuméraire est spécifique aux clones AVR. Les candidats d'épissage alternatif pour les cas où l'exon surnuméraire est spécifique au clone VIR n'ont pas été inclus dans la tentative de confirmation durant le temps de ma thèse.

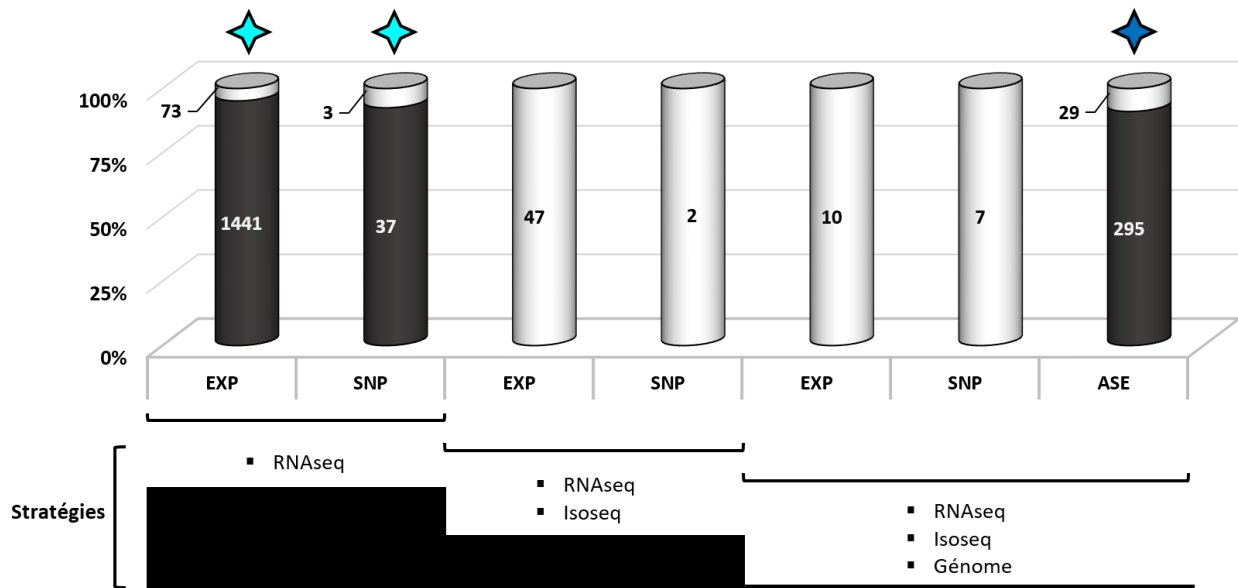
## 5. Discussion

Dans ce chapitre, j'ai étudié la voie pour le contournement de la résistance *Vat-1* par les pucerons qui consiste, en accord avec le **diagramme en Z**, en une **évolution du gène d'avirulence (AVR)** chez les clones interagissant avec la protéine NBS-LRR VAT-1. Mes travaux ont consisté à explorer, au moyen d'**analyses comparatives de données d'expression** de deux clones AVR (C4 et C9) et d'un clone virulent (VIR) C6, **trois hypothèses pour la transition AVR-VIR** : **(i) l'absence de transcrite** chez le clone VIR, **(ii) des polymorphismes spécifiques** du clone VIR et **(iii), des évènements d'épissage alternatif** chez le clone VIR. Ces trois cribles m'ont permis d'établir des **listes de candidats à l'AVR**, dont une partie a été amplifiée chez une gamme de douze clones d'*A. gossypii*, reflétant une diversité génétique importante où trois races d'hôtes sont représentées (Cucurbitacées, Malvacées et Solanacées) et clairement caractérisés pour leur caractère AVR/VIR, de façon à faire ressortir des listes réduites composées de candidats pertinents.

Dans cette partie organisée en **5 points**, **(i)** j'offre une vue d'ensemble des candidats d'AVR que j'ai identifiés via différentes approches *in silico*, **(ii)** je discute la pertinence des stratégies testées et leur cohérence sur la base des listes de candidats qu'elles ont produits, **(iii)** je traite le cas particulier d'un candidat de type 'polymorphisme' au profil AVR/VIR confirmé pour les 6 clones spécialisés sur les cucurbitacées, **(iv)** je discute de intérêt d'avoir implémentée une base de données de protéines salivaires d'hémiptères durant ma thèse ainsi que de l'usage que j'en ai eu, **(v)** enfin, je propose de ré-exploiter des stratégies de recherche *in silico* que j'ai implémentées afin d'identifier des candidats additionnels.

### 5.1. Vue d'ensemble des candidats d'AVR identifiés

Au total, j'ai implémenté **sept stratégies** de recherche *in silico* redondantes sur les hypothèses testées, intégrant différents types de données **en fonction de la date de leur mise à disposition** : des données **RNAseq** de têtes de pucerons (**2018**), des données **Isoseq** de têtes de pucerons (**2020**) et un **génom**e de référence annoté pour le clone AVR C9 (**2022**). Un résumé des résultats obtenus est présenté dans la **figure 26**.



**Figure 26** – Bilan des candidats d’AVR identifiés par approches *in silico*

Nombre de candidats (en % du total) identifiés par approches *in silico* et amplifiés sur la collection des 12 clones AVR/VIR du GAFL (**cylindres blancs**) et non testés durant ma thèse (**cylindres noirs**).

Les arguments utilisés pour une amplification sur la collection sont désignés par les étoiles colorées : utilisation de programmes de prédiction de motifs d’adressage à l’exportation par voie salivaire (**bleu clair**), ou utilisation de la base de données salivaire (**bleu foncé**).

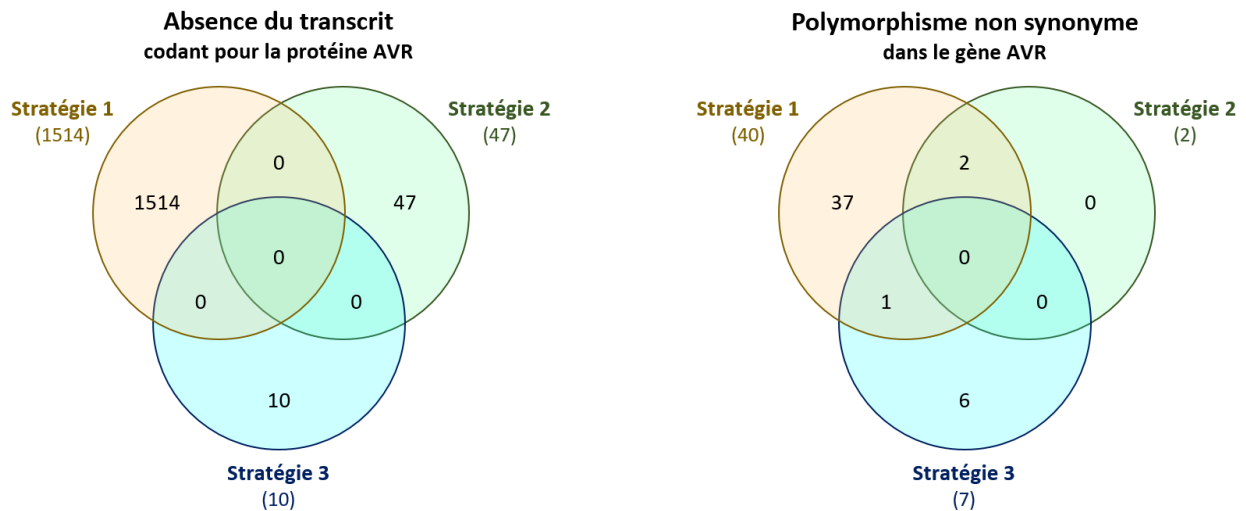
**Trois hypothèses pour la transition AVR-VIR** : l’absence de transcrit chez le clone VIR C6 (**EXP**), des polymorphismes spécifiques du clone VIR (**SNP**) et des évènements d’épissage alternatif chez le VIR (**ASE**).

Trois approches *in silico* m’ont permis d’identifier un total **1571 candidats d’AVR non exprimés chez le clone VIR**, dont 130 (8%) ont été retenus, sur la base de prédiction motifs d’adressage à l’exportation par voie salivaire, pour être amplifiés sur la collection de clones. Également, trois approches m’ont permis d’identifier un total de **95 candidats d’AVR avec un polymorphisme chez le clone VIR**, dont 12 (12%) ont été retenus, pour être amplifiés sur la collection de clones. Enfin, une approche m’a permis d’identifier un total de **324 candidats d’AVR avec un évènement d’épissage alternatif**, dont 29 (10%) ont été retenus, identifiés dans la base de données salivaire, pour être amplifiés sur la collection de clones en vue d’établir une sous-liste de candidats solides.

## 5.2 Pertinence des stratégies et cohérence entre elles

Les hypothèses où la transition AVR-VIR est causée par (i) l’absence de transcrit chez le clone VIR C6 et (ii), des polymorphismes spécifiques du clone VIR, ont été traitées par trois stratégies *in silico* combinant différents types de données, chaque stratégie ayant abouti à la production d’un nombre de candidats présenté dans la **figure 27**.





**Figure 27** – Nombre de candidats d’AVR identifiés par différents usages de données *in silico*

Trois stratégies, intégrant différents types de données en fonction de la date de leur mise à disposition, ont été implémentées pour identifier des candidats : absents chez le clone VIR (à gauche) et avec un polymorphisme chez le clone VIR (à droite). La **stratégie 1** est exclusivement basée l’utilisation des données de RNAseq séquencées à partir d’ARN de têtes de pucerons. La **stratégie 2** combine l’utilisation des données de RNAseq et des données d’Iseq produites à partir des même ARN de têtes. La **stratégie 3** combine l’utilisation des données de RNAseq et en plus le génome de référence annoté pour le clone C9.

Notamment, concernant l’hypothèse de la non expression du facteur d’AVR, **aucune intersection** n’existe entre les trois listes de candidats [Figure 27, gauche]. Aussi, la **stratégie 1**, reposant uniquement sur l’utilisation des données RNAseq de têtes de pucerons, a produit la liste de candidats la plus conséquente (1514). Dans cette liste, 73 candidats (4,8%) identifiés comme facteurs putativement sécrété, ont été amplifiés sur une extraction indépendante pour confirmer leur profil AVR-VIR. Cependant, aucun de ces candidats n’a été confirmé, **soulignant l’importance du nombre de faux positifs** générés avec la **stratégie 1**. De la même manière, aucun des candidats issus des listes produites par **les stratégies 2 et 3** (respectivement 48 et 10) n’est ressorti positif à la tentative de confirmation.

En comparaison, la recherche de candidats sous l’hypothèse que la transition AVR-VIR est gouvernée par des SNP à produit des listes réduites de candidats en comparaison de celles issues du traitement de l’hypothèse de la non expression du facteur d’AVR (*cf.* stratégie 1). La recherche d’intersections entre les listes de candidats produites dans le traitement de l’hypothèse ‘polymorphisme’ a mis en évidence entre un et deux candidats communs à deux listes. Toutefois, aucun n’a été redécouvert par les trois stratégies, ayant produit un total de 49 candidats [Figure 27, droite]. Un sous-ensemble de 12 candidats (24%), sélectionné sur la base de prédiction de motifs

d'adressage à l'exportation par voie salivaire (stratégie 1), ou d'identification dans la base de données salivaires (stratégie 3), ont été reséquencés à partir d'extractions indépendantes pour les trois clones C4, C9 et C6. Un seul de ces candidats avait le profil AVR-VIR attendu. Notamment, ce candidat, le transcrit Isoseq 55596, avait été communément identifié dans les listes produites par les stratégies 1 et 2 [Figure 22].

En l'absence d'un génome de référence, l'assemblage *de novo* d'ARN-Seq reste la seule option possible pour étudier un transcriptome (Grabherr et al., 2013). La reconstruction précise de transcrits complets à partir de données RNAseq pose des problèmes, notamment dus: **(i)** à des taux de couverture variables entre les transcrits, **(ii)** des transcrits encodés par des loci proches en terme de séquence pouvant se chevaucher et être fusionnés par erreur formant ainsi un transcrit chimérique, **(iii)** des introductions d'ambiguïtés d'assemblage dues à des séquences répétées dans différents gènes (Grabherr et al., 2011). De ce fait, les erreurs d'assemblages *de novo*, à partir de données RNAseq, sont la source principale des faux positifs observés. Au final, le taux de faux positifs très important, et ce, quelle que soit la stratégie implémentée. Les résultats de mes travaux de recherche confirment que l'absence de génome de référence est **fortement pénalisante**, sinon rédhibitoire, pour la recherche de transcrits absents ou de polymorphismes. La tentative de redécouverte de SNP combinant les données RNAseq et Isoseq que j'ai mis en place ne pouvait logiquement permettre de ne distinguer que des SNP réellement homozygotes de polymorphismes consensus. Par conséquent, les taux de faux négatifs et de faux positifs s'avèrent très élevés.

La **démarche conventionnelle**, implémentée dans la **stratégie 3**, permet *à priori* de génotyper et de prédire de manière fiable l'impact de polymorphismes à l'échelle de la protéine, mais aussi de s'assurer que les loci sur lesquels portent les comptages d'expression sont réellement des gènes. A nouveau, de nombreux faux positifs ont été générés, comme en témoignent les non-confirmations. A partir de ces constats, il me paraît raisonnable d'expliquer la très forte prédominance des faux positifs par une imprécision dans les annotations structurales, source d'erreurs dans la sélection des unités de comptage de mappings.

L'hypothèse où la transition AVR-VIR est causée par **un évènement d'épissage alternatif** a été traitée dans une seule stratégie, produisant une liste de **324 candidats** (dont 295 avec un exon surnuméraire dans l'allèle VIR et 29 avec un exon surnuméraire dans l'allèle AVR). Au cours de ma thèse, seuls les 29 candidats avec un exon surnuméraire dans l'allèle AVR ont été testés sur la collection de clones mais aucun n'a été confirmé.

### 5.3 Intérêt et usage de la base de données salivaire

Les mécanismes de sécrétion chez les insectes piqueurs-suceurs reposent sur des motifs moléculaires qui assignent des protéines à des voies de sécrétion (Bos et al., 2010). Des programmes de prédiction de ces motifs moléculaires d'adressage à l'export sont couramment utilisés dans des études sur les pucerons et pour la constitution de catalogues d'effecteurs salivaires putatifs (Boulain et al., 2018, 2019; Y. Zhang et al., 2017). Cependant, **la valeur prédictive de ses programmes est limitée** pour les insectes, et les pucerons en particulier, car ces programmes ont été entraînés sur des organismes éloignés, notamment des archées et des bactéries (Juan et al., 2019) et la prédiction des signaux d'importation est fiable dans les mitochondries et les chloroplastes (Emanuelsson et al., 2007) et se trouve restreinte à quelques espèces modèles *A. thaliana*, *H. sapiens* et *D. melanogaster*. En conséquence, des protéines détectées dans des échantillons de salive collectés sur milieu artificiel ne sont pas forcément prédites comme sécrétées et inversement, des protéines issues de glandes salivaires peuvent être prédites à tort comme sécrétés. De plus, la connaissance des mécanismes de sécrétion n'est pas exhaustive et la possibilité que des mécanismes de sécrétion encore inconnus existent chez les pucerons n'est donc pas exclue.

La base de données de protéines salivaires a été pensée et implémentée au cours de ma thèse **pour se substituer aux programmes de prédiction** de motifs d'adressage à l'exportation par voie salivaire traditionnellement utilisés dans les études sur les pucerons.

### 5.4 Un candidat pour la race d'hôte des cucurbitacées

De manière intéressante, le transcrit Iseq 55596 précédemment identifié comme candidat de type 'polymorphisme' a été identifié par les stratégies 1 et 2 [Figure 27, diagramme en haut à droite]. Son amplification chez une gamme de **douze clones** d'*A. gossypii* couvrant une **importante diversité génétique**, un **spectre d'hôte conséquent** et au sein de laquelle les phénotypes de **reconnaissance ou non de l'éliciteur aphidien** par la protéine **VAT-1** sont illustrés [Figure 8] montre une **discordance générale entre les profils AVR/VIR attendus et ceux donnés par le reséquençage** à partir d'extractions indépendantes. En effet, un tiers de l'ensemble des profils AVR/VIR attendus (4/12) n'ont pas été redécouverts, relativisant la pertinence du transcrit Iseq 55596 comme candidat à l'AVR. Toutefois et de manière intéressante, les profils AVR/VIR observés et attendus sont **100% concordants au sein de la race d'hôte des cucurbitacées**, représentant la moitié de la gamme d'hôte testée, où 5 clones sont AVR (C4, C9, NM1, CUC1 et GWD2) et un seul clone est VIR (C6). Les 6 autres clones testés sont spécialisés soit sur les malvacées, soit sur les solanacées. Parmi eux, 2 sont AVR

(Hib22 et Hib23) et 4 sont VIR (Burk, Aub1, Psp4, Pot1). Cependant, le reséquençage indique qu'ils ont tous les 6 le profil AVR. Sur la base de ce constat, il serait intéressant de **tester l'hypothèse que la virulence implique d'autres effecteurs salivaires chez les clones spécialisés sur d'autres familles botaniques**. Une tentative de validation du candidat d'AVR 55596 *in planta*, selon une stratégie proposée par Cécile Desbiez (INRAE, Pathologie végétale) apporterait un premier élément de réponse. Cette stratégie est présentée **plus en détail dans la partie Discussion générale et conclusion** de mon manuscrit de thèse.

## 5.5 Identification de candidats additionnels

Les différentes stratégies de recherche *in silico* **pourront être ré-exploitées afin d'identifier des candidats additionnels**. En effet, nous avons pensé et implémenté mes pipelines en hiérarchisant les différentes étapes des chaînes de traitement, en fonction de la pertinence des hypothèses biologiques qu'elles représentent au travers de leurs filtres associés : dans chaque stratégie (*i.e.* chaque chaîne de traitement) ces étapes sont **organisées par pertinence décroissante**. De cette façon, le retrait du dernier filtre d'une chaîne de traitement, correspondant donc à **la levée d'une contrainte**, va refavoriser la sensibilité de candidats au détriment de la précision d'identification et donc permettre de **produire de nouvelles listes de candidats** en conservant **la pertinence** au regard des hypothèses testées.

**Dans la stratégie 1**, présentée en **figures 18 et 19**, un nombre de candidats conséquent a été identifié. Dans mes travaux, j'ai eu recours à des **programmes de prédiction** de motifs d'adressage à l'exportation par voie salivaire pour réduire ces listes à un ensemble restreint de candidats putativement salivaires. Alternativement, la substitution de ces programmes de prédiction par la base de données salivaires comme crible de preuves de sécrétion confèrerait un double avantage pour la sélection de candidats pertinents : **(i)** gagner en fiabilité quant à la qualification de nature salivaire/sécrété et **(ii)**, écarter les faux positifs liée aux problèmes d'assemblages de transcriptomes *de novo* à partir de données RNAseq. Concernant la recherche de candidats 'de non expression', **85 candidats** identifiés dans le 'set total' **[Figure 18]** avaient été identifiés dans la base de données salivaires mais n'ont pas encore été testés **[Table 3]**. Concernant la recherche de candidats 'de polymorphisme', **37 candidats** avaient été identifiés dans le 'set total' (dont 2 également dans la base de données salivaire) mais n'ont pas non plus été testés **[Table 3]**.

En outre, le retrait de l'étape de sélection des SNP communs aux deux clones AVR C4 et C9 **[Figure 19]** permettrait de **prendre en compte une diversité allélique dans l'AVR**, reflétant ainsi une

réalité biologique tout à fait plausible. Dans des analyses complémentaires, les 2623 CDS issus du filtre 'Recherche de SNP non synonymes' ont été traduits en protéines et blastés contre la base de données salivaire, révélant ainsi **178 candidats additionnels** [Table 3].

**Table 3** – Bilan des candidats d'AVR restant à confirmer

▲ : candidats déjà identifiés mais non traités durant ma thèse ; ■ : candidats additionnels identifiés à partir d'analyses complémentaires mobilisant la base de données salivaire ; ● : candidats identifiés à partir d'analyses complémentaires ne mobilisant pas la base de données salivaire.

Clones	Non expression	Polymorphisme	Epissage alternatif
Stratégie 1	85▲	37▲ + 178■	-
Stratégie 2	131●		-
Stratégie 3	0	21●	295▲

Dans la **stratégie 2**, présentée en **figure 20**, une première étape avait permis de réduire un jeu de données de départ de plus de 100 000 transcrits Iseq AVR à 227 transcrits codant pour des protéines communes aux deux clones AVR et sans homologue VIR exact, conditions minimums requises pour définir un candidat. Par la suite, sur la base de ces résultats d'alignement BLASTP, j'avais identifié **(i)** 46 protéines de C9 sans homologue chez le clone VIR C6 qui ont été traitées comme des candidats de type 'non expression' et **(ii)**, 48 protéines de C9 partageant un fort taux d'homologie avec une protéine VIR qui ont été traitées comme des candidats 'de type polymorphisme'. Cependant, cette stratégie, basée sur l'inférence de catégories de candidats (*i.e.* **(i)** ou **(ii)**) à partir de résultats d'alignement BLAST a présenté des limites. En effet, **131 protéines candidates** sur 227, soit plus de la moitié, n'ont été attribuées à aucune de ces deux catégories et n'ont par conséquent pas été traitées au cours de ma thèse bien qu'elles représentent **un intérêt potentiel en tant que candidats d'AVR**. Considérer ces 131 protéines candidates supplémentaires pour lesquelles des taux d'homologie variables avec des protéines VIR ont été identifiés constituerait l'opportunité de traiter une nouvelle catégorie de candidats [Table 3] présentant un nombre important de polymorphisme et ou des INDEL. Dans un premier temps, la discrimination des candidats de catégorie 'non expression' pourra être faite par amplification sur des extractions d'ARN de C4, C9 et C6. Le reséquençage des candidats non confirmés de cette manière devrait alors permettre de discriminer des candidats de 'polymorphisme'.

Dans la **stratégie 3**, présentée en **figure 23 (partie de droite)**, seuls les **SNP bialléliques** ont été pris en compte (pour des questions pratiques) pour définir les candidats de type 'polymorphisme'. Toutefois, la levée de cette restriction permettrait **de prendre en compte une diversité allélique dans l'AVR**, reflétant ainsi une réalité biologique tout à fait plausible. J'ai effectué des analyses

complémentaires en repartant de l'étape de filtrage du vcf brut [figure 23, partie de droite en haut]. Cette fois-ci, seuls les SNP multialléliques ont été sélectionnés et filtrés pour les mêmes critères de qualité (cf. page 91). De cette façon, 495 gènes représentant 543 SNP, ont été identifiés. Le codage du filtre logique pour le profil des SNP multialléliques définissant les candidats d'AVR étant particulièrement fastidieux, ce filtre n'a pas été mis en place. Ensuite, l'annotation des SNP avec Allmine a révélé 25 SNP dans des zones codantes et entraînant un changement au niveau protéique (24 non synonymes et 1 absence de codon stop), correspondant à **21 gènes d'intérêt supplémentaires** [Table 3] pour des tentatives de confirmation ultérieures sur la collection de clones du GAFL. Dans un premier temps, cette liste de gènes candidats supplémentaires pourra être confirmée *in silico* par la visualisation des données d'alignement par IGV, permettant de discriminer les SNP multialléliques des faux positifs.

Concernant la méthode de recherche des candidats avec des **événements d'épissage alternatifs (EEA)** développée dans la **stratégie 3** et présentée en **figure 25**, 295 candidats identifiés avec un exon surnuméraire VIR [Figure 3] n'ont pas été soumis à une tentative de confirmation sur la collection de clones. Compte tenu du nombre de candidats à tester et dans un souci d'efficacité, une solution serait d'amplifier ces candidats par PCR puis de séquencer les amplicons par NGS et de valider leur profil AVR-VIR *in silico*.

Le retrait de l'étape de **recherche d'homologues dans la base de données salivaire** permettrait d'obtenir 436 (465-29) candidats supplémentaires avec un exon surnuméraire spécifique à la forme AVR et 2610 (2905-295) candidats supplémentaires avec un exon surnuméraire spécifique à la forme VIR, en acceptant d'ignorer la caractéristique salivaire des candidats d'AVR [Figure 25]. Il est important de souligner que la **caractéristique salivaire des effecteurs aphidiens** constitue l'**hypothèse la plus forte** et la plus largement admise, et l'ignorer pose théoriquement problème pour identifier des candidats d'AVR pertinents.

Cependant les mécanismes de sécrétion des pucerons sont **loin d'être entièrement compris**. Par conséquent, la prédiction du caractère salivaire des gènes candidats avec une base de données salivaire **est limitée**. L'usage de la base de données salivaires, bien que cohérent pour réduire une liste de candidats d'AVR en vue d'une potentielle validation, constitue néanmoins **un risque de production de faux positifs ainsi que de faux négatifs**. Pour cette raison, le retrait de l'étape de recherche d'homologues dans la base de données salivaire n'a pas ici de caractère rédhibitoire et constitue **une opportunité d'identifier des candidats pertinents supplémentaires**.

Aussi, en raison de la difficulté d'identifier précisément des EEA, j'ai choisi de restreindre le profil des gènes candidat à l'AVR à ceux exprimant des isoformes dont l'exon attendu responsable de la transition AVR-VIR ne flanquait pas les régions UTR. Si ce profil de candidat avait le double avantage d'être l'un des EEA les plus répandus chez les animaux et d'être le plus simple à évaluer avec la méthode que j'ai implémentée, deux autres EEA assez répandus mériteraient également d'être considérés : le choix du site d'EA 5', et le choix du site EA 3' (Nilsen & Graveley, 2010). Préalablement à cela, il faudra retravailler le jeu de données Isoseq pour s'assurer que les transcrits sont tous reconstitués dans leur intégralité, les deux EEA précédemment évoqués impliquant les UTR.





## **Discussion générale et conclusion**

## 1. Vue d'ensemble

Mes travaux de thèse avaient **pour objectif d'explorer la diversité potentielle des voies d'adaptation d'*A. gossypii* aux Cucurbitacées ainsi qu'au gène NLR *Vat***. Dans cette ultime partie, Je propose une relecture des deux chapitres qui font l'objet de ma thèse selon 3 axes : **(i)** l'adaptation d'*A. gossypii* aux cucurbitacées, centrée sur le phloème, **espace clé** dans les interactions plante-puceron, **(ii)** le déclenchement de la résistance *Vat-1* et **(iii)**, l'adaptation à la résistance *Vat-1*. A cette occasion, j'exposerai les principales avancées apportées par mes travaux, les limites des différentes analyses que j'ai effectuées, et les perspectives qu'ouvrent mes travaux de thèse.

## 2. L'adaptation d'*A. gossypii* aux cucurbitacées

De manière générale, les pucerons doivent **manipuler les défenses de la plante** pour établir une relation compatible avec le phloème et ainsi s'alimenter durablement (Will et al., 2007). Remarquablement, les cucurbitacées possèdent **deux systèmes phloémiens** : le **phloème fasciculaire (FP)** est principalement impliqué dans le transport des sucres, tandis que le **phloème extra-fasciculaire (EFP)** est largement associé aux mécanismes de défense et à la signalisation en général (R Turgeon, 2016; B. Zhang et al., 2010).

La piqûre des pucerons, perçue par la plante comme une blessure, entraîne des dépôts de callose et de lignine dans les parois des cellules adjacentes au chemin des stylet, ainsi qu'une obstruction du phloème (Will et al., 2007, 2009). Une telle obstruction a également été observée après des blessures mécaniques dans le FP et l'EFP, mais plus tard et dans une moindre mesure concernant l'EFP (Peng & Walker, 2018; B. Zhang et al., 2010) soulevant la question de l'utilisation de l'EFP par la race hôte d'*A. gossypii* spécialisée sur les cucurbitacées comme source d'alimentation alternative. Cette question a fait l'objet du **chapitre I** de ma thèse. L'approche, développée et mise en œuvre au cours de ma thèse en partenariat avec la plateforme de métabolomique liée à l'Université d'Avignon, était centrée sur la **détection de métabolites** fortement associés à chacun des deux phloèmes du melon **dans les pucerons**. **Nous avons montré que les clones d'*A. gossypii* appartenant à la race hôte spécialisée sur les cucurbitacées exploitent le FP lorsqu'ils s'alimentent sur melon non-*Vat***.

Dans ce cadre, la stratégie des clones d'*Aphis gossypii* adaptés aux cucurbitacée rejoint très probablement le **paradigme de l'évolution pour les interactions plante-pathogène** (Jones & Dangl, 2006), reposant sur l'évolution d'effecteurs des agents pathogènes, qui annule la PTI (PAMP reconnus par des motifs apoplastiques de PRR de la plante). Cette manipulation des défenses de la plante par

les pucerons, au travers d'un répertoire de composés salivaires, correspond à **la sensibilité déclenchée par des effecteurs (ETS)** décrite dans le diagramme en Z (Dangl et al., 2013). Par conséquent, la mise en place de l'ETS doit permettre aux clones spécialisés sur les cucurbitacées de se nourrir dans le FP. Ainsi, de nombreux efforts de recherche se concentrent sur l'identification et la caractérisation **des effecteurs aphidiens qui sont responsables de l'ETS** (Mondal, 2020; Mutti et al., 2008; Y. Zhang, Liu, Francis, et al., 2022; Y. Zhang, Liu, Fu, et al., 2022) et doivent empêcher le colmatage du phloème en réponse aux blessures infligées par les tentatives de pénétration effectuées par les pucerons au moyen de leurs stylets.

Le séquençage haut débit du transcriptome de tout ou partie de la collection de clones du GAFL, de telle sorte que les trois races d'hôtes 'Cucurbitacées', 'Malvacées' et 'Solanacées' soient représentées, permettrait au travers d'analyses comparatives d'expression, d'identifier des gènes candidats au rôle d'effecteur dans l'adaptation aux cucurbitacées. Typiquement, le mapping des banques de reads RNAseq de chacun des clones sur le génome de référence de C9 permettra d'identifier des gènes spécifiquement exprimés chez les clones spécialisés chez les cucurbitacées ou présentant une combinaison de polymorphismes spécifiques de cette race d'hôte. De manière encourageante, les données RNAseq de têtes pour 4 clones de la race d'hôte des cucurbitacées (C4, C6, C9 et CUC1) sont déjà disponibles au GAFL. Pour la suite, les extractions d'ARN pour des clones des deux autres races d'hôte pourront cette fois être réalisées sur pucerons entiers, le procédé de collecte de têtes de pucerons développé et mis en œuvre (Leslie Dutartre en 2013-2014) étant long et fastidieux, et les analyses d'expression pourront être réduites aux gènes exprimés dans la tête des clones cités plus haut.

### 3. Le déclenchement de la résistance *Vat-1*

Le paradigme de l'évolution pour les interactions plantes-pathogènes (Jones & Dangl, 2006) intervient également dans le déclenchement de la résistance liée à un gène de type NLR. Dans ce cadre, une part importante de mon travail de thèse a été consacrée à la recherche de l'effecteur aphidien qui interagit avec la protéine NLR VAT-1, chez l'accension de melon résistante PI 161375 [**chapitre II**]. Dans cette optique, mes travaux ont consisté à faire des **analyses comparatives de données d'expression** de clones AVR et VIR afin d'établir des listes de protéines candidates à l'avirulence. J'ai implémenté un total de sept stratégies de recherche, reposant sur des données et combinaisons de données variées, pour évaluer trois hypothèses pour la transition AVR-VIR : **(i)** l'absence de transcrite chez le clone VIR, **(ii)** des polymorphismes spécifiques du clone VIR et **(iii)**, des événements d'épissage alternatif chez le clone VIR.

Au final, même si mes travaux n'ont pas abouti à l'identification de l'effecteur responsable du déclenchement de la résistance *Vat*, l'hypothèse résumée par le diagramme en Z sur laquelle ils reposaient **reste une hypothèse forte** pour expliquer le déclenchement de cette résistance. Aussi, **plusieurs orientations et pistes d'amélioration intéressantes se dessinent**. Ces dernières pourront être approfondies lors de recherches ultérieures : **(i)** l'exploitation d'un vivier de candidats, identifiés mais n'ayant pas fait l'objet de tentatives de confirmation lors de ma thèse ou issus de listes élargies (*cf.* partie 5.5) pourra s'inscrire dans une tentative de validation *in planta* **(ii)** la prise de recul sur les résultats obtenus a permis de faire émerger quatre nouvelles hypothèses pour expliquer la transition AVR-VIR, **(iii)**, l'établissement du protéome salivaire des clones AVR vs VIR en vue d'établir une stratégie de recherche de candidats plus performante et proche de la réalité biologique.

### 3.1 Objectif : validation *in planta* du gène d'AVR

Si aujourd'hui, des résistances aux pucerons héritables ont été décrites chez une quinzaine d'espèces cultivées (Catherine Dogimont et al., 2010), peu ont été formellement caractérisées **au niveau moléculaire** : le gène **NLR *Mi1-2*** chez la tomate qui confère la résistance à la fois à *Macrosiphum euphorbiae* ainsi qu'à d'autres hémiptères (Rossi et al., 1998) et aux nématodes à galles (Vos et al., 1998), le gène **NLR *Vat-1*** chez le melon qui confère la résistance à *A. gossypii* ainsi qu'aux virus qu'il transmet (Catherine Dogimont et al., 2014), le gène ***SL11*** chez *A. thaliana* qui confère une résistance à *Myzus persicae* (Kloth et al., 2017; Kloth & Kormelink, 2020) et le gène **NLR *Adnr1*** chez le blé qui confère la résistance à *Diuraphis noxia* (Nicolis & Venter, 2018). Également, deux autres gènes de résistance aux hémiptères ont été caractérisés **au niveau moléculaire**, **Bph14** et le gène **NLR *Bph26*** chez le riz qui confèrent la résistance à la cicadelle *Nilaparvata lugens* (Du et al., 2009; Tamura et al., 2014). Au total, quatre de ces gènes codent pour des protéines du récepteur de répétition riche en leucine (NLR) du site de liaison aux nucléotides, responsables de l'immunité déclenchée par des effecteurs de pucerons. Dans le cadre de cette famille de gènes, on attend qu'un effecteur d'AVR, introduit dans la plante via la salive, élicite la résistance. Encore aujourd'hui, **aucun gène d'AVR n'a été caractérisé chez les hémiptères**.

Mes travaux de thèse **ne m'ont pas permis d'identifier l'effecteur** qui élicite la résistance *Vat*. Toutefois, les stratégies de recherches bioinformatiques que j'ai implémentées pour produire des candidats d'AVR en considérant différentes hypothèses de transition AVR-VIR pourront **servir de base à un nouveau projet**. En effet, parmi les candidats que j'ai pu identifier, beaucoup n'ont pas été amplifiés sur la collection de clones au cours de ma thèse.

Dans la prolongation de mes travaux, une démarche de **validation *in planta* du gène d'AVR** a été pensée par mon équipe, pour tenter d'identifier le tout premier couple effecteur d'AVR / protéine NLR chez les pucerons. Dans cette démarche, les protéines candidates d'AVR seront exprimées *in planta* via l'utilisation d'un **vecteur d'expression viral**, ce dernier étant **disponible au laboratoire de Pathologie Végétale**. Les CDS correspondant aux protéines candidates seront intégrés dans un clone viral, puis ces constructions chimériques seront inoculées par biolistique à des melons transgéniques contenant l'allèle *Vat-1*. Si la protéine candidate **est bien l'éliciteur de la résistance (AVR)**, nous attendons **une absence de multiplication du virus dans les melons *Vat*** (*i.e.* absence de symptômes foliaires) mais se produira dans les melons non-*Vat*. A l'inverse, si le facteur d'AVR est absent, une multiplication du virus est attendue sur les melons *Vat* et non-*Vat*.

Faute de candidat ayant un profil AVR-VIR 100% concordant pour les douze clones de la collection du GAFL, cette étape de validation fonctionnelle n'a **pas été initiée durant ma thèse**. Toutefois, j'ai pu identifier **un candidat, '55596'**, pour lequel les profils AVR/VIR observés et attendus sont 100% concordants au sein de la race d'hôte des cucurbitacées (soit représentant la moitié de la gamme d'hôte testée). **Il serait intéressant d'éprouver la méthode de validation *in planta* imaginée par mon équipe avec ce candidat.**

### 3.2 D'autres hypothèses pour expliquer la transition AVR-VIR ?

Le fait que la **stratégie 3**, bien qu'étant **classique**, n'ait pas permis d'identifier des candidats fiables **me conduit à penser que mon jeu d'hypothèses de départ n'était pas complet**, bien que reflétant les cas les plus largement étudiés. Dans mes travaux, j'ai confronté le paradigme des interactions plantes-pathogènes à la relation plantes-pucerons et mes réflexions étaient strictement circonscrites dans ce cadre. Toutefois, d'autres réalités biologiques, **en accord ou non avec ce paradigme**, méritant d'être explorées ne sont pas couvertes par le jeu d'hypothèses de travail qui ont fait l'objet de ce chapitre. Dans cette section, je propose **quatre hypothèses additionnelles pour expliquer la transition AVR-VIR**, exposées par pertinence décroissante à la lumière des connaissances actuelles : **(i)** la rétention de l'effecteur salivaire d'AVR dans l'acrostyle chez les clones VIR empêche son introduction dans la plante et donc sa reconnaissance par la protéine cytoplasmique NLR VAT-1 (Boualem et al., 2016; Deshoux et al., 2022), **(ii)** la surproduction d'un facteur de VIR bloque la réponse immunitaire déclenchée par un effecteur d'AVR (ETI) (da Cunha et al., 2007; Laugé et al., 1998; Ste-Croix et al., 2021), **(iii)** la régulation de la présence de séquences 'signals' change l'adressage subcellulaire de facteurs d'AVR (Göhre et al., 2013) et **(iv)**, des modifications post-traductionnelles du

facteur d'AVR entraînent sa perte de reconnaissance (Anderson et al., 2006; Lei et al., 2020; Yeam et al., 2010).

#### **HYPOTHESE (i) : RETENTION DE L'EFFECTEUR SALIVAIRE D'AVR DANS L'ACROSTYLE.**

Une première hypothèse intéressante à évaluer, car étant un dérivatif au paradigme de Dangl, et pouvant donc refléter une diversité potentielle des voies d'adaptation des pucerons à une résistance de type NLR, serait une **implication de l'acrostyle** dans la détermination du rôle AVR ou VIR d'un effecteur salivaire. Les travaux de recherches menés par Deshoux et al., 2022 ont montré, chez *M. persicae* et *A. pisum*, **la capacité de rétention de l'effecteur aphidien salivaire Mp10 par l'acrostyle**. L'acrostyle est une structure anatomique cuticulaire située à l'extrémité du canal commun des stylets maxillaires des pucerons (Uzest et al., 2010) **à l'interface avec la lumière du phloème** lorsqu'une interaction compatible plante-puceron est établie. Cette conclusion est étayée par un corps de preuves indirectes obtenues par des approches différentes et dont les résultats convergent: **(i)** la démonstration de la liaison spécifique de Mp10 à l'acrostyle par microscopie à épifluorescence, **(ii)** la démonstration de l'affaiblissement de cette liaison effecteur-acrostyle au moyen de l'interférence par ARN dirigée contre la protéine cuticulaire Stylin-03 dont l'interaction spécifique avec l'effecteur puceron Mp10 a été montré chez la levure et **(iii)**, la perturbation du comportement alimentaire des pucerons ayant préalablement ingéré des petits ARN interférents inhibant la production de la protéine Stylin-03. Sur la base de leurs résultats, Deshoux et al., 2022 proposent **une interprétation inédite** de l'avantage de la rétention de Mp10 par l'acrostyle pour le puceron qui est en lien avec mes travaux de thèses. De façon originale, l'acrostyle pourrait avoir un **rôle dans la modulation du flux d'effecteur Mp10**. L'accumulation de Mp10 à la pointe des stylets serait suivie d'un relâchement dans les tissus végétaux **sous forme concentrée**, ce qui conférerait à cet effecteur son potentiel d'action favorable au puceron (Deshoux et al., 2022). Sur la base de ces résultats, je propose une **nouvelle hypothèse pour la transition AVR-VIR** où la protéine d'AVR est exprimée, et ce, de manière équivalente chez les clones AVR et VIR et où **le niveau de concentration de cette protéine dans la salive sécrétée dans les cellules végétales, modulé par des protéines de l'acrostyle, détermine son rôle AVR ou VIR**. En accord avec Boissot et al., 2016, cette nouvelle hypothèse suggère que, dans certains cas, la résistance au virus induite par la reconnaissance d'un facteur aphidien pourrait être partielle (Boissot, Thomas, et al., 2016). Le matériel disponible au GAFL (ressources de bioinformatique) **ne nous permet pas de tester une telle hypothèse**. Pour poursuivre dans cette direction, il faudrait dans un premier temps établir **un catalogue des protéines salivaires** en interaction avec les différentes **protéines Stylin** de l'acrostyle. **Un début de piste** pourrait consister à étudier **dans la levure** l'interaction spécifique d'une ou plusieurs protéines cuticulaires Stylin avec des protéines salivaires d'*A. gossypii* préalablement collectées sur milieu artificiel.

**HYPOTHESE (ii) : SURPRODUCTION D'UN FACTEUR DE VIR BLOQUANT L'ETI.** Une seconde hypothèse, intégrant cette fois le modèle de référence pour les interactions plantes-bioagresseurs, consisterait en une surproduction d'un facteur de VIR bloquant l'ETI, le facteur d'AVR quant à lui étant exprimé de manière équivalente chez les clones AVR et VIR. **Les trois hypothèses** sur lesquelles je me suis concentré dans le présent chapitre présentaient **une même vision** de la mécanique de reconnaissance d'un facteur aphidien par un gène NLR. Dans cette vision, **la transition AVR-VIR consiste en l'absence d'une protéine AVR** impliquée dans une réponse immunitaire déclenchée par l'effecteur (ETI). En d'autres termes, dans cette vision, la présence d'une protéine AVR était incompatible avec la mise en place de l'ETS. Ce cas de figure s'inscrit bien le cadre des interactions plante-pathogène mais n'exploite pas toutes les possibilités offertes par le diagramme en Z (Jones & Dangl, 2006). En effet, toujours en accord avec ce diagramme, **la protéine AVR pourrait être présente** et c'est alors **une protéine additionnelle, VIR**, qui serait responsable du **blocage de l'ETI** (da Cunha et al., 2007; Laugé et al., 1998; Ste-Croix et al., 2021). Sur la base de cette réflexion, je propose une **nouvelle hypothèse pour la transition AVR-VIR** où la protéine d'AVR est exprimée, et ce, de manière équivalente chez les clones AVR et VIR et où une protéine spécifique du clone VIR C6 est surproduite, **bloquant l'ETI et rétablissant l'ETS**. L'évaluation de cette hypothèse se baserait sur des **analyses d'expression différentielle de gènes**. Actuellement, les données d'expression issues du séquençage d'ARN de têtes de pucerons constituent une seule répétition biologique, or les analyses d'expression différentielle en requièrent plusieurs. Par conséquent, **cette hypothèse n'est actuellement par évaluable** et d'autres extractions d'ARN de têtes de pucerons, bien que fastidieuses, permettraient de lever ce verrou.

**HYPOTHESE (iii) : RETENTION DU FACTEUR D'AVR DANS LES GLANDES SALIVAIRES.**

Dans l'état de l'art sur les effecteurs aphidiens présenté en **partie 3.3 de l'introduction** de mon manuscrit ainsi que dans la **partie 'Hypothèses de travail'** du présent chapitre, et en accord avec le paradigme largement admis pour les interactions plante-puceron, j'avais clairement défini qu'un effecteur aphidien était **sécrété par les glandes salivaires** dans la lumière du canal salivaire par un mécanisme d'adressage à la sécrétion, puis **injecté dans les cellules végétales**, et ce, qu'il soit AVR ou VIR. **Alternativement**, je propose que la transition AVR-VIR soit gouvernée par la sécrétion ou la rétention du facteur d'AVR. Dans cette nouvelle hypothèse (Göhre et al., 2013), la forme AVR serait injectée dans le cytoplasme des cellules végétale où la reconnaissance avec la protéine NLR VAT-1 a lieu, et la forme VIR correspondrait à une **rétention du facteur AVR dans les glandes salivaires du puceron**. Dans cet axe de réflexion, l'adressage des protéines vers leur destination subcellulaire est

connu pour être lié à la **présence de séquences 'signal'**. Celles-ci peuvent être régulées, notamment par un épissage alternatif du pré-ARNm (Göhre et al., 2013). Dans la mesure où les mécanismes de sécrétion chez les insectes piqueurs-suceurs semblent reposer sur des **motifs moléculaires conservés** (Bos et al., 2010), il est également envisageable que l'assignation des protéines à des voies de sécrétion puisse être régulée par des **polymorphismes** dans la séquence signal. **Cette hypothèse a en fait été testée** puisqu'elle implique un polymorphisme non synonyme au début d'une séquence protéique.

#### **HYPOTHESE (iv) : MODIFICATION POST-TRADUCTIONNELLE DU FACTEUR D'AVR.**

Cette quatrième hypothèse s'inscrit **dans un paradigme éloigné de mon cadre d'étude**, où les effecteurs ont évolué pour 'imiter' des substrats d'enzymes eucaryotes et sont, par conséquent, modifiés après leur traduction, dans la cellule hôte. Dans ce cadre, des études suggèrent que la VIR et l'AVR d'effecteurs bactériens de type III dépendent de modifications post-traductionnelles médiées par l'hôte lorsque ces effecteurs sont injectés dans la plante (Anderson et al., 2006; Lei et al., 2020; Yeam et al., 2010). Actuellement, rien de tel n'a été décrit chez des insectes et par conséquent **la validité de cette hypothèse est non probable**. Cependant, il me semblait pertinent d'évoquer ce cas de figure, **puisque qu'il ne peut pas être non plus écarté**.

### **3.3 Identification de protéomes salivaire AVR/VIR : vers une nouvelle stratégie de recherche de candidats**

Une autre part importante de mon travail, faisant l'objet du **chapitre II** de ma thèse, concernait l'implémentation d'une base de données de protéines salivaire **pour se substituer aux programmes de prédiction** de motifs d'adressage à l'exportation par voie salivaire qui sont traditionnellement utilisés dans les études sur les pucerons. En effet, **leur valeur sélective est limitée** pour les pucerons, générant un taux important de faux positifs ainsi que de faux négatifs. Une **première version** (2014) d'une telle base de données était disponible dans mon équipe à mon arrivée en thèse. Celle-ci comportait une liste de protéines exprimées dans les glandes salivaires de *Myzus persicae* (315) et d'*Acyrtosiphon pisum* (181). J'ai implémenté **une seconde version** (2020) de cette base de données. Cette nouvelle version représente un **enrichissement important** avec des séquences salivaires, ou putativement salivaires (*i.e.* identifiées à partir de glandes salivaires), identifiées entre 2014 et 2020. Aussi, cette seconde version comprend **trois espèces de pucerons additionnelles** (*i.e.* *M. euphorbiae*, *D. noxia* et *S. graminum*) et a été élargie aux *Delphacidae*, avec *N. lugens*, une référence pour l'étude du protéome salivaire chez les hémiptères. Encore, cette base de donnée dans sa nouvelle version inclus des effecteurs salivaires notables, dont **Mp1** chez *M. persicae* (Rodriguez et al., 2014, 2017),



**Armet** chez *A. pisum* (Wang, Dai, et al., 2015), **MIF1** chez *A. pisum* et *M. persicae* (Naessens et al., 2015) et, **Me10** chez *M. euphorbiae* (Atamian et al., 2013) parmi les plus étudiés.

L'usage de cette base de données a permis de **réduire les listes candidats produites tout en conservant la caractéristique salivaire** des effecteurs aphidiens. Cette caractéristique constitue l'hypothèse la plus forte et la plus largement admise, cependant, les données disponibles ne permettaient pas de l'implémenter avec fiabilité *in silico*. Par conséquent, j'ai géré le risque associé à cette imprécision en positionnant le filtre 'salivaire' en fin de chaîne de traitement (*cf.* partie 5.4).

Pour y remédier, l'idéal serait de mener **des analyses comparatives de protéomes salivaires** de clones VIR et AVR. Ainsi, nous serions certains du caractère salivaire des protéines candidates. Une telle piste avait été envisagée par mon équipe avant mon arrivée en thèse. Cependant, cette entreprise avait été abandonnée en raison des quantités insuffisantes de la salive recueillie dans les milieux nutritifs artificiels. En raison de l'enjeu lié au succès de cette expérimentation, je pense qu'il est nécessaire de la reconsidérer. La mise en place d'un milieu nutritif incorporant du résorcinol (Chaudhary et al., 2015), un neurostimulant ayant fait ses preuves pour induire la salivation des pucerons, pourrait permettre de **lever ce verrou** et ainsi permettre de **collecter des quantités suffisantes de salive** pour procéder à une identification par spectrométrie de masse des protéomes de clones AVR et VIR. Typiquement, la salive est recueillie sur des milieux liquides nutritifs recouverts de parafilm où des pucerons engagés s'alimentent (Cherqui & Tjallingii, 2000; Miles, 1965). Il est ensuite possible de collecter séparément, sur la base de leur consistance respective, la salive aqueuse contenant putativement des effecteurs et celle gélifiée (responsable de la formation de la gaine salivaire) et ainsi d'avoir un premier filtrage des échantillons avant même la détection par spectrométrie de masse. Disposer d'une telle ressource serait un atout majeur dans la continuité de ma thèse, pour identifier des candidats d'AVR autrement plus pertinents mais aussi pour enrichir et améliorer la qualité des annotations du génome de référence de C9.

## 4. Le contournement de la résistance *Vat-1*

### 4.1 L'EFP n'est pas impliqué dans le contournement de *Vat-1*

L'étude de la diversité des interactions *Vat/A. gossypii* où la résistance au CMV déclenchée par les pucerons a révélé l'existence d'un contournement de *Vat-1* par les pucerons. Dans le cadre de la résistance *Vat*, trois catégories **de clones spécialisés sur les cucurbitacées** sont définis (Boissot, Thomas, et al., 2016) : **(i)** ceux (*i.e.* NM1, C9, etc.) qui élicitent la résistance *Vat-1* et ne colonisent pas

les plantes *Vat-1* [Figure 7.A.1], (ii) ceux (*i.e.* C6) qui n'élicitent pas la résistance *Vat-1* et colonisent les plantes [Figure 7.A.2] et (iii), ceux (*i.e.* GDW2) qui élicitent la résistance *Vat-1* mais colonisent les plantes *Vat-1* [Figure 7.B]. Cette troisième catégorie de clone **est dite contournante** vis-à-vis de la résistance *Vat-1*, alors déclenchée.

Remarquablement, la majorité des interactions melon/puceron observées reflétaient ce cas de figure, non explicable avec le modèle de Jones et Dangl, indiquant que **l'adaptation des pucerons à l'ETI par un polymorphisme dans un effecteur salivaire n'est pas la norme** (Boissot, Thomas, et al., 2016). Alternativement, Boissot et al., avait émis l'hypothèse du **rôle du phloème particulier des cucurbitacées dans l'adaptation d'*A. gossypii* à *Vat-1***. Cette hypothèse **originale** était **cohérente** avec la **concordance des temps évolutifs** entre les pucerons et les cucurbitacées ainsi qu'avec **l'intérêt nutritif potentiel** de la sève de l'EFP, riche en azote. Cette question scientifique a fait l'objet d'un article durant ma thèse [chapitre I]. Nous avons montré que **le contournement de la résistance *Vat-1* par le clone GDW2 n'impliquait pas l'usage de l'EFP**.

#### 4.1.1 Des effecteurs pour expliquer le contournement de *Vat-1* ?

De manière intéressante, la démonstration de la non implication de l'EFP dans l'adaptation de certains clones à la résistance *Vat-1* déclenchée [chapitre I] semble **replacer le rôle des effecteurs aphidiens au centre de la question de l'adaptation d'*A. gossypii* à *Vat-1***.

Aussi, la vision large du terme effecteur que j'avais proposée dans l'introduction de mon manuscrit a été considéré **uniquement du point de vue immunitaire [chapitre II]**, où la reconnaissance d'un effecteur par une protéine NLR est attendue et où l'aspect fonctionnel (ou non) de l'effecteur aphidien n'est pas requis pour l'issue de l'interaction. **Alternativement**, une vision des effecteurs aphidiens comme **atténuateur des réponses physiologiques caractéristiques de l'ETI** au profit des performances biologiques des pucerons mériterait d'être explorée.

Dans ce nouveau cadre de réflexion, il serait intéressant de considérer le contournement de la résistance *Vat-1* par des clones d'*A. gossypii* adaptés dans le cas où la résistance au virus transmis par pucerons est déclenchée (ETI). En opposition aux effecteurs d'AVR qui ont fait l'objet du chapitre II de ma thèse, **je propose ici la définition d'effecteur de contournement**, désignant un facteur aphidien d'origine salivaire dont la fonction est antagoniste aux voies de signalisation en aval du déclenchement de *Vat-1* par un effecteur d'AVR et permettant au puceron de compenser les effets délétères, et tout particulièrement le colmatage du FP, consécutifs à la mise en place de l'ETI.

#### 4.1.2 Portrait-robot d'un effecteur de contournement

La résistance *Vat* est associée à une variété de **réponses physiologiques caractéristiques de l'ETI** (Boissot, Schoeny, et al., 2016). Celles-ci comprennent notamment un pic de production de médiateurs physiologiques tels que **des espèces réactives de l'oxygène et des ions calciques** auxquels sont typiquement attribués les **dépôts de callose et de lignine** dans les parois des cellules sondées par les stylets des pucerons et aussi dans les cellules adjacentes non blessées grâce à des vagues de potentiel calcique (Villada et al., 2009).

Dans le cadre de la résistance *Vat-1*, et **lorsque cette résistance est déclenchée**, les réponses physiologiques du point de vue du virus, et du puceron sont **couplées** (Boissot, Thomas, et al., 2016) : la résistance est déclenchée par un effecteur d'AVR salivaire, d'origine aphidienne, et révélée par l'ETI. Sur la base de ces connaissances, un effecteur 'de contournement par le puceron' **ne doit pas être impliqué dans l'aspect 'virus' de la réponse associée à *Vat-1*** (*i.e.* blocage et destruction rapide de la charge virale à l'échelle de la cellule piquée), puisque l'ETI est déclenchée par un effecteur d'AVR, mais au contraire, est **attendu responsable de l'établissement et du maintien durable de la connexion stylet-phloème**. Aussi, puisque l'EFP n'est probablement pas impliqué dans le contournement de *Vat-1*, il me semble raisonnable d'écarter les facteurs salivaires aphidiens liés à la détoxification de métabolites secondaires de cette définition d'effecteur de contournement.

D'après les résultats apportés par nos recherches faisant l'objet du **chapitre I**, seul le **FP** pourrait être concerné. De ce point de vue, **l'interférence avec la signalisation calcique** de la plante inhibant les dépôts de callose serait une fonction candidate de premier choix pour un effecteur de contournement. De manière similaire, la **suppression du colmatage des tubes criblés** du FP pourrait faciliter l'ingestion continue de la sève par les pucerons (Hattori et al., 2012; Walker, 2021). A ce titre, la fonction de **chélation d'ions calciques** est souvent mise en avant dans la littérature sur les pucerons et autres hémiptères associée à ce domaine de recherche (Carolan et al., 2009; N. Cui et al., 2019; Tian et al., 2021; Wang, Dai, et al., 2015; Will et al., 2007). De manière intéressante, **Armet**, l'un des effecteurs aphidien les mieux étudiés à ce jour, a été identifié comme protéine chélatrice d'ions calciques putative (N. Cui et al., 2019; Wang, Dai, et al., 2015). Cependant, contrairement au profil d'effecteur d'AVR dans lequel sa fonction ne présage pas de sa capacité à déclencher la résistance, **l'aspect fonctionnel est primordial** pour l'identification d'effecteurs candidats de contournement. Par conséquent, il est au préalable nécessaire de produire **des annotations structurales et fonctionnelles solides** pour pouvoir envisager la recherche d'effecteurs de contournement.

## 5. Conclusion de ma thèse

Le projet qui est par la suite devenu l'objet de ma thèse est né de l'intérêt commun des deux entreprises internationales spécialisées dans la sélection de semences potagères (ENZA ZADEN et RIJK ZWANN). Il s'inscrit dans **les objectifs de l'Unité GAFL** qui concernent en particulier l'étude des mécanismes d'adaptation des pucerons, thème fondateur pour la **gestion durable des résistances**. L'objet biologique retenu, le couple *Aphis gossypii*/melon, permet d'aborder de façon originale, **l'adaptation à deux niveaux** : à la résistance non hôte et à la résistance hôte conférée par un gène NBS-LRR.

Mes travaux de thèse, qui font l'objet du présent manuscrit, se situent à **la croisée de disciplines variées** et ont réuni **l'expertise d'une dizaine d'acteurs de recherche** en interactions plante-bioagresseurs, bioinformatique, biologie moléculaire, physiologie végétale et métabolomique. C'est la **transdisciplinarité** qui a fait l'originalité de ce projet et qui nous a permis de nous saisir de la question des voies d'adaptations à une résistance majeure **dans sa complexité** et en **testant les limites du cadre de réflexion** faisant référence, formalisé par le **diagramme en Z**. Ma thèse a donné lieu à **une publication** qui, pour la première fois, prend en compte simultanément la résistance NLR VAT-1, l'existence de clones contournant et la complexité structurale du phloème du melon. Bien que mes travaux n'aient  **finalement pas abouti à l'identification de l'effecteur** en interaction avec *Vat-1*, ils ouvrent **des perspectives intéressantes** pour la suite et posent notamment la question de la vraie nature de l'effecteur interagissant avec *Vat-1*.

L'objectif de mes travaux de thèse étaient de la **produire des connaissances** sur les interactions plantes-pucerons. Toutefois, il est envisageable qu'ils puissent aussi, **à terme**, bénéficier à **des objectifs finalisés** tel que l'identification de nouveaux **QTL complétifs de l'effet de *Vat-1***, comme un petit nombre de QTL à effet faible l'a déjà été, et ainsi élargir la gamme de clones d'*A. gossypii* auxquelles des variétés élites sont résistantes. L'identification de l'effecteur aphidien qui interagit avec *Vat-1* sera au même titre une source importante de connaissances inédites du point de vue des interactions plante-puceron mais pourra également converger avec **des intérêts d'ordre agronomique** et dans des perspectives de **développement durable** en favorisant la résistance des plantes. Dans une réalité plus proche de mon sujet, les résultats de mes recherches pourraient **contribuer à la compréhension de la différenciation en race d'hôte** chez les pucerons.



# Annexes

## 1. Article : Tables supplémentaires

Les tables supplémentaires S1 à S8 sont disponibles sur Recherche Data Gouv :

<https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.57745/3LI3FJ>

Nom du fichier : Supplementary Tables.xlsx

Ci-dessous sont données les légendes de ces tables supplémentaires.

### Table S1 Quantification of seven phloem-associated metabolites in six melon lines

Stachyose, trehalose, sucrose, raffinose and galactinol were analysed by GC-MS and quantified as the area under the chromatogram peak (no units) corrected by leaf sample weight (g). Cucurbitacin B and Glycosylated Cucurbitacin B were analysed by LC-MS and expressed as nM.mg<sup>-1</sup>. PI 161375 and PI 482398 melons carry different *Vat* resistance alleles, Védrentais and Charentais T are susceptible to aphids. Their isogenic and near-isogenic lines with resistance donors are noted [Védrentais]<sup>R\_PI 161375</sup> and [Charentais]<sup>R\_PI 482398</sup>

### Table S2 Quantification of five phloem-associated metabolites in healthy versus aphid-infested melon plants

Védrentais and its isogenic [Védrentais]<sup>R\_PI 161375</sup> plants were infested for 17h by NM1 aphid clone - triggering the *Vat*-mediated resistance- in both experiments, in experiment **(1)** by GWD2 aphids - triggering too the *Vat*-mediated resistance- and in experiment **(2)** by C6 aphids -no triggering the *Vat*-mediated resistance. Stachyose, sucrose and galactinol were analysed by GC-MS in melon leaves and expressed as peak area (no units) corrected by leaf sample weight (g). Cucurbitacin B and Glycosylated Cucurbitacin B were analysed by LC-MS and expressed as nMole.mg<sup>-1</sup>.

### Table S3 Quantification of seven metabolites in Okra and Védrentais melon plants

In Okra and Védrentais melon line. Sucrose, trehalose, galactinol, raffinose and stachyose were analysed by GC-MS and expressed as peak area/mg. Cucurbitacin B (Cb) and Glycosylated Cucurbitacin B (Cb-Glc) were analysed by LC-MS and expressed as nM/g.

### Table S4 Seven melon-associated metabolites in aphids during fasting

C6 adult aphids were reared on Védrentais melon plants, then fasted by batches of 20 to 50 individuals for 3h, 6h and 23h. t0 corresponds to the aphids sampled at rearing (before fasting). Stachyose, sucrose and galactinol were analysed by GC-MS in aphid batches and expressed as peak area (no units)

corrected by the aphid batch weight (mg). Cucurbitacin B and Glycosylated Cucurbitacin B were not detected by LC-MS.

**Table S5 Stachyose, sucrose and galactinol in three aphid clones fed on *Vat*/non-*Vat* melons**

Aphid clones were fed on Védraçais and its isogenic [Védraçais]<sup>R<sub>PI161375</sub></sup> melon lines for 17h. NM1 aphids -triggering the *Vat*-mediated resistance- were used in both experiments, with either GWD2 aphids -triggering the resistance in [Védraçais]<sup>R<sub>PI161375</sub></sup>- in experiment **(1)**, or C6 aphids -non-triggering the resistance in [Védraçais]<sup>R<sub>PI161375</sub></sup>- in experiment **(2)**. Results are ratios of [metabolite quantity in an aphid batch]/[metabolite quantity in the leaf on which they fed].

**Table S6 Quantification of stachyose, sucrose and galactinol in three aphid clones fed on *Vat*/non-*Vat* melon**

Aphid clones were fasted for 4h and then fed on Védraçais and its isogenic [Védraçais]<sup>R<sub>PI161375</sub></sup> melon lines for 17h. NM1 aphids -triggering the *Vat*-mediated resistance- were used in both experiments, with either GWD2 aphids -triggering the resistance in [Védraçais]<sup>R<sub>PI161375</sub></sup>- in experiment **(1)**, or C6 aphids -non-triggering the resistance in [Védraçais]<sup>R<sub>PI161375</sub></sup>- in experiment **(2)**. Stachyose, sucrose and galactinol were analysed by GC-MS in aphid batches and expressed as peak area (no units) corrected by the batch weight (mg).

**Table S7 Cucurbitacin B and Glycosylated Cucurbitacin B in aphids fasted and fed on *Vat*/non-*Vat* melons**

Aphids were fasted for 4h and then fed on Védraçais and its isogenic [Védraçais]<sup>R<sub>PI161375</sub></sup> melon lines for 17h. NM1 aphids -triggering the *Vat*-mediated resistance and GWD2 aphids -triggering the resistance in [Védraçais]<sup>R<sub>PI161375</sub></sup>. Results are expressed as peak area (no units) corrected by the aphid batch weight (mg).

**Table S8 Weight (mg) of three aphid clones fasted and fed on *Vat*/non-*Vat* melons**

Aphids were fasted for 4h and then fed on Védraçais and its isogenic [Védraçais]<sup>R<sub>PI161375</sub></sup> melon lines for 17h. NM1 aphids -triggering the *Vat*-mediated resistance- were used in both experiments, with either GWD2 aphids -triggering the resistance in [Védraçais]<sup>R<sub>PI161375</sub></sup>- in experiment **(1)**, or C6 aphids -non-triggering the resistance in [Védraçais]<sup>R<sub>PI161375</sub></sup>- in experiment **(2)**.



## 2. Scripts

### Script 1

**1/ Descriptif** : Elimination des séquences nucléotidiques/protéiques strictement identiques.  
**2/ Entrée** : un fichier FASTA avec des séquences dupliquées  
**3/ Sorties** : le fichier FASTA avec 100% de séquences uniques  
**4/ Complément** : lorsque des séquences sont redondantes, seul l'identifiant de la première séquence lue par le script sera conservé.

```
#!/bin/bash

### Usage: bash rmdup.sh [fastafilename]

### Build outname
f=$1
base=${f%.fa*}
brut=${base##*/}
outname=${brut}.rmdup.fasta

### Removes identical sequences from a fasta file
awk 'BEGIN{RS=">"; FS="\n"; ORS=""}
      (FNR==1){next}
      { name=$1; seq=$0; gsub(/^[^\n]*\n/, "", seq) }
      !(seen[seq]++){ print ">" $0 }' $1 > $outname
```

## Script 2

- 1/ Descriptif** : conservation du cadre de lecture le plus long pour chaque transcrit dans un fichier FASTA
- 2/ Entrées** : fichier FASTA généré par Transdecoder avec les différents cadres de lecture pour chaque transcrit
- 3/ Sortie** : fichier FASTA avec le cadre de lecture le plus long pour chaque transcrit
- 4/ Complément** : avant lancement, formater le fichier d'entrée avec: `sed '/^>/s/.\{3\}$//'`

```
#!/usr/bin/perl -w

my $line="";
my $seq="";
my $all_before="";
my %list=();
my $match_tmp="";
my $isnohit=0;
my $total_hits=0;
my $total_nohits=0;

$line=<>;
$seq=<>;
chomp($line);
$match_tmp=$line;
$match_tmp=~ s/_[0-9]{1,} .*$/g;
#print "$match_tmp";
$list{length($seq)}=$seq;

while($line=<>){
    if($line=~ /^>){
        $seq=<>;
        if($line =~ /$match_tmp/){
            $list{length($seq)}=$seq;
        }else{
            foreach my $max (sort { $b <=> $a } keys(%list) ){
                print "$match_tmp\n", $list{$max};
                last;
            }
            %list=();
            chomp($line);
            $match_tmp=$line;
            $match_tmp=~ s/_[0-9]{1,} .*$/g;
            $list{length($seq)}=$seq;
        }
    }
}

foreach my $max (sort { $b <=> $a } keys(%list) ){
    print "$match_tmp\n", $list{$max};
    last;
}

exit (0);
```

### Script 3

**1/ Descriptif** : filtrage d'une table BLAST produite avec le programme noblast (Lagnel et al., 2009).

**2/ Entrées** : **1** : fichier tabulé (txt) de résultats bruts de NOBLAST [noblast\_raw\_output\_to\_filter], **2** : sélection des N meilleurs alignement pour un hit donné [Hit\_rank], **3** : % d'alignement réciproque minimum entre la 'query' et la 'subject' [%\_Ali\_coverage\_for\_Q\_OR\_S], **4** : % d'identité minimum [%\_Identity], **5** : % d'identité 'relâché' entre acides aminés (deux acides aminés différents sont considérés « identiques » du point de vue s'ils ont des propriétés similaires) [%\_Positive], **6** : % de gaps maximum [%\_Gaps].

**3/ Sortie** : Fichier tabulé (txt) des résultats de BLAST filtré pour les paramètres d'entrée 2-6. Le nom du fichier de sortie est celui du fichier d'entrée avec, en suffixe, les paramètres 2-6 utilisés pour le filtrer.

**4/ Complément** : affichage en Shell d'un récapitulatif des paramètres d'entrée, du nom du fichier de sortie ainsi que du nombre d'éléments de la 'query', et de ceux de la 'subject', ayant un match.

```
#!/bin/bash

### Usage: bash noblast.filter.sh [noblast_raw_output_to_filter] [Hit_rank]
[%_Ali_coverage_for_Q_OR_S] [%_Identity] [%_Positive] [%_Gaps]

### In parameters
rawblast=$1
hit=$2
ali=$3
pid=$4
pos=$5
gap=$6

### Building outname
noext=${rawblast%.txt}
nopath=${noext##*/}
tmp=${nopath}.tmp
tmp2=${nopath}.tmp2
outname=${nopath}.hit$hit.ali$ali.pid$pid.pos$pos.gap$gap.txt

### AWK command for filtering
echo "***** Running *****"
awk -F'\t' -v hit=$hit -v ali=$ali -v pid=$pid -v pos=$pos -v gap=$gap '{if ($3<=hit &&
($12>=ali && $13>=ali) && $22>=pid && $24>=pos && $26<=gap) print
$5,$7,$9,$10,$11,$12,$13,$22,$24,$19,$26,$8}' $rawblast > $tmp

sed $'1
i\\nQ_Name\tSID\tQ_size\tS_size\tAli_size\t%Ali_Q\t%Ali_S\t%Identity\t%Positive\tBITscore\
t%Gaps\tS_name' $tmp > $tmp2
cat $tmp2 | awk '{for(i=1;i<12;i++){printf "%s\t",$i}};for(i=12;i<=NF;i++){printf "%s_",
$i}; printf "\n"}' | awk '{sub(/OS=.*/, ""); print}' | sed 's/.$//' > $outname
sed -i 's//g' $outname

### Print filtering report
echo "***** Report *****"
echo "====> Hit rank : $hit"
echo "====> % Ali coverage for Q and S : $ali"
echo "====> % Identity : $pid"
echo "====> % Positif : $pos"
echo "====> % Gaps : $gap"
echo "====> Outname : $outname"
echo "====> Nb of matching Q : `cat $outname | awk '{print $1}' | sort | uniq | wc -l`"
echo "====> Nb of matching S : `cat $outname | awk '{print $2}' | sort | uniq | wc -l`"
```

## Script 4

**1/ Descriptif** : Pipeline Snakemake pour le traitement de données RNAseq à l'aide d'un génome de référence. Des données fastq brutes au comptage de gènes et appel de SNP.

**2/ Entrées** : Données fastq brutes, génome de référence et son annotation structurale GFF

**3/ Sorties** : Tables de comptage des reads mappés sur les gènes, fichiers d'appel de SNP

**4/ Complément** : Ce script a été co-édité par Jacques Lagnel et Sonia Elbelt (GAFL, équipe I2B). J'ai en parti suivi sa construction lors de réunions de façon à m'appropriier son contenu. En raison de sa taille conséquente et dans un souci de clarté, ce script est uniquement disponible sur le git de l'Unité GAFL : [https://forgemia.inra.fr/gafl/pipelines\\_snakemake/rnaseq\\_de\\_snps](https://forgemia.inra.fr/gafl/pipelines_snakemake/rnaseq_de_snps). A soumettre sur le gestionnaire de travaux SLURM.

## Script 5

**1/ Descriptif** : filtrage du profil des SNP bialléliques pour les candidats d'AVR comme décrit dans la figure 11.

**2/ Entrées** : Fichier VCF à filtrer

**3/ Sortie** : Fichier VCF filtré

**4/ Complément** : A soumettre sur le gestionnaire de travaux SLURM

```
#!/bin/bash

export SINGULARITY_BINDPATH="/work2/project/gafl"
GATK='/work2/project/gafl/tools/containers/gatk4_v4.1.4.1.sif'
inputvcf=$1

# 1) Selection of variants: for C6 SNP homozygous non-reference (HomVar)
singularity exec $GATK gatk SelectVariants \
--variant $inputvcf \
-select '(vc.getGenotype("C6").isHomVar() && ((vc.getGenotype("C4").isHomRef() &&
(vc.getGenotype("C9").isHomRef() || vc.getGenotype("C9").isHet())) ||
(vc.getGenotype("C9").isHomRef() && (vc.getGenotype("C4").isHomRef() ||
vc.getGenotype("C4").isHet()))))' \
--output C6homoVar_candidates.vcf.gz

# 2) Selection of variants: for C6 SNP homozygous reference (HomRef)
singularity exec $GATK gatk SelectVariants \
--variant $inputvcf \
-select '(vc.getGenotype("C6").isHomRef() && ((vc.getGenotype("C4").isHomVar() &&
(vc.getGenotype("C9").isHomVar() || vc.getGenotype("C9").isHet())) ||
(vc.getGenotype("C9").isHomVar() && (vc.getGenotype("C4").isHomVar() ||
vc.getGenotype("C4").isHet()))))' \
--output C6homoRef_candidates.vcf.gz

# 3) Selection of variants: for C4-C9 Heterozygous (Het)
singularity exec $GATK gatk SelectVariants \
--variant $inputvcf \
-select '((vc.getGenotype("C6").isHomRef() || vc.getGenotype("C6").isHomVar()) &&
(vc.getGenotype("C4").isHet() && vc.getGenotype("C9").isHet()))' \
--output C6Het_candidates.vcf.gz

# 4) Combine vcf obtained in steps 1,2 and 3 above
singularity exec $GATK gatk MergeVcfs \
-I C6homoVar_candidates.vcf.gz \
-I C6homoRef_candidates.vcf.gz \
-I C6Het_candidates.vcf.gz \
-O C4C6C9_candidates.vcf.gz

# 4) Selection of variants: Keep SNP for C6
singularity exec $GATK gatk SelectVariants \
--variant ${name}_C4C6C9_candidates.vcf.gz \
--exclude-sample-name 'C9' \
--exclude-sample-name 'C4' \
--output C6_candidates.vcf.gz
```

## Script 6

**1/ Descriptif** : identification de gènes candidats d'AVR avec des exons surnuméraires chez les clones AVR ou VIR.

**2/ Entrées** : données Isoseq, génome de référence et son annotation structurale GFF, base de données salivaire 'database\_salivary\_proteins\_hemipterans' au format FASTA.

**3/ Sortie** :

**4/ Complément** : Ce script repose en grande partie sur l'utilisation de la suite logicielle Bedtools, dédiée à la comparaison, la manipulation et l'annotation des caractéristiques génomiques (Quinlan & Hall, 2010). En raison de sa taille conséquente et dans un souci de clarté, ce script est uniquement disponible sur le GitLab communautaire d'INRAE : <https://forgemia.inra.fr/>

Nom du projet :

[https://forgemia.inra.fr/pierre.sadon/agossypii\\_alternative\\_splicing\\_effector\\_candidates\\_identification](https://forgemia.inra.fr/pierre.sadon/agossypii_alternative_splicing_effector_candidates_identification)

Nom du script : ASE\_effector\_candidates\_identification.slurm



## **Références bibliographiques**



- Acevedo, F. E., Rivera-Vega, L. J., Chung, S. H., Ray, S., & Felton, G. W. (2015). Cues from chewing insects - the intersection of DAMPs, HAMPs, MAMPs and effectors. *Current Opinion in Plant Biology*, *26*, 80–86. <https://doi.org/10.1016/j.pbi.2015.05.029>
- Åhman, I., Kim, S. Y., & Zhu, L. H. (2019). Plant Genes Benefitting Aphids—Potential for Exploitation in Resistance Breeding. *Frontiers in Plant Science*, *10*(November), 1–14. <https://doi.org/10.3389/fpls.2019.01452>
- Allen, R. L., Bittner-Eddy, P. D., Grenville-Briggs, L. J., Meitz, J. C., Rehmany, A. P., Rose, L. E., & Beynon, J. L. (2004). Host-parasite coevolutionary conflict between *Arabidopsis* and downy mildew. *Science*, *306*(5703), 1957–1960. <https://doi.org/10.1126/science.1104022>
- An, D., Cao, H. X., Li, C., Humbeck, K., & Wang, W. (2018). Isoform sequencing and State-Of-Art applications for unravelling complexity of plant transcriptomes. *Genes*, *9*(1). <https://doi.org/10.3390/genes9010043>
- Anderson, J. C., Pascuzzi, P. E., Xiao, F., Sessa, G., & Martin, G. B. (2006). Host-mediated phosphorylation of type III effector AvrPto promotes *Pseudomonas* virulence and avirulence in tomato. *Plant Cell*, *18*(2), 502–514. <https://doi.org/10.1105/tpc.105.036590>
- Andrews, M. C., Callaghan, A., Field, L. M., Williamson, M. S., & Moores, G. D. (2004). Identification of mutations conferring insecticide-insensitive AChE in the cotton-melon aphid, *Aphis gossypii* Glover. *Insect Molecular Biology*, *13*(5), 555–561. <https://doi.org/10.1111/j.0962-1075.2004.00517.x>
- Atamian, H. S., Chaudhary, R., Cin, V. D., Bao, E., Girke, T., & Kaloshian, I. (2013). In planta expression or delivery of potato aphid *Macrosiphum euphorbiae* effectors Me10 and Me23 enhances aphid fecundity. *Molecular Plant-Microbe Interactions*, *26*(1), 67–74. <https://doi.org/10.1094/MPMI-06-12-0144-FI>
- Ayliffe, M., Singh, R., & Lagudah, E. (2008). Durable resistance to wheat stem rust needed. *Current Opinion in Plant Biology*, *11*(2), 187–192. <https://doi.org/10.1016/j.pbi.2008.02.001>
- Baker, D. A. B. (2000). Long-distance vascular transport of endogenous hormones in plants and their role in source-sink regulation. *Israel Journal of Plant Sciences*, *48*(3), 199–203. <https://doi.org/10.1560/QA6D-YP8C-DP8G-AG6K>
- Balch, R. E. (1952). Studies of the balsam woolly aphid *Adelges piceae* (Ratz.) and its effects on balsam fir, *Abies balsamea* (L.) Mill. . *Can. Dep. Agric. Pub.*, *867*, 1–76.
- Bass, C., Puinean, A. M., Zimmer, C. T., Denholm, I., Field, L. M., Foster, S. P., Gutbrod, O., Nauen, R., Slater, R., & Williamson, M. S. (2014). The evolution of insecticide resistance in the peach potato aphid, *Myzus persicae*. *Insect Biochemistry and Molecular Biology*, *51*(1), 41–51. <https://doi.org/10.1016/j.ibmb.2014.05.003>
- Betz, R., Walter, S., & Requena, N. (2016). Alternative splicing - an elegant way to diversify the function of repeat-containing effector proteins? *The New Phytologist*, *212*(2), 306–309. <https://doi.org/10.1111/nph.14157>
- Biała, A., Zess, E. K., De La Concepcion, J. C., Franceschetti, M., Pennington, H. G., Yoshida, K., Upson, J. L., Chanclud, E., Wu, C. H., Langner, T., Maqbool, A., Varden, F. A., Derevnina, L., Belhaj, K., Fujisaki, K., Saitoh, H., Terauchi, R., Banfield, M. J., & Kamoun, S. (2018). Lessons in effector and NLR biology of plant-microbe systems. *Molecular Plant-Microbe Interactions*, *31*(1), 34–45. <https://doi.org/10.1094/MPMI-08-17-0196-FI>
- Bigiard, J., Colcombet, J., & Hirt, H. (2015). Signaling mechanisms in pattern-triggered immunity (PTI). *Molecular Plant*, *8*(4), 521–539. <https://doi.org/10.1016/j.molp.2014.12.022>
- Blob, B., Heo, J. ok, & Helariutta, Y. (2018). Phloem differentiation: an integrative model for cell specification. *Journal of Plant Research*, *131*(1), 31–36. <https://doi.org/10.1007/s10265-017-0999-0>
- Boissot, N., Schoeny, A., & Vanlerberghe-Masutti, F. (2016). Vat, an amazing gene conferring resistance to aphids and viruses they carry: From molecular structure to field effects. *Frontiers in Plant Science*, *7*(September2016), 1–18. <https://doi.org/10.3389/fpls.2016.01420>
- Boissot, N., Thomas, S., Chovelon, V., & Lecoq, H. (2016). NBS-LRR-mediated resistance triggered by aphids: Viruses do not adapt; aphids adapt via different mechanisms. *BMC Plant Biology*, *16*(1),

- 1–12. <https://doi.org/10.1186/s12870-016-0708-5>
- Boller, T., & Felix, G. (2009). A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. *Annual Review of Plant Biology*, *60*(1), 379–406. <https://doi.org/10.1146/annurev.arplant.57.032905.105346>
- Bonaventure, G. (2018). Plants Recognize Herbivorous Insects by Complex Signalling Networks. In *Annual Plant Reviews online* (Vol. 47). <https://doi.org/10.1002/9781119312994.apr0505>
- Bos, J. I. B., Prince, D., Pitino, M., Maffei, M. E., Win, J., & Hogenhout, S. A. (2010). A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (green peach aphid). *PLoS Genetics*, *6*(11). <https://doi.org/10.1371/journal.pgen.1001216>
- Boualem, A., Dogimont, C., & Bendahmane, A. (2016). The battle for survival between viruses and their host plants. *Current Opinion in Virology*, *17*, 32–38. <https://doi.org/10.1016/j.coviro.2015.12.001>
- Boulain, H., Legeai, F., Guy, E., Morlière, S., Douglas, N. E., Oh, J., Murugan, M., Smith, M., Jaquiéry, J., Peccoud, J., White, F. F., Carolan, J. C., Simon, J. C., & Sugio, A. (2018). Fast evolution and lineage-specific gene family expansions of aphid salivary effectors driven by interactions with host-plants. *Genome Biology and Evolution*, *10*(6), 1554–1572. <https://doi.org/10.1093/gbe/evy097>
- Boulain, H., Legeai, F., Jaquiéry, J., Guy, E., Morlière, S., Simon, J. C., & Sugio, A. (2019). Differential Expression of Candidate Salivary Effector Genes in Pea Aphid Biotypes With Distinct Host Plant Specificity. *Frontiers in Plant Science*, *10*(October), 1–12. <https://doi.org/10.3389/fpls.2019.01301>
- Bradley, R. H. E., Sylvester, E. S., & Wade, C. V. (1962). Note on the Movements of the Mandibular and Maxillary Stylets of the Aphid, *Myzus persicae* (Sulzer). *The Canadian Entomologist*, *94*(6), 653–654. <https://doi.org/10.4039/Ent94653-6>
- Brant, E. J., & Budak, H. (2018). Plant small non-coding RNAs and their roles in biotic stresses. *Frontiers in Plant Science*, *9*(July), 1–9. <https://doi.org/10.3389/fpls.2018.01038>
- Cao, C. W., Zhang, J., Gao, X. W., Liang, P., & Guo, H. L. (2008). Overexpression of carboxylesterase gene associated with organophosphorous insecticide resistance in cotton aphids, *Aphis gossypii* (Glover). *Pesticide Biochemistry and Physiology*, *90*(3), 175–180. <https://doi.org/10.1016/j.pestbp.2007.11.004>
- Cao, H. H., Liu, H. R., Zhang, Z. F., & Liu, T. X. (2016). The green peach aphid *Myzus persicae* perform better on pre-infested Chinese cabbage *Brassica pekinensis* by enhancing host plant nutritional quality. *Scientific Reports*, *6*(February), 1–12. <https://doi.org/10.1038/srep21954>
- Cao, T. (2010). Metabolic engineering of raffinose-family oligosaccharides in the phloem reveals alterations in patterns of carbon partitioning and enhances resistance to green peach aphid. University of North Texas.
- Cao, T., Lahiri, I., Singh, V., Louis, J., Shah, J., & Ayre, B. G. (2013). Metabolic engineering of raffinose-family oligosaccharides in the phloem reveals alterations in carbon partitioning and enhances resistance to green peach aphid. *Frontiers in Plant Science*, *4*(JUL), 1–13. <https://doi.org/10.3389/fpls.2013.00263>
- Carletto, J., Lombaert, E., Chavigny, P., Brévault, T., Lapchin, L., & Vanlerberghe-Masutti, F. (2009). Ecological specialization of the aphid *Aphis gossypii* Glover on cultivated host plants. *Molecular Ecology*, *18*(10), 2198–2212. <https://doi.org/10.1111/j.1365-294X.2009.04190.x>
- Carlsbecker, A., & Helariutta, Y. (2005). Phloem and xylem specification: Pieces of the puzzle emerge. *Current Opinion in Plant Biology*, *8*(5), 512–517. <https://doi.org/10.1016/j.pbi.2005.07.001>
- Carolan, J. C., Caragea, D., Reardon, K. T., Mutti, N. S., Dittmer, N., Pappan, K., Cui, F., Castaneto, M., Poulain, J., Dossat, C., Tagu, D., Reese, J. C., Reeck, G. R., Wilkinson, T. L., & Edwards, O. R. (2011). Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrtosiphon pisum*): A dual transcriptomic/proteomic approach. *Journal of Proteome Research*, *10*(4), 1505–1518. <https://doi.org/10.1021/pr100881q>
- Carolan, J. C., Fitzroy, C. I. J., Ashton, P. D., Douglas, A. E., & Wilkinson, T. L. (2009). The secreted salivary proteome of the pea aphid *Acyrtosiphon pisum* characterised by mass spectrometry.

- Proteomics*, 9(9), 2457–2467. <https://doi.org/10.1002/pmic.200800692>
- Casteel, C. L., Walling, L. L., & Paine, T. D. (2006). Behavior and biology of the tomato psyllid, *Bactericera cockerelli*, in response to the Mi-1.2 gene. *Entomologia Experimentalis et Applicata*, 121(1), 67–72. <https://doi.org/10.1111/j.1570-8703.2006.00458.x>
- Cayla, T., Batailler, B., Hir, R. Le, Revers, F., Anstead, J. A., Thompson, G. A., Grandjean, O., & Dinant, S. (2015). Live imaging of companion cells and sieve elements in arabidopsis leaves. *PLoS ONE*, 10(2), 1–22. <https://doi.org/10.1371/journal.pone.0118122>
- Chaudhary, R., Atamian, H. S., Shen, Z., Briggs, S. P., & Kaloshian, I. (2014). GroEL from the endosymbiont *Buchnera aphidicola* betrays the aphid by triggering plant defense. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24), 8919–8924. <https://doi.org/10.1073/pnas.1407687111>
- Chaudhary, R., Atamian, H. S., Shen, Z., Briggs, S. P., & Kaloshian, I. (2015). Potato aphid salivary proteome: Enhanced salivation using resorcinol and identification of aphid phosphoproteins. *Journal of Proteome Research*, 14(4), 1762–1778. <https://doi.org/10.1021/pr501128k>
- Chaudhary, R., Peng, H. C., He, J., MacWilliams, J., Teixeira, M., Tsuchiya, T., Chesnais, Q., Mudgett, M. B., & Kaloshian, I. (2019). Aphid effector Me10 interacts with tomato TFT7, a 14-3-3 isoform involved in aphid resistance. *New Phytologist*, 221(3), 1518–1528. <https://doi.org/10.1111/nph.15475>
- Chen, Jian Q., Rahbe, Y., Delobel, B., Sauvion, N., Guillaud, J., & Febvay, G. (1997). Melon resistance to the aphid *Aphis gossypii*: behavioural analysis and chemical correlations with nitrogenous compounds. *Entomologia Experimentalis et Applicata*, 85(1), 33–44. <https://doi.org/10.1046/j.1570-7458.1997.00232.x>
- Chen, Jian Qun, Martin, B., Rahbé, Y., & Fereres, A. (1997). Early intracellular punctures by two aphid species on near-isogenic melon lines with and without the virus aphid transmission (Vat) resistance gene. *European Journal of Plant Pathology*, 103(6), 521–536. <https://doi.org/10.1023/A:1008610812437>
- Chen, Y., Singh, A., Kaithakottil, G. G., Mathers, T. C., Gravino, M., Mugford, S. T., van Oosterhout, C., Swarbreck, D., & Hogenhout, S. A. (2020). An aphid RNA transcript migrates systemically within plants and is a virulence factor. *Proceedings of the National Academy of Sciences*, 117(23), 201918410. <https://doi.org/10.1073/pnas.1918410117>
- Cherqui, A., & Tjallingii, W. F. (2000). Salivary proteins of aphids, a pilot study on identification, separation and immunolocalisation. *Journal of Insect Physiology*, 46(8), 1177–1186. [https://doi.org/10.1016/S0022-1910\(00\)00037-8](https://doi.org/10.1016/S0022-1910(00)00037-8)
- Cho, S. M., Kang, E. Y., Kim, M. S., Yoo, S. J., Im, Y. J., Kim, Y. C., Yang, K. Y., Kim, K. Y., Kim, K. S., Choi, Y. S., & Cho, B. H. (2010). Jasmonate-dependent expression of a galactinol synthase gene is involved in priming of systemic fungal resistance in *Arabidopsis Thaliana*. *Botany*, 88(5), 452–461. <https://doi.org/10.1139/B10-009>
- Choi, W.-G., Hilleary, R., Swanson, S. J., Kim, S.-H., & Gilroy, S. (2016). Rapid, Long-Distance Electrical and Calcium Signaling in Plants. *Annual Review of Plant Biology*, 67(1), 287–307. <https://doi.org/10.1146/annurev-arplant-043015-112130>
- Choi, W. G., Miller, G., Wallace, I., Harper, J., Mittler, R., & Gilroy, S. (2017). Orchestrating rapid long-distance signaling in plants with Ca<sup>2+</sup>, ROS and electrical signals. *Plant Journal*, 90(4), 698–707. <https://doi.org/10.1111/tpj.13492>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Ruden, D. M., & Lu, X. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster*. *June*, 1–13.
- Collier, S. M., & Moffett, P. (2009). NB-LRRs work a “bait and switch” on pathogens. *Trends in Plant Science*, 14(10), 521–529. <https://doi.org/10.1016/j.tplants.2009.08.001>
- Cook, D. E., Mesarich, C. H., & Thomma, B. P. H. J. (2015). Understanding Plant Immunity as a Surveillance System to Detect Invasion. *Annual Review of Phytopathology*, 53(1), 541–563. <https://doi.org/10.1146/annurev-phyto-080614-120114>
- Crafts, A. S. (1932). Phloem anatomy, exudation, and transport of organic nutrients in cucurbits.

*Plant Physiology*, 7(2), 1–2.

- Cronshaw, J., Area, S., & Contents, P. (1981). Phloem structure and function. *Ann. Rev. Plant Physiol.*
- Cui, L., Yin, W., Dong, S., & Wang, Y. (2012). Analysis of polymorphism and transcription of the effector gene *Avr1b* in *Phytophthora sojae* isolates from China virulent to *Rps1b*. *Molecular Plant Pathology*, 13(2), 114–122. <https://doi.org/10.1111/j.1364-3703.2011.00733.x>
- Cui, N., Lu, H., Wang, T., Zhang, W., Kang, L., & Cui, F. (2019). Armet, an aphid effector protein, induces pathogen resistance in plants by promoting the accumulation of salicylic acid. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1767). <https://doi.org/10.1098/rstb.2018.0314>
- da Cunha, L., Sreerekha, M. V., & Mackey, D. (2007). Defense suppression by virulence effectors of bacterial phytopathogens. In *Current Opinion in Plant Biology* (Vol. 10, Issue 4, pp. 349–357). <https://doi.org/10.1016/j.pbi.2007.04.018>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., Mcvean, G., Durbin, R., Project, G., & Vcf, T. (2018). *The variant call format and VCFtools*. 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Dangl, J. L., Horvath, D. M., & Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science*, 341(6147), 746–751. <https://doi.org/10.1126/science.1236011>
- Davies, P. J. (2010). Plant hormones: Biosynthesis, signal transduction, action! *Plant Hormones: Biosynthesis, Signal Transduction, Action!*, October, 1–802. <https://doi.org/10.1007/978-1-4020-2686-7>
- De Rybel, B., Mähönen, A. P., Helariutta, Y., & Weijers, D. (2016). Plant vascular development: From early specification to differentiation. *Nature Reviews Molecular Cell Biology*, 17(1), 30–40. <https://doi.org/10.1038/nrm.2015.6>
- De Schepper, V., De Swaef, T., Bauweraerts, I., & Steppe, K. (2013). Phloem transport: A review of mechanisms and controls. *Journal of Experimental Botany*, 64(16), 4839–4850. <https://doi.org/10.1093/jxb/ert302>
- Dedryver, C. A., Le Ralec, A., & Fabre, F. (2010). The conflicting relationships between aphids and men: A review of aphid damage and control strategies. *Comptes Rendus - Biologies*, 333(6–7), 539–553. <https://doi.org/10.1016/j.crv.2010.03.009>
- Deeken, R., Geiger, D., Fromm, J., Koroleva, O., Ache, P., Langenfeld-Heyser, R., Sauer, N., May, S. T., & Hedrich, R. (2002). Loss of the *AKT2/3* potassium channel affects sugar loading into the phloem of *Arabidopsis*. *Planta*, 216(2), 334–344. <https://doi.org/10.1007/s00425-002-0895-1>
- Denison, F. C., Paul, A. L., Zupanska, A. K., & Ferl, R. J. (2011). 14-3-3 Proteins in Plant Physiology. *Seminars in Cell and Developmental Biology*, 22(7), 720–727. <https://doi.org/10.1016/j.semcdb.2011.08.006>
- Deshoux, M., Monsion, B., Pichon, E., Jim, J., Cayrol, B., Thébaud, G., Mugford, S. T., Hogenhout, S. A., & Fereres, A. (2022). Role of Acrostyle Cuticular Proteins in the Retention of an Aphid Salivary Effector. *International Journal of Molecular Sciences*. <https://doi.org/https://doi.org/10.3390/ijms232315337>
- Dinant, S., Bonnemain, J. L., Girousse, C., & Kehr, J. (2010). Phloem sap intricacy and interplay with aphid feeding. *Comptes Rendus - Biologies*, 333(6–7), 504–515. <https://doi.org/10.1016/j.crv.2010.03.008>
- Dinant, S., & Lemoine, R. (2010). The phloem pathway: New issues and old debates. *Comptes Rendus - Biologies*, 333(4), 307–319. <https://doi.org/10.1016/j.crv.2010.01.006>
- Dobin, A., Gingeras, T. R., Spring, C., Flores, R., Sampson, J., Knight, R., Chia, N., & Technologies, H. S. (2016). Mapping RNA-seq with STAR. *Curr Protoc Bioinformatics*, 51(4), 586–597. <https://doi.org/10.1002/0471250953.bi1114s51.Mapping>
- Dogimont, C., Chovelon, V., Tual, S., Boissot, N., Rittener, V., Giovanazzo, N., & Bendahmane, A. (2008). Molecular diversity at the *Vat/Pm-W* resistance locus in melon. *Cucurbitaceae 2008*, 219–228.
- Dogimont, Catherine, Bendahmane, A., Chovelon, V., & Boissot, N. (2010). Host plant resistance to

- aphids in cultivated crops: Genetic and molecular bases, and interactions with aphid populations. *Comptes Rendus - Biologies*, 333(6–7), 566–573.  
<https://doi.org/10.1016/j.crvi.2010.04.003>
- Dogimont, Catherine, Chovelon, V., Pauquet, J., Boualem, A., & Bendahmane, A. (2014). The *Vat* locus encodes for a CC-NBS-LRR protein that confers resistance to *Aphis gossypii* infestation and *A. gossypii*-mediated virus resistance. *Plant Journal*, 80(6), 993–1004.  
<https://doi.org/10.1111/tpj.12690>
- Döring, T. F. (2014). How aphids find their host plants, and how they don't. *Annals of Applied Biology*, 165(1), 3–26. <https://doi.org/10.1111/aab.12142>
- Douglas, A. E. (2006). Phloem-sap feeding by animals: Problems and solutions. *Journal of Experimental Botany*, 57(4), 747–754. <https://doi.org/10.1093/jxb/erj067>
- Du, B., Zhang, W., Liu, B., Hu, J., Wei, Z., Shi, Z., He, R., Zhu, L., Chen, R., Han, B., & He, G. (2009). Identification and characterization of Bph14, a gene conferring resistance to brown planthopper in rice. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52), 22163–22168. <https://doi.org/10.1073/pnas.0912139106>
- Dubreuil, G., Deleury, E., Crochard, D., Simon, J. C., & Coustau, C. (2014). Diversification of MIF immune regulators in aphids: Link with agonistic and antagonistic interactions. *BMC Genomics*, 15(1), 1–12. <https://doi.org/10.1186/1471-2164-15-762>
- Eleftherianos, I., Vamvatsikos, P., Ward, D., & Gravanis, F. (2006). Changes in the levels of plant total phenols and free amino acids induced by two cereal aphids and effects on aphid fecundity. *Journal of Applied Entomology*, 130(1), 15–19. <https://doi.org/10.1111/j.1439-0418.2005.01017.x>
- Elsayed, G. (2011). Plant secondary substances and insects behaviour. *Archives of Phytopathology and Plant Protection*, 44(16), 1534–1549. <https://doi.org/10.1080/03235408.2010.507957>
- Elzinga, D. a, Vos, M. De, & Jander, G. (2014). Suppression of Plant Defenses by a *Myzus persicae* (Green Peach Aphid) Salivary Effector Protein. 27(7), 747–756. <https://doi.org/10.1094/MPMI-01-14-0018-R>
- Emanuelsson, O., Brunak, S., von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, 2(4), 953–971.  
<https://doi.org/10.1038/nprot.2007.131>
- Fiehn, O. (2003). Metabolic networks of *Cucurbita maxima* phloem. *Phytochemistry*, 62(6), 875–886.  
[https://doi.org/10.1016/S0031-9422\(02\)00715-X](https://doi.org/10.1016/S0031-9422(02)00715-X)
- Finlay, K. J., & Luck, J. E. (2011). Response of the bird cherry-oat aphid (*Rhopalosiphum padi*) to climate change in relation to its pest status, vectoring potential and function in a crop-vector-virus pathosystem. *Agriculture, Ecosystems and Environment*, 144(1), 405–421.  
<https://doi.org/10.1016/j.agee.2011.08.011>
- Flor, H. H. (1971). Current status of the gene-fob-gene concept. 275–296.
- Francis, F., Haubruge, E., Gaspar, C., & Dierickx, P. J. (2001). Glutathione S-transferases of *Aulacorthum solani* and *Acyrtosiphon pisum*: Partial purification and characterization. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 129(1), 165–171. [https://doi.org/10.1016/S1096-4959\(01\)00329-3](https://doi.org/10.1016/S1096-4959(01)00329-3)
- Francis, F., Vanhaelen, N., & Haubruge, E. (2005). Glutathione S-transferases in the adaptation to plant secondary metabolites in the *Myzus persicae* aphid. *Archives of Insect Biochemistry and Physiology*, 58(3), 166–174. <https://doi.org/10.1002/arch.20049>
- Fricaux-dutartre, L., Bouchet, J. P., Beucher, C., Boissot, N., & Dogimont, C. (2014). Identification of virulence gene candidates in the non-model species *Aphis gossypii* using a RNASeq approach (p. 84143).
- Fromm, J., & Lautner, S. (2007). Electrical signals and their physiological significance in plants. *Plant, Cell and Environment*, 30(3), 249–257. <https://doi.org/10.1111/j.1365-3040.2006.01614.x>
- Fudal, I., Ross, S., Brun, H., Besnard, A. L., Ermel, M., Kuhn, M. L., Balesdent, M. H., & Rouxel, T. (2009). Repeat-Induced Point Mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*. *Molecular Plant-Microbe Interactions*, 22(8), 932–941.

- <https://doi.org/10.1094/MPMI-22-8-0932>
- Fukuda, H., & Ohashi-Ito, K. (2019). Vascular tissue development in plants. In *Current Topics in Developmental Biology* (1st ed., Vol. 131). Elsevier Inc.  
<https://doi.org/10.1016/bs.ctdb.2018.10.005>
- Fukumorita, T., & Chino, M. (1982). Sugar, Amino Acid and Inorganic Contents in Rice Phloem Sap. *Plant and Cell Physiology*, 23(2), 273–283. <https://doi.org/10.1093/oxfordjournals.pcp.a076347>
- Fuller, S. J., Chavigny, P., Lapchin, L., & Vanlerberghe-Masutti, F. (1999). Variation in clonal diversity in glasshouse infestations of the aphid, *Aphis gossypii* Glover in southern France. *Molecular Ecology*, 8(11), 1867–1877. <https://doi.org/10.1046/j.1365-294X.1999.00782.x>
- Furch, A. C. U., Van Bel, A. J. E., & Will, T. (2015). Aphid salivary proteases are capable of degrading sieve-tube proteins. *Journal of Experimental Botany*, 66(2), 533–539.  
<https://doi.org/10.1093/jxb/eru487>
- Furk, C., Powell, D. F., & Heyd, S. (1980). Pirimicarb resistance in the melon and cotton aphid, *Aphis gossypii* Glover. *Plant Pathology*, 29(4), 191–196. <https://doi.org/10.1111/j.1365-3059.1980.tb01211.x>
- Gachon, C. M. M., Langlois-Meurinne, M., & Saindrenan, P. (2005). Plant secondary metabolism glycosyltransferases: The emerging functional analysis. *Trends in Plant Science*, 10(11), 542–549.  
<https://doi.org/10.1016/j.tplants.2005.09.007>
- Garzo, E., Soria, C., Gomez-Guillamon, M. L., & Fereres, A. (2002). Feeding behavior of *Aphis gossypii* on resistant accessions of different melon genotypes (*Cucumis melo*). *Phytoparasitica*, 30(2), 129–140. <https://doi.org/10.1007/BF02979695>
- Garzo, Elisa, Fernández-Pascual, M., Morcillo, C., Fereres, A., Gómez-Guillamón, M. L., & Tjallingii, W. F. (2018). Ultrastructure of compatible and incompatible interactions in phloem sieve elements during the stylet penetration by cotton aphids in melon. *Insect Science*, 25(4), 631–642.  
<https://doi.org/10.1111/1744-7917.12447>
- Gaupels, F., & Ghirardo, A. (2013). The extrafascicular phloem is made for fighting. *Frontiers in Plant Science*, 4(JUN), 2–5. <https://doi.org/10.3389/fpls.2013.00187>
- Gaupels, F., Sarioglu, H., Beckmann, M., Hause, B., Spannagl, M., Draper, J., Lindermayr, C., & Durner, J. (2012). Deciphering systemic wound responses of the pumpkin extrafascicular phloem by metabolomics and stable isotope-coded protein labeling. *Plant Physiology*, 160(4), 2285–2299.  
<https://doi.org/10.1104/pp.112.205336>
- Gilroy, S., Białasek, M., Suzuki, N., Górecka, M., Devireddy, A. R., Karpiński, S., & Mittler, R. (2016). ROS, calcium, and electric signals: Key mediators of rapid systemic signaling in plants. *Plant Physiology*, 171(3), 1606–1615. <https://doi.org/10.1104/pp.16.00434>
- Giordanengo, P., Brunissen, L., Rusterucci, C., Vincent, C., Van Bel, A., Dinant, S., Girousse, C., Faucher, M., & Bonnemain, J. L. (2010). Compatible plant-aphid interactions: How aphids manipulate plant responses. *Comptes Rendus - Biologies*, 333(6–7), 516–523.  
<https://doi.org/10.1016/j.crv.2010.03.007>
- Göhre, V., Haag, C., & Feldbrügge, M. (2013). RNA Biology in Fungal Phytopathogens. *PLoS Pathogens*, 9(10), 9–11. <https://doi.org/10.1371/journal.ppat.1003617>
- Goławska, S., Kapusta, I., & Łukasik, I. (2008). Effect of phenolics on the pea aphid, *Acyrtosiphon pisum* (Harris) population on *Pisum sativum* L. (Fabaceae). 71–77.
- Goławska, S., Łukasik, I., Goławski, A., Kapusta, I., & Janda, B. (2010). Alfalfa (*Medicago sativa* L.) apigenin glycosides and their effect on the pea aphid *Acyrtosiphon pisum*. *Polish Journal of Environmental Studies*, 19(5), 913–919.
- Goławska, S., Łukasik, I., & Leszczyński, B. (2008). Effect of alfalfa saponins and flavonoids on pea aphid. *Entomologia Experimentalis et Applicata*, 128(1), 147–153.  
<https://doi.org/10.1111/j.1570-7458.2008.00709.x>
- González, V. M., Aventín, N., Centeno, E., & Puigdomènech, P. (2014). Interspecific and intraspecific gene variability in a 1-Mb region containing the highest density of NBS-LRR genes found in the melon genome. *BMC Genomics*, 15(1), 1–15. <https://doi.org/10.1186/1471-2164-15-1131>
- Graherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L.,

- Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacoheh, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacoheh, N., Gnirke, A., Rhind, N., Di Palma, F., W. Birren, B., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644–652. <https://doi.org/10.1038/nbt.1883>. Trinity
- Gray, W. M. (2004). Hormonal regulation of plant growth and development. *PLoS Biology*, *2*(9). <https://doi.org/10.1371/journal.pbio.0020311>
- Ham, B.-K., & Lucas, W. J. (2017). Phloem-Mobile RNAs as Systemic Signaling Agents. *Annual Review of Plant Biology*, *68*(1), 173–195. <https://doi.org/10.1146/annurev-arplant-042916-041139>
- Haq, F. U., Ali, A., Khan, M. N., Shah, S. M. Z., Kandel, R. C., Aziz, N., Adhikari, A., Choudhary, M. I., Ur-Rahman, A., El-Seedi, H. R., & Musharraf, S. G. (2019). Metabolite Profiling and Quantitation of Cucurbitacins in Cucurbitaceae Plants by Liquid Chromatography coupled to Tandem Mass Spectrometry. *Scientific Reports*, *9*(1), 1–11. <https://doi.org/10.1038/s41598-019-52404-1>
- Haritatos, E., Keller, F., & Turgeon, R. (1996). Raffinose oligosaccharide concentrations measured in individual cell and tissue types in Cucumis melo L. leaves: Implications for phloem loading. *Planta*, *198*(4), 614–622. <https://doi.org/10.1007/BF00262649>
- Harmel, N., Létocart, E., Cherqui, A., Giordanengo, P., Mazzucchelli, G., Guillonéau, F., De Pauw, E., Haubruge, E., & Francis, F. (2008). Identification of aphid salivary proteins: A proteomic investigation of *Myzus persicae*. *Insect Molecular Biology*, *17*(2), 165–174. <https://doi.org/10.1111/j.1365-2583.2008.00790.x>
- Harvey, K. L., Jarocki, V. M., Charles, I. G., & Djordjevic, S. P. (2019). The diverse functional roles of elongation factor tu (Ef-tu) in microbial pathogenesis. *Frontiers in Microbiology*, *10*(OCT), 1–19. <https://doi.org/10.3389/fmicb.2019.02351>
- Hattori, M., Nakamura, M., Komatsu, S., Tsuchihara, K., Tamura, Y., & Hasegawa, T. (2012). Molecular cloning of a novel calcium-binding protein in the secreted saliva of the green rice leafhopper *Nephotettix cincticeps*. *Insect Biochemistry and Molecular Biology*, *42*(1), 1–9. <https://doi.org/10.1016/j.ibmb.2011.10.001>
- Hayafune, M., Berisio, R., Marchetti, R., Silipo, A., Kayama, M., Desaki, Y., Arima, S., Squeglia, F., Ruggiero, A., Tokuyasu, K., Molinaro, A., Kaku, H., & Shibuya, N. (2014). Chitin-induced activation of immune signaling by the rice receptor CEBiP relies on a unique sandwich-type dimerization. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(3). <https://doi.org/10.1073/pnas.1312099111>
- Hayashi, H., & Chino, M. (1986). Collection of pure phloem sap from wheat and its chemical composition. *Plant and Cell Physiology*, *27*(7), 1387–1393. <https://doi.org/10.1093/oxfordjournals.pcp.a077237>
- Heath, M. C. (2000). Hypersensitive response-related death. *Plant Molecular Biology*, *44*(3), 321–334. <https://doi.org/10.1023/A:1026592509060>
- Hedrich, R., Salvador-Recatalà, V., & Dreyer, I. (2016). Electrical Wiring and Long-Distance Plant Communication. *Trends in Plant Science*, *21*(5), 376–387. <https://doi.org/10.1016/j.tplants.2016.01.016>
- Heo, J. ok, Blob, B., & Helariutta, Y. (2017). Differentiation of conductive cells: a matter of life and death. *Current Opinion in Plant Biology*, *35*, 23–29. <https://doi.org/10.1016/j.pbi.2016.10.007>
- Hogenhout, S. A., & Bos, J. I. B. (2011). Effector proteins that modulate plant-insect interactions. *Current Opinion in Plant Biology*, *14*(4), 422–428. <https://doi.org/10.1016/j.pbi.2011.05.003>
- Hogenhout, S. A., Van Der Hoorn, R. A. L., Terauchi, R., & Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. *Molecular Plant-Microbe Interactions*, *22*(2), 115–122. <https://doi.org/10.1094/MPMI-22-2-0115>
- Hong, F., Mo, S. H., Lin, X. Y., Niu, J., Yin, J., & Wei, D. (2020). The PacBio Full-Length Transcriptome of

- the Tea Aphid as a Reference Resource. *Frontiers in Genetics*, 11(November), 1–9.  
<https://doi.org/10.3389/fgene.2020.558394>
- Houterman, P. M., Ma, L., Van Ooijen, G., De Vroomen, M. J., Cornelissen, B. J. C., Takken, F. L. W., & Rep, M. (2009). The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *Plant Journal*, 58(6), 970–978.  
<https://doi.org/10.1111/j.1365-313X.2009.03838.x>
- Hullé, M., Coeur d’Acier, A., Bankhead-Dronnet, S., & Harrington, R. (2010). Aphids in the face of global changes. *Comptes Rendus - Biologies*, 333(6–7), 497–503.  
<https://doi.org/10.1016/j.crv.2010.03.005>
- Ivanoff, S. S. (1944). Resistance of cantaloupes to downy mildew and the melon aphid. *Journal of Heredity*, 35(2), 35–39. <https://doi.org/10.1093/oxfordjournals.jhered.a105345>
- Ji, R., Ye, W., Chen, H., Zeng, J., Li, H., Yu, H., Li, J., & Lou, Y. (2017). A salivary endo- $\beta$ -1,4-glucanase acts as an effector that enables the brown planthopper to feed on rice. *Plant Physiology*, 173(3), 1920–1932. <https://doi.org/10.1104/pp.16.01493>
- Jiang, Y., Zhang, C. X., Chen, R., & He, S. Y. (2019). Challenging battles of plants with phloem-feeding insects and prokaryotic pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 116(47), 23390–23397. <https://doi.org/10.1073/pnas.1915396116>
- Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323–329.  
<https://doi.org/10.1038/nature05286>
- Juan, J., Armenteros, A., Tsigos, K. D., Søndersby, C. K., Petersen, T. N., Winther, O., Brunak, S., Heijne, G. Von, & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(April). <https://doi.org/10.1038/s41587-019-0036-z>
- Kaloshian, I., & Walling, L. (2016a). Plant Immunity: Connecting the Dots Between Microbial and Hemipteran Immune Responses. *Management of Insect Pests to Agriculture: Lessons Learned from Deciphering Their Genome, Transcriptome and Proteome*, 1–290.  
<https://doi.org/10.1007/978-3-319-24049-7>
- Kaloshian, I., & Walling, L. L. (2005). Hemipterans as Plant Pathogens. *Annual Review of Phytopathology*, 43(1), 491–521. <https://doi.org/10.1146/annurev.phyto.43.040204.135944>
- Kaloshian, I., & Walling, L. L. (2016b). Hemipteran and dipteran pests: Effectors and plant host immune regulators. *Journal of Integrative Plant Biology*, 58(4), 350–361.  
<https://doi.org/10.1111/jipb.12438>
- Kamoun, S. (2006). A Catalogue of the Effector Secretome of Plant Pathogenic Oomycetes. *Annual Review of Phytopathology*, 44(1), 41–60.  
<https://doi.org/10.1146/annurev.phyto.44.070505.143436>
- Kamphuis, L. G., Zulak, K., Gao, L. L., Anderson, J., & Singh, K. B. (2013). Plant-aphid interactions with a focus on legumes. *Functional Plant Biology*, 40(12), 1271–1284.  
<https://doi.org/10.1071/FP13090>
- Kanvil, S., Pham, J., Lopez-Cobollo, R., Selby, M., Bennett, M., Beckingham, C., Powell, G., & Turnbull, C. (2017). Cucurbit extrafascicular phloem has strong negative impacts on aphids and is not a preferred feeding site. *Plant Cell and Environment*, 40(11), 2780–2789.  
<https://doi.org/10.1111/pce.13053>
- Kapos, P., Devendrakumar, K. T., & Li, X. (2019). Plant NLRs: From discovery to application. *Plant Science*, 279(March 2018), 3–18. <https://doi.org/10.1016/j.plantsci.2018.03.010>
- Kehr, J., & Buhtz, A. (2008). Long distance transport and movement of RNA through the phloem. *Journal of Experimental Botany*, 59(1), 85–92. <https://doi.org/10.1093/jxb/erm176>
- Kettles, G., & Kaloshian, I. (2016). The potato aphid salivary effector Me47 is a glutathione-S-transferase involved in modifying plant responses to aphid infestation. *Frontiers in Plant Science*, 7(AUG2016), 1–12. <https://doi.org/10.3389/fpls.2016.01142>
- Kim, M. S., Cho, S. M., Kang, E. Y., Im, Y. J., Hwangbo, H., Kim, Y. C., Ryu, C. M., Yang, K. Y., Chung, G. C., & Cho, B. H. (2008). Galactinol is a signaling component of the induced systemic resistance caused by *Pseudomonas chlororaphis* O6 root colonization. *Molecular Plant-Microbe*



- Interactions*, 21(12), 1643–1653. <https://doi.org/10.1094/MPMI-21-12-1643>
- Kishaba, A. N., Bohn, G. W., & Toba, H. H. (1971). *Resistance to Aphis gossypii in Muskmelon. 1*, 935–937.
- Klingler, J., Powell, G., Thompson, G. A., & Isaacs, R. (1997). Phloem specific aphid resistance in Cucumis melo line AR 5: effects on feeding behaviour and performance of *Aphis gossypii*. *Entomologia Experimentalis et Applicata*, 86(1), 79–88. <https://doi.org/10.1046/j.1570-7458.1998.00267.x>
- Kloth, K. J., Busscher-Lange, J., Wiegers, G. L., Kruijer, W., Buijs, G., Meyer, R. C., Albrechtsen, B. R., Bouwmeester, H. J., Dicke, M., & Jongsma, M. A. (2017). SIEVE ELEMENT-LINING CHAPERONE1 restricts aphid feeding on arabidopsis during heat stress. *Plant Cell*, 29(10), 2450–2464. <https://doi.org/10.1105/tpc.16.00424>
- Kloth, K. J., & Kormelink, R. (2020). Defenses against Virus and Vector: A Phloem-Biological Perspective on RTM- and SLI1-Mediated Resistance to Potyviruses and Aphids. *Viruses*, 12(2). <https://doi.org/10.3390/v12020129>
- Knoblauch, M., Knoblauch, J., Mullendore, D. L., Savage, J. A., Babst, B. A., Beecher, S. D., Dodgen, A. C., Jensen, K. H., & Holbrook, N. M. (2016). Testing the Münch hypothesis of long distance phloem transport in plants. *ELife*, 5(JUN2016), 1–16. <https://doi.org/10.7554/eLife.15341>
- Knoblauch, M., & Oparka, K. (2012). The structure of the phloem - Still more questions than answers. *Plant Journal*, 70(1), 147–156. <https://doi.org/10.1111/j.1365-313X.2012.04931.x>
- Knoblauch, M., & Peters, W. S. (2010). Münch, morphology, microfluidics - our structural problem with the phloem. *Plant, Cell and Environment*, 33(9), 1439–1452. <https://doi.org/10.1111/j.1365-3040.2010.02177.x>
- Knoblauch, M., Peters, W. S., Bell, K., Ross-Elliott, T. J., & Oparka, K. J. (2018). Sieve-element differentiation and phloem sap contamination. *Current Opinion in Plant Biology*, 43, 43–49. <https://doi.org/10.1016/j.pbi.2017.12.008>
- Koboldt, D. C., Larson, D. E., & Wilson, R. K. (2014). *Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection*. 1–22. <https://doi.org/10.1002/0471250953.bi1504s44>. Using
- Konno, K. (2011). Plant latex and other exudates as plant defense systems: Roles of various defense chemicals and proteins contained therein. *Phytochemistry*, 72(13), 1510–1530. <https://doi.org/10.1016/j.phytochem.2011.02.016>
- Lacombe, B., & Achard, P. (2016). Long-distance transport of phytohormones through the plant vascular system. *Current Opinion in Plant Biology*, 34, 1–8. <https://doi.org/10.1016/j.pbi.2016.06.007>
- Lagnel, J., Feriche-linares, R., Bertrand, P., Marande, W., Loiseau, A., Sayeh, A., Manno, M., & Boissot, N. (2020). Hybrid de novo genome assembly using Oxford Nanopore Technology , 10X Genomics Linked-Reads sequencing , and Bionano optical map of a non-model species : the case of the melon-cotton aphid *Aphis gossypii*. *Journées Ouvertes de Biologie, Informatique et Mathématique*.
- Lagnel, J., Tsigenopoulos, C. S., & Iliopoulos, I. (2009). NOBLAST and JAMBLAST: New options for BLAST and a Java application manager for BLAST results. *Bioinformatics*, 25(6), 824–826. <https://doi.org/10.1093/bioinformatics/btp067>
- Lambert, I., Paysant-Le Roux, C., Colella, S., & Martin-Magniette, M. L. (2020). DiCoExpress: A tool to process multifactorial RNAseq experiments from quality controls to co-expression analysis through differential analysis based on contrasts inside GLM models. *Plant Methods*, 16(1), 1–10. <https://doi.org/10.1186/s13007-020-00611-7>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Laskowska, I., Leszczynski, B., & Markowski, J. (1999). Activity of glutathione transferase and reductase in tissues of bird cherry-oat aphid during its host-plant alternation. *Experimental and Toxicologic Pathology*, 51(4–5), 357–359. [https://doi.org/10.1016/S0940-2993\(99\)80021-9](https://doi.org/10.1016/S0940-2993(99)80021-9)
- Laugé, R., Joosten, M. H. A. J., Haanstra, J. P. W., Goodwin, P. H., Lindhout, P., & De Wit, P. J. G. M. (1998). Successful search for a resistance gene in tomato targeted against a virulence factor of a

- fungal pathogen. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15), 9014–9018. <https://doi.org/10.1073/pnas.95.15.9014>
- Lecoq, H., Labonne, G., & Pitrat, M. (1980). Specificity of resistance to virus transmission by aphids in *Cucumis melo*. *Ann. Phytopathol*, 12(2), 139–144.
- Lecoq, Hervé, & Desbiez, C. (2012). Viruses of Cucurbit Crops in the Mediterranean Region. An Ever-Changing Picture. In *Advances in Virus Research* (Vol. 84). <https://doi.org/10.1016/B978-0-12-394314-9.00003-8>
- Lei, L., Stevens, D. M., & Coaker, G. (2020). Phosphorylation of the Pseudomonas Effector AvrPtoB by Arabidopsis SnRK2.8 Is Required for Bacterial Virulence. *Molecular Plant*, 13(10), 1513–1522. <https://doi.org/10.1016/j.molp.2020.08.018>
- Leonhardt, H. (1940). Beiträge zur Kenntnis der Lachniden, der wichtigsten Tannenhonigtauerzeuger. *Zeitschrift Für Angewandte Entomologie*, 27(2), 208–272. <https://doi.org/10.1111/j.1439-0418.1940.tb00493.x>
- Lewis, J. D., Knoblauch, M., & Turgeon, R. (2022). The Phloem as an Arena for Plant Pathogens. *Annual Review of Phytopathology*, 60(1), 77–96. <https://doi.org/10.1146/annurev-phyto-020620-100946>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, Q., Xie, Q. G., Smith-Becker, J., Navarre, D. A., & Kaloshian, I. (2006). Mi-1-mediated aphid resistance involves salicylic acid and mitogen-activated protein kinase signaling cascades. *Molecular Plant-Microbe Interactions*, 19(6), 655–664. <https://doi.org/10.1094/MPMI-19-0655>
- Lokeshwari, D., Kumar, N. K. K., Manjunatha, H., & Shivashankar, S. (2016). Biochemical Characterization of Detoxifying Enzymes in Dimethoate-Resistant Strains of Melon Aphid, <i>Aphis gossypii</i> (Hemiptera: Aphididae). *Advances in Entomology*, 04(03), 167–182. <https://doi.org/10.4236/ae.2016.43018>
- Lopez-Cobollo, R. M., Filippis, I., Bennett, M. H., & Turnbull, C. G. N. (2016). Comparative proteomics of cucurbit phloem indicates both unique and shared sets of proteins. *Plant Journal*, 88(4), 633–647. <https://doi.org/10.1111/tpj.13288>
- Luderer, R., Takken, F. L. W., De Wit, P. J. G. M., & Joosten, M. H. A. J. (2002). *Cladosporium fulvum* overcomes Cf-2-mediated resistance by producing truncated AVR2 elicitor proteins. *Molecular Microbiology*, 45(3), 875–884. <https://doi.org/10.1046/j.1365-2958.2002.03060.x>
- Maekawa, T., Kufer, T. A., & Schulze-Lefert, P. (2011). NLR functions in plant and animal immune systems: So far and yet so close. *Nature Immunology*, 12(9), 818–826. <https://doi.org/10.1038/ni.2083>
- Mattson, W. J. (1980). Herbivory in Relation to Plant Nitrogen Content. *Annual Review of Ecology and Systematics*, 11(1), 119–161. <https://doi.org/10.1146/annurev.es.11.110180.001003>
- Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & Depristo, M. A. (2010). The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. 1297–1303. <https://doi.org/10.1101/gr.107524.110.20>
- McLean, D. L., & Kinsey, M. G. (1964). A Technique for Electronically Recording Aphid Feeding and Salivation. *Nature*, 202. <https://doi.org/https://doi.org/10.1038/2021358a0>
- Mesarich, C. H., Bowen, J. K., Hamiaux, C., & Templeton, M. D. (2015). Repeat-containing protein effectors of plant-associated organisms. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2015.00872>
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H., & Michelmore, R. W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell*, 15(4), 809–834. <https://doi.org/10.1105/tpc.009308>
- Miles, P. W. (1965). Studies on the Salivary Physiology of Plant-bugs : the Salivary Secretions of Aphids. *Journal of Insect Physiology*, 11, 1261–1268.
- Miles, P. W., & Peng, Z. (1989). Studies on the salivary physiology of plant bugs: Detoxification of phytochemicals by the salivary peroxidase of aphids. *Journal of Insect Physiology*, 35(11), 865–

872. [https://doi.org/10.1016/0022-1910\(89\)90102-9](https://doi.org/10.1016/0022-1910(89)90102-9)
- Mistral, P., Vanlerberghe-Masutti, F., Elbelt, S., & Boissot, N. (2021). *Aphis gossypii/Aphis frangulae* collected worldwide: Microsatellite markers data and genetic cluster assignment. *Data in Brief*, *36*, 106967. <https://doi.org/10.1016/j.dib.2021.106967>
- Mithöfer, A., & Boland, W. (2008). Recognition of herbivory-associated molecular patterns. *Plant Physiology*, *146*(3), 825–831. <https://doi.org/10.1104/pp.107.113118>
- Mittler, T. E. (1957). Studies on the feeding and nutrition of *Tuberolachnus Salignus* (Gmelin) (Homoptera, Aphididae):: 1. the Uptake of Phloem Sap. *J. Exp. Biol.*, *34*(3), 334–341. <http://jeb.biologists.org/content/34/3/334.short>
- Mondal, H. A. (2020). Aphid saliva: a powerful recipe for modulating host resistance towards aphid clonal propagation. *Arthropod-Plant Interactions*, *0123456789*. <https://doi.org/10.1007/s11829-020-09769-2>
- Mugford, S. T., Barclay, E., Drurey, C., Findlay, K. C., & Hogenhout, S. A. (2016). An immunosuppressive aphid saliva protein is delivered into the cytosol of plant mesophyll cells during feeding. *Molecular Plant-Microbe Interactions*, *29*(11), 854–861. <https://doi.org/10.1094/MPMI-08-16-0168-R>
- Mullendore, D. L., Windt, C. W., van As, H., & Knoblauch, M. (2010). Sieve tube geometry in relation to phloem flow. *Plant Cell*, *22*(3), 579–593. <https://doi.org/10.1105/tpc.109.070094>
- Mutti, N. S., Louis, J., Pappan, L. K., Pappan, K., Begum, K., Chen, M. S., Park, Y., Dittmer, N., Marshall, J., Reese, J. C., & Reeck, G. R. (2008). A protein from the salivary glands of the pea aphid, *Acyrtosiphon pisum*, is essential in feeding on a host plant. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(29), 9965–9969. <https://doi.org/10.1073/pnas.0708958105>
- Mutti, N. S., Park, Y., Reese, J. C., & Reeck, G. R. (2006). RNAi Knockdown of a Salivary Transcript Leading to Lethality in the Pea Aphid, *Acyrtosiphon pisum*. *Journal of Insect Science*, *6*(38), 1–7. <https://doi.org/10.1673/031.006.3801>
- Nadwodnik, J., & Lohaus, G. (2008). Subcellular concentrations of sugar alcohols and sugars in relation to phloem translocation in *Plantago major*, *Plantago maritima*, *Prunus persica*, and *Apium graveolens*. *Planta*, *227*(5), 1079–1089. <https://doi.org/10.1007/s00425-007-0682-0>
- Naessens, E., Dubreuil, G., Giordanengo, P., Baron, O. L., Minet-Kebdani, N., Keller, H., & Coustau, C. (2015). A Secreted MIF Cytokine Enables Aphid Feeding and Represses Plant Immune Responses. *Current Biology*, *25*(14), 1898–1903. <https://doi.org/10.1016/j.cub.2015.05.047>
- Nagare, M., Ayachit, M., Agnihotri, A., Schwab, W., & Joshi, R. (2021). Glycosyltransferases: the multifaceted enzymatic regulator in insects. *Insect Molecular Biology*, *30*(2), 123–137. <https://doi.org/10.1111/imb.12686>
- Nalam, V., Louis, J., & Shah, J. (2019). Plant defense against aphids, the pest extraordinaire. *Plant Science*, *279*(April 2018), 96–107. <https://doi.org/10.1016/j.plantsci.2018.04.027>
- Nicholson, S. J., Nickerson, M. L., Dean, M., Song, Y., Hoyt, P. R., Rhee, H., Kim, C., & Puterka, G. J. (2015). The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC Genomics*, *16*(1), 1–16. <https://doi.org/10.1186/s12864-015-1525-1>
- Nicolis, V., & Venter, E. (2018). Silencing of a unique integrated domain nucleotide-binding leucine-rich repeat gene in wheat abolishes *Diuraphis noxia* resistance. *Molecular Plant-Microbe Interactions*, *31*(9), 940–950. <https://doi.org/10.1094/MPMI-11-17-0262-R>
- Nielsen, H. (2017). *Predicting Secretory Proteins with SignalP*. *1611*, 59–73. <https://doi.org/10.1007/978-1-4939-7015-5>
- Nilsen, T. W., & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, *463*(7280), 457–463. <https://doi.org/10.1038/nature08909>
- Noda, H., Kawai, S., Koizumi, Y., Matsui, K., Zhang, Q., Furukawa, S., Shimomura, M., & Mita, K. (2008). Annotated ESTs from various tissues of the brown planthopper *Nilaparvata lugens*: A genomic resource for studying agricultural pests. *BMC Genomics*, *9*, 1–18. <https://doi.org/10.1186/1471-2164-9-117>
- Nombela, G., Williamson, V. M., & Muñiz, M. (2003). The root-knot nematode resistance gene Mi-1.2

- of tomato is responsible for resistance against the whitefly *Bemisia tabaci*. *Molecular Plant-Microbe Interactions*, 16(7), 645–649. <https://doi.org/10.1094/MPMI.2003.16.7.645>
- Nürnberg, T., Brunner, F., Kemmerling, B., & Piater, L. (2004). Innate immunity in plants and animals: Striking similarities and obvious differences. *Immunological Reviews*, 198, 249–266. <https://doi.org/10.1111/j.0105-2896.2004.0119.x>
- Omid, A., Keilin, T., Glass, A., Leshkowitz, D., & Wolf, S. (2007). Characterization of phloem-sap transcription profile in melon plants. *Journal of Experimental Botany*, 58(13), 3645–3656. <https://doi.org/10.1093/jxb/erm214>
- Oparka, K. J., & Turgeon, R. (1999). Sieve elements and companion cells - Traffic control centers of the phloem. *Plant Cell*, 11(4), 739–750. <https://doi.org/10.1105/tpc.11.4.739>
- Peng, H. C., & Walker, G. P. (2018). Sieve element occlusion provides resistance against *Aphis gossypii* in TGR-1551 melons. *Insect Science*, 27(1), 33–48. <https://doi.org/10.1111/1744-7917.12610>
- Pirone, T. P., & Perry, K. L. (2002). Aphids : Non-persistent Transmission. *Advances in Botanical Research*.
- Pitino, M., & Hogenhout, S. A. (2013). Aphid protein effectors promote aphid colonization in a plant species-specific manner. *Molecular Plant-Microbe Interactions*, 26(1), 130–139. <https://doi.org/10.1094/MPMI-07-12-0172-FI>
- Pitrat, M., & Lecoq, H. (1980). Inheritance of Resistance to Cucumber Mosaic Virus Transmission by *Aphis gossypii* in Cucumis melo. *Phytopathology*, 70(10), 958. <https://doi.org/10.1094/phyto-70-958>
- Ponsen, M. B. (1972). The Site of Potato Leafroll Virus Multiplication in Its Vector, *Myzus persicae*.
- Powell, G., Tosh, C. R., & Hardie, J. (2006). HOST PLANT SELECTION BY APHIDS: Behavioral, Evolutionary, and Applied Perspectives. *Annual Review of Entomology*, 51(1), 309–330. <https://doi.org/10.1146/annurev.ento.51.110104.151107>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Qutob, D., Tedman-Jones, J., Dong, S., Kuflu, K., Pham, H., Wang, Y., Dou, D., Kale, S. D., Arredondo, F. D., Tyler, B. M., & Gijzen, M. (2009). Copy number variation and transcriptional polymorphisms of *Phytophthora sojae* RXLR effector genes Avr1a and Avr3a. *PLoS ONE*, 4(4). <https://doi.org/10.1371/journal.pone.0005066>
- Rafiqi, M., Bernoux, M., Ellis, J. G., & Dodds, P. N. (2009). In the trenches of plant pathogen recognition: Role of NB-LRR proteins. *Seminars in Cell and Developmental Biology*, 20(9), 1017–1024. <https://doi.org/10.1016/j.semcd.2009.04.010>
- Rajasree, R. S., Sibi, P. I., Francis, F., & William, H. (2016). Phytochemicals of cucurbitaceae family – A review. *International Journal of Pharmacognosy and Phytochemical Research*, 8(1), 113–123.
- Ramsey, J. S., Rider, D. S., Walsh, T. K., De Vos, M., Gordon, K. H. J., Ponnala, L., MacMil, S. L., Roe, B. A., & Jander, G. (2010). Comparative analysis of detoxification enzymes in *Acyrtosiphon pisum* and *Myzus persicae*. *Insect Molecular Biology*, 19(SUPPL. 2), 155–164. <https://doi.org/10.1111/j.1365-2583.2009.00973.x>
- Ramsey, John S., Wilson, A. C. C., de Vos, M., Sun, Q., Tamborindeguy, C., Winfield, A., Malloch, G., Smith, D. M., Fenton, B., Gray, S. M., & Jander, G. (2007). Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. *BMC Genomics*, 8(May 2014). <https://doi.org/10.1186/1471-2164-8-423>
- Rao, W., Zheng, X., Liu, B., Guo, Q., Guo, J., Wu, Y., Shangguan, X., Wang, H., Wu, D., Wang, Z., Hu, L., Xu, C., Jiang, W., Huang, J., Shi, S., & He, G. (2019). Secretome analysis and in planta expression of salivary proteins identify candidate effectors from the brown planthopper *Nilaparvata lugens*. *Molecular Plant-Microbe Interactions*, 32(2), 227–239. <https://doi.org/10.1094/MPMI-05-18-0122-R>
- Rice, P., Longden, L., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, 16(6), 276–277. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2)
- Richards, S., Gibbs, R. A., Gerardo, N. M., Moran, N., Nakabachi, A., Stern, D., Tagu, D., Wilson, A. C., Muzny, D., Kovar, C., Cree, A., Chacko, J., Chandrabose, M. N., Dao, M. D., Dinh, H. H., Gabisi,

- R. A., Hines, S., Hume, J., Jhangian, S. N., ... Hunter, W. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biology*, 8(2). <https://doi.org/10.1371/journal.pbio.1000313>
- Richardson, P. T., Baker, D. A., & Ho, L. C. (1982). The chemical composition of cucurbit vascular exudates. *Journal of Experimental Botany*, 33(6), 1239–1247. <https://doi.org/10.1093/jxb/33.6.1239>
- Rodriguez, P. A., Escudero-Martinez, C., & Bos, J. I. B. (2017). An aphid effector targets trafficking protein VPS52 in a host-specific manner to promote virulence. *Plant Physiology*, 173(3), 1892–1903. <https://doi.org/10.1104/pp.16.01458>
- Rodriguez, P. A., Stam, R., Warbroek, T., & Bos, J. I. B. (2014). Mp10 and Mp42 from the aphid species *Myzus persicae* trigger plant defenses in nicotiana benthamiana through different activities. *Molecular Plant-Microbe Interactions*, 27(1), 30–39. <https://doi.org/10.1094/MPMI-05-13-0156-R>
- Rossi, M., Goggin, F. L., Milligan, S. B., Kaloshian, I., Ullman, D. E., & Williamson, V. M. (1998). The nematode resistance gene Mi of tomato confers resistance against the potato aphid. *Proceedings of the National Academy of Sciences of the United States of America*, 95(17), 9750–9754. <https://doi.org/10.1073/pnas.95.17.9750>
- Rouxel, T., & Balesdent, M.-H. (2010). Avirulence Genes. *Encyclopedia of Life Sciences*. <https://doi.org/10.1002/9780470015902.a0021267>
- Sabater-Muñoz, B., Legeai, F., Rispe, C., Bonhomme, J., Dearden, P., Dossat, C., Duclert, A., Gauthier, J. P., Ducray, D. G., Hunter, W., Dang, P., Kambhampati, S., Martinez-Torres, D., Cortes, T., Moya, A., Nakabachi, A., Philippe, C., Prunier-Leterme, N., Rahbé, Y., ... Tagu, D. (2006). Large-scale gene discovery in the pea aphid *Acyrtosiphon pisum* (Hemiptera). *Genome Biology*, 7(3). <https://doi.org/10.1186/gb-2006-7-3-r21>
- Salinier, J., Lefebvre, V., Besombes, D., Burck, H., Causse, M., Daunay, M. C., Dogimont, C., Goussopoulos, J., Gros, C., Maisonneuve, B., McLeod, L., Tobal, F., & Stevens, R. (2022). The INRAE Centre for Vegetable Germplasm: Geographically and Phenotypically Diverse Collections and Their Use in Genetics and Plant Breeding. *Plants*, 11(3). <https://doi.org/10.3390/plants11030347>
- Sandström, J., & Pettersson, J. (1994). Amino acid composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrtosiphon pisum*) performance. *Journal of Insect Physiology*, 40(11), 947–955. [https://doi.org/10.1016/0022-1910\(94\)90133-3](https://doi.org/10.1016/0022-1910(94)90133-3)
- Sandström, J., Telang, A., & Moran, N. A. (2000). Nutritional enhancement of host plants by aphids - A comparison of three aphid species on grasses. *Journal of Insect Physiology*, 46(1), 33–40. [https://doi.org/10.1016/S0022-1910\(99\)00098-0](https://doi.org/10.1016/S0022-1910(99)00098-0)
- Schnieder, N., Känel, A., Zimmermann, M., Kriebs, K., Witte, A., Wrobel, L. S., Twyman, R. M., Prüfer, D., Furch, A. C. U., & Noll, G. A. (2022). So similar yet so different: The distinct contributions of extrafascicular and fascicular phloem to transport and exudation in cucumber plants. *Journal of Plant Physiology*, 271. <https://doi.org/10.1016/j.jplph.2022.153643>
- Schoeny, A., Desbiez, C., Millot, P., Wipf-Scheibel, C., Nozeran, K., Gognalons, P., Lecoq, H., & Boissot, N. (2017). Impact of Vat resistance in melon on viral epidemics and genetic structure of virus populations. *Virus Research*, 241, 105–115. <https://doi.org/10.1016/j.virusres.2017.05.024>
- Schulz, A. (1998). *Phloem. Structure Related to Function*. [https://doi.org/10.1007/978-3-642-80446-5\\_16](https://doi.org/10.1007/978-3-642-80446-5_16)
- Sengupta, S., Mukherjee, S., Basak, P., & Majumder, A. L. (2015). Significance of galactinol and raffinose family oligosaccharide synthesis in plants. *Frontiers in Plant Science*, 6(AUG), 1–11. <https://doi.org/10.3389/fpls.2015.00656>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Sparkes, A., De Baetselier, P., Roelants, K., De Trez, C., Magez, S., Van Ginderachter, J. A., Raes, G., Bucala, R., & Stijlemans, B. (2017). The non-mammalian MIF superfamily. *Immunobiology*, 222(3), 473–482. <https://doi.org/10.1016/j.imbio.2016.10.006>

- Stahl, E., Hilfiker, O., & Reymond, P. (2018). Plant–arthropod interactions: who is the winner? *Plant Journal*, *93*(4), 703–728. <https://doi.org/10.1111/tpj.13773>
- Ste-Croix, D. T., St-Marseille, A. F. G., Lord, E., Bélanger, R. R., Brodeur, J., & Mimee, B. (2021). Genomic profiling of virulence in the soybean cyst nematode using single-nematode sequencing. *Phytopathology*, *111*(1), 137–148. <https://doi.org/10.1094/PHYTO-08-20-0348-FI>
- Sun, Y., Guo, H., & Ge, F. (2020). *Medicago truncatula* –pea aphid interaction in the context of global climate change. *The Model Legume Medicago Truncatula*, 369–376. <https://doi.org/10.1002/9781119409144.ch45>
- Tamura, Y., Hattori, M., Yoshioka, H., Yoshioka, M., Takahashi, A., Wu, J., Sentoku, N., & Yasui, H. (2014). Map-based cloning and characterization of a brown planthopper resistance gene BPH26 from *Oryza sativa* L. ssp. indica cultivar ADR52. *Scientific Reports*, *4*, 1–8. <https://doi.org/10.1038/srep05872>
- Telang, A., Sandström, J., Dyreson, E., & Moran, N. A. (2002). Feeding damage by *Diuraphis noxia* results in a nutritionally enhanced phloem diet *Aparna*. *Entomologia Experimentalis et Applicata*, 403–412. <https://doi.org/10.1023/a:1003617523820>
- Thomas, S., Boissot, N., & Vanlerberghe-Masutti, F. (2012). What do spring migrants reveal about sex and host selection in the melon aphid? *BMC Evolutionary Biology*, *12*(1). <https://doi.org/10.1186/1471-2148-12-47>
- Thomas, S., Dogimont, C., & Boissot, N. (2012). Association between *Aphis gossypii* genotype and phenotype on melon accessions. *Arthropod-Plant Interactions*, *6*(1), 93–101. <https://doi.org/10.1007/s11829-011-9155-2>
- Thomas, S., Vanlerberghe-Masutti, F., Mistral, P., Loiseau, A., & Boissot, N. (2016). Insight into the durability of plant resistance to aphids from a demo-genetic study of *Aphis gossypii* in melon crops. *Evolutionary Applications*, *9*(6), 756–768. <https://doi.org/10.1111/eva.12382>
- Thompson, G. A., & van Bel, A. J. E. (2012). Phloem: Molecular Cell Biology, Systemic Communication, Biotic Interactions. In *Phloem: Molecular Cell Biology, Systemic Communication, Biotic Interactions*. <https://doi.org/10.1002/9781118382806>
- Thompson, S. N. (2003). Trehalose - The Insect “Blood” Sugar. In *Advances in Insect Physiology* (Vol. 31, Issue 03). [https://doi.org/10.1016/S0065-2806\(03\)31004-5](https://doi.org/10.1016/S0065-2806(03)31004-5)
- Thorpe, P., Cock, P. J. A., & Bos, J. (2016). Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets. *BMC Genomics*, *17*(1), 1–18. <https://doi.org/10.1186/s12864-016-2496-6>
- Tian, T., Ji, R., Fu, J., Li, J., Wang, L., Zhang, H., Yang, S., Ye, W., Fang, J., & Zhu-Salzman, K. (2021). A salivary calcium-binding protein from *Laodelphax striatellus* acts as an effector that suppresses defense in rice. *Pest Management Science*, *77*(5), 2272–2281. <https://doi.org/10.1002/ps.6252>
- Tjallingii, W. Fred. (2006). Salivary secretions by aphids interacting with proteins of phloem wound responses. *Journal of Experimental Botany*, *57*(4), 739–745. <https://doi.org/10.1093/jxb/erj088>
- Tjallingii, W F. (1978). *Electronic recording of penetration behaviour by aphids*. 24.
- Torkey, H. M., Abou-Yousef, H. M., Azeiz, A., & Farid, H. E. A. (2009). Insecticidal effect of Cucurbitacin E Glycoside isolated from *Citrullus colocynthis* against *Aphis craccivora*. *Australian Journal of Basic and Applied Sciences*, *3*(4), 4060–4066.
- Toyota, M., Spencer, D., Sawai-toyota, S., Jiaqi, W., & Zhang, T. (2018). *Glutamate triggers long-distance, calcium-based plant defense signaling*. 6(September), 1112–1115.
- Turgeon, R. (2016). Phloem Biology of the Cucurbitaceae. *Plant Genetics and Genomics: Crops and Models*, October, 291–311. <https://doi.org/10.1007/7397>
- Turgeon, Robert, & Oparka, K. (2010). The secret phloem of pumpkins. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(30), 13201–13202. <https://doi.org/10.1073/pnas.1008134107>
- Turgeon, Robert, & Wolf, S. (2009). Phloem Transport: Cellular Pathways and Molecular Trafficking. *Annual Review of Plant Biology*, *60*(1), 207–221. <https://doi.org/10.1146/annurev.arplant.043008.092045>
- Turnbull, C. G. N., & Lopez-Cobollo, R. M. (2013). Heavy traffic in the fast lane: Long-distance

- signalling by macromolecules. *New Phytologist*, 198(1), 33–51.  
<https://doi.org/10.1111/nph.12167>
- Twayana, M., Girija, A. M., Mohan, V., & Shah, J. (2022). Phloem: At the center of action in plant defense against aphids. *Journal of Plant Physiology*, 273(October 2021), 153695.  
<https://doi.org/10.1016/j.jplph.2022.153695>
- Tyler, B. M., & Gijzen, M. (2014). The *Phytophthora sojae* Genome Sequence: Foundation for a Revolution. *Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens*, 1–239.  
<https://doi.org/10.1007/978-3-662-44056-8>
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, 40(15), 1–12.  
<https://doi.org/10.1093/nar/gks596>
- Uzest, M., Gargani, D., Dombrovsky, A., Cazevieille, C., Cot, D., & Blanc, S. (2010). The “acrostyle”: A newly described anatomical structure in aphid stylets. *Arthropod Structure and Development*, 39(4), 221–229. <https://doi.org/10.1016/j.asd.2010.02.005>
- Valluru, R., & Van den Ende, W. (2011). Myo-inositol and beyond - Emerging networks under stress. *Plant Science*, 181(4), 387–400. <https://doi.org/10.1016/j.plantsci.2011.07.009>
- Van Baaren, J., Le Lann, C., & Van Alphen, J. J. M. (2010). Aphid Biodiversity under Environmental Change. *Aphid Biodiversity under Environmental Change*, 55–68. <https://doi.org/10.1007/978-90-481-8601-3>
- Van Bel, A. J.E. (2003). The phloem, a miracle of ingenuity. *Plant, Cell and Environment*, 26(1), 125–149. <https://doi.org/10.1046/j.1365-3040.2003.00963.x>
- Van Bel, Aart J.E. (1999). Evolution, polymorphology and multifunctionality of the phloem system. *Perspectives in Plant Ecology, Evolution and Systematics*, 2(2), 163–184.  
<https://doi.org/10.1078/1433-8319-00069>
- Van Bel, Aart J.E., & Gaupels, F. (2004). Pathogen-induced resistance and alarm signals in the phloem. *Molecular Plant Pathology*, 5(5), 495–504. <https://doi.org/10.1111/J.1364-3703.2004.00243.X>
- van Bel, Aart J.E., & Will, T. (2016). Functional evaluation of proteins in watery and gel saliva of aphids. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2016.01840>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. In *Current Protocols in Bioinformatics* (Issue SUPL.43).  
<https://doi.org/10.1002/0471250953.bi1110s43>
- Van Der Hoorn, R. A. L., & Kamoun, S. (2008). From guard to decoy: A new model for perception of plant pathogen effectors. *Plant Cell*, 20(8), 2009–2017. <https://doi.org/10.1105/tpc.108.060194>
- Van Emden, H. F., & Harrington, R. (2007). *Aphids as Crop Pests* (H. F. Van Emden & R. Harrington (eds.)). <https://doi.org/10.1079/9781780647098.0173>
- 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(5), 1–12.  
<https://doi.org/10.1371/journal.pone.0174398>
- Villada, E. S., González, E. G., López-Sesé, A. I., Castiel, A. F., & Gómez-Guillamón, M. L. (2009). Hypersensitive response to *Aphis gossypii* Glover in melon genotypes carrying the *Vat* gene. *Journal of Experimental Botany*, 60(11), 3269–3277. <https://doi.org/10.1093/jxb/erp163>
- Vos, P., Simons, G., Jesse, T., Wijbrandi, J., Heinen, L., Hogers, R., Frijters, A., Groenendijk, J., Diergaarde, P., Reijans, M., Fierens-Onstenk, J., De Both, M., Peleman, J., Liharska, T., Hontelez, J., & Zabeau, M. (1998). The tomato Mi-1 gene confers resistance to both root-knot nematodes and potato aphids. *Nature Biotechnology*, 16(13), 1365–1369. <https://doi.org/10.1038/4350>
- Walker, G. P. (2021). Sieve element occlusion: Interactions with phloem sap-feeding insects. A review. *Journal of Plant Physiology*, 269(December 2021), 153582.  
<https://doi.org/10.1016/j.jplph.2021.153582>
- Walling, L. L. (2008). Avoiding effective defenses: Strategies employed by phloem-feeding insects. *Plant Physiology*, 146(3), 859–866. <https://doi.org/10.1104/pp.107.113142>

- Walz, C., Giavalisco, P., Schad, M., Juenger, M., Klose, J., & Kehr, J. (2004). Proteomics of curcubit phloem exudate reveals a network of defence proteins. *Phytochemistry*, *65*(12), 1795–1804. <https://doi.org/10.1016/j.phytochem.2004.04.006>
- Wang, W., Dai, H., Zhang, Y., Chandrasekar, R., Luo, L., Hiromasa, Y., Sheng, C., Peng, G., Chen, S., Tomich, J. M., Reese, J., Edwards, O., Kang, L., Reeck, G., & Cui, F. (2015). Armet is an effector protein mediating aphid-plant interactions. *FASEB Journal*, *29*(5), 2032–2045. <https://doi.org/10.1096/fj.14-266023>
- Wang, W., Luo, L., Lu, H., Chen, S., Kang, L., & Cui, F. (2015). Angiotensin-converting enzymes modulate aphid-plant interactions. *Scientific Reports*. <https://doi.org/10.1038/srep08885>
- Waterhouse, R. M., Zdobnov, E. M., & Kriventseva, E. V. (2011). Correlating traits of gene retention, sequence divergence, duplicability and essentiality in vertebrates, arthropods, and fungi. *Genome Biology and Evolution*, *3*(1), 75–86. <https://doi.org/10.1093/gbe/evq083>
- Webster, C. G., Pichon, E., Munster, M. Van, Monsion, B., Deshoux, M., & Gargani, D. (2018). crossm Identification of Plant Virus Receptor Candidates in the Stylets. *Journal of Virology*, *92*(14), 1–15.
- Westerink, N., Brandwagt, B. F., De Wit, P. J. G. M., & Joosten, M. H. A. J. (2004). Cladosporium fulvum circumvents the second functional resistance gene homologue at the Cf-4 locus (Hcr9-4E) by secretion of a stable avr4E isoform. *Molecular Microbiology*, *54*(2), 533–545. <https://doi.org/10.1111/j.1365-2958.2004.04288.x>
- Will, T., Furch, A. C. U., & Zimmermann, M. R. (2013). How phloem-feeding insects face the challenge of phloem-located defenses. *Frontiers in Plant Science*, *4*(AUG), 1–12. <https://doi.org/10.3389/fpls.2013.00336>
- Will, T., Kornemann, S. R., Furch, A. C. U., Fred Tjallingil, W., & Van Bel, A. J. E. (2009). Aphid watery saliva counteracts sieve-tube occlusion: A universal phenomenon? *Journal of Experimental Biology*, *212*(20), 3305–3312. <https://doi.org/10.1242/jeb.028514>
- Will, T., Steckbauer, K., Hardt, M., & van Bel, A. J. E. (2012). Aphid Gel Saliva: Sheath Structure, Protein Composition and Secretory Dependence on Stylet-Tip Milieu. *PLoS ONE*, *7*(10). <https://doi.org/10.1371/journal.pone.0046903>
- Will, T., Tjallingii, W. F., Thönnessen, A., & Van Bel, A. J. E. (2007). Molecular sabotage of plant defense by aphid saliva. *PNAS*, *104*(25), 10536–10541. [https://doi.org/10.1016/0047-2670\(76\)85028-9](https://doi.org/10.1016/0047-2670(76)85028-9)
- Will, T., & Vilcinskas, A. (2015). The structural sheath protein of aphids is required for phloem feeding. *Insect Biochemistry and Molecular Biology*, *57*, 34–40. <https://doi.org/10.1016/j.ibmb.2014.12.005>
- Windt, C. W., Vergeldt, F. J., De Jager, P. A., & Van As, H. (2006). MRI of long-distance water transport: A comparison of the phloem and xylem flow characteristics and dynamics in poplar, castor bean, tomato and tobacco. *Plant, Cell and Environment*, *29*(9), 1715–1729. <https://doi.org/10.1111/j.1365-3040.2006.01544.x>
- Yates, A. D., & Michel, A. (2018). Mechanisms of aphid adaptation to host plant resistance. *Current Opinion in Insect Science*, *26*, 41–49. <https://doi.org/10.1016/j.cois.2018.01.003>
- Yeaman, I., Nguyen, H. P., & Martin, G. B. (2010). Phosphorylation of the Pseudomonas syringae effector AvrPto is required for FLS2/BAK1-independent virulence activity and recognition by tobacco. *Plant Journal*, *61*(1), 16–24. <https://doi.org/10.1111/j.1365-313X.2009.04028.x>
- Yoo, B.-C., Kragler, F., Varkonyi-Gasic, E., Haywood, V., Archer-Evans, S., Lee, Y. M., Lough, T. J., & Lucas, W. J. (2004). A systemic small RNA signaling. *The Plant Cell*, *16*(August), 1979–2000. <https://doi.org/10.1105/tpc.104.023614.specialized>
- Yousaf, H. K., Shan, T., Chen, X., Ma, K., Shi, X., Desneux, N., Biondi, A., & Gao, X. (2018). Impact of the secondary plant metabolite Cucurbitacin B on the demographical traits of the melon aphid, *Aphis gossypii*. *Scientific Reports*, *8*(1), 1–10. <https://doi.org/10.1038/s41598-018-34821-w>
- Zhang, B., Tolstikov, V., Turnbull, C., Hicks, L. M., & Fiehn, O. (2010). Divergent metabolome and proteome suggest functional independence of dual phloem transport systems in cucurbits. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(30),



- 13532–13537. <https://doi.org/10.1073/pnas.0910558107>
- Zhang, C., Yu, X., Ayre, B. G., & Turgeon, R. (2012). The origin and composition of cucurbit “Phloem” exudate. *Plant Physiology*, *158*(4), 1873–1882. <https://doi.org/10.1104/pp.112.194431>
- Zhang, W., Wang, S., Yang, J., Kang, C., Huang, L., & Guo, L. (2022). Glycosylation of plant secondary metabolites: Regulating from chaos to harmony. *Environmental and Experimental Botany*, *194*(October 2021), 104703. <https://doi.org/10.1016/j.envexpbot.2021.104703>
- Zhang, Y., Fan, J., Sun, J., Francis, F., & Chen, J. (2017). Transcriptome analysis of the salivary glands of the grain aphid, *Sitobion avenae*. *Scientific Reports*, *7*(1), 1–14. <https://doi.org/10.1038/s41598-017-16092-z>
- Zhang, Y., Fan, J., Sun, J. rui, & Chen, J. lian. (2015). Cloning and RNA interference analysis of the salivary protein C002 gene in *Schizaphis graminum*. *Journal of Integrative Agriculture*, *14*(4), 698–705. [https://doi.org/10.1016/S2095-3119\(14\)60822-4](https://doi.org/10.1016/S2095-3119(14)60822-4)
- Zhang, Y., Liu, X., Francis, F., Xie, H., Fan, J., Wang, Q., Liu, H., Sun, Y., & Chen, J. (2022). The salivary effector protein Sg2204 in the greenbug *Schizaphis graminum* suppresses wheat defence and is essential for enabling aphid feeding on host plants. *Plant Biotechnology Journal*, 2187–2201. <https://doi.org/10.1111/pbi.13900>
- Zhang, Y., Liu, X., Fu, Y., Crespo-Herrera, L., Liu, H., Wang, Q., Zhang, Y., & Chen, J. (2022). Salivary Effector Sm9723 of Grain Aphid *Sitobion miscanthi* Suppresses Plant Defense and Is Essential for Aphid Survival on Wheat. *International Journal of Molecular Sciences*, *23*(13). <https://doi.org/10.3390/ijms23136909>
- Zhao, C., Ma, C., Luo, J., Niu, L., Hua, H., Zhang, S., & Cui, J. (2021). Potential of cucurbitacin b and epigallocatechin gallate as biopesticides against *Aphis gossypii*. *Insects*, *12*(1), 1–16. <https://doi.org/10.3390/insects12010032>
- Zhu, C., Liu, T., Chang, Y. N., & Duan, C. G. (2019). Small RNA functions as a trafficking effector in plant immunity. *International Journal of Molecular Sciences*, *20*(11). <https://doi.org/10.3390/ijms20112816>
- Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D. G., Boller, T., & Felix, G. (2006). Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated Transformation. *Cell*, *125*(4), 749–760. <https://doi.org/10.1016/j.cell.2006.03.037>
- Züst, T., & Agrawal, A. A. (2016). Mechanisms and evolution of plant resistance to aphids. *Nature Plants*, *2*(January), 1–9. <https://doi.org/10.1038/nplants.2015.206>