



**HAL**  
open science

# Diversité, caractéristiques évolutives et rôles des effecteurs salivaires du puceron du pois dans l'interaction avec ses plantes hôtes

Hélène Boulain

► **To cite this version:**

Hélène Boulain. Diversité, caractéristiques évolutives et rôles des effecteurs salivaires du puceron du pois dans l'interaction avec ses plantes hôtes. Biologie végétale. Agrocampus Ouest, 2017. Français. NNT : 2017NSARA082 . tel-01814755

**HAL Id: tel-01814755**

**<https://theses.hal.science/tel-01814755>**

Submitted on 13 Jun 2018

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

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

UNIVERSITE

BRETAGNE

LOIRE

AGRO  
CAMPUS

OUEST

**Hélène BOULAIN** • 15 décembre 2017

Thèse AGROCAMPUS OUEST  
sous le label de l'Université Bretagne Loire  
pour obtenir le grade de  
**DOCTEUR D'AGROCAMPUS OUEST**  
Spécialité Écologie et Évolution

ÉCOLE DOCTORALE • Écologie, Géosciences, Agronomie et  
Alimentation (EGAAL)

LABORATOIRE D'ACCUEIL • UMR 1349 Institut de génétique,  
environnement et protection des plantes (IGEPP)

**Diversité, caractéristiques  
évolutives et rôles  
des effecteurs salivaires  
du puceron du pois  
dans l'interaction  
avec ses plantes hôtes**

**Yannick OUTREMAN**

Professeur, AGROCAMPUS OUEST, UMR INRA-UR1-AO IGEPP /  
*président*

**Virginie COURTIER-ORGOGOZO**

Directrice de recherche, CNRS / *rapporteuse*

**Philippe REYMOND**

Professeur, Université de Lausanne / *rapporteur*

**Christine COUSTAU**

Directrice de recherche, CNRS / *examinatrice*

**Boris SZUREK**

Chargé de recherche, IRD / *examineur*

**Akiko SUGIO**

Chargée de recherche, UMR INRA-UR1-AO IGEPP /  
*directrice de thèse*

**Jean-Christophe SIMON**

Co-directeur de recherche, UMR INRA-UR1-AO IGEPP /  
*co-directeur de thèse*



## REMERCIEMENTS

Je tiens tout d'abord à remercier mes deux directeurs de thèse Akiko Sugio et Jean-Christophe Simon. Merci à vous d'avoir pris le temps de m'aider dans la rédaction de ce mémoire, aux dépens de vos vacances et heures de sommeil... ! Durant ces trois années de thèse vous m'avez offert la possibilité de rencontrer des collaborateurs étrangers passionnant et de présenter mon travail à de nombreux colloques internationaux, merci car cela m'a vraiment enrichie ! Merci aussi de m'avoir poussé à passer 5 mois de ma thèse en Angleterre, c'était tout simplement génial !

Ensuite je remercie Endrick Guy, maître incontesté de la bactériologie et de la biomol, tu m'as tout appris. Merci pour ces fous-rires partagés dans le bureau durant ces deux années passées ensemble.

Merci à "the amazing Fabrice Legeai", la bioinfo n'a pas de secrets pour toi, heureusement que tu as été présent et toujours réactif pendant ces 3 ans pour m'aider dès que j'en avais besoin !

Merci à Patoche d'être toujours dispo pour nous accompagner dans toutes ces démarches administratives compliquées ! Que ferait-on sans toi !

Merci à tous les membres du labo de Saskia Hogenhout, Saskia, Sam, Beccy, Sylvain, Yazhoo, Adi, Zigmunds, Thomas, et Gabriele de m'avoir si bien accueilli à Norwich. Mention spéciale à Sam, mon deuxième maître en biomol, merci pour ta bonne humeur, ta patience et ta fameuse rengaine du vendredi soir : "Beer o'clock ?"

Merci à mes formidables colloques' du 14 Fairfax Road, Nuno, Marc, et Andrew pour ces 4 mois d'échanges et de rigolade et à Billy, Tom et Sylvain de m'avoir accueilli le dernier mois. Sans oublier tous les amis rencontrés durant ces 5 mois, Jemima, Agathe, Antoine, Torsten et Jofus. Vous avez tous participé à rendre cette expérience anglaise incroyable !!

Merci à Jim et Nadine de m'avoir fait découvrir l'Irlande, les pubs de Dublin et la protéomique !

Merci aux membres de mon comité de thèse pour leur bienveillance et les discussions que nous avons partagées.

Un grand merci à tous les techniciens de l'équipe EGI qui ont participé à mes travaux de thèse, Grégory, Stéphanie, Gaetan, Jeff, et Fred, le comptage de pucerons c'est quand même mieux à plusieurs !

Merci à tous les membres de l'équipe EGI pour ces trois années passées dans la bonne humeur.

Merci à ma famille de m'avoir toujours soutenue pendant ces longues années d'études !

Merci aux amis, surtout aux biches Dadou, Mayon, Anne-so, de n'avoir pas changé, d'être toujours là, depuis bientôt 10 ans que l'on s'est rencontrées sur les bancs de la fac ! NEM NEM

Merci à Elisabeth Huguet, de m'avoir, un jour en licence 3, ouvert les portes de son labo, me faisant ainsi découvrir le monde merveilleux des interactions hôtes-parasites. Vous êtes LA personne à qui je dois d'être là, sans notre rencontre je n'aurais peut-être jamais fait de thèse... MERCI POUR TOUT.

Et enfin, merci à Jérémie.... POUR TOUT !

J'espère que mon manuscrit de thèse sera un assez beau cadeau en ce 3 Novembre, jour de tes.... (non jvais pas le dire), alors JOYEUX ANNIVERSAIRE !!!! (pas eu le temps de prévoir d'autres cadeaux de toute façon ^^)

# TABLE DES MATIERES

I. Introduction générale.....	1
1. Evolution des insectes herbivores .....	2
1.1 Coévolution antagoniste .....	2
1.2 Les sauts d'hôte .....	5
2. Bases moléculaires de l'adaptation des insectes herbivores .....	6
2.1 Chimioréception .....	6
2.2 Défenses des plantes et effecteurs des insectes herbivores .....	7
3. Interactions plantes-pucerons .....	11
3.1 Caractéristiques générales des pucerons .....	11
3.2 Dialogue moléculaire lors des interactions plantes-pucerons .....	16
4. Modèle d'étude : <i>Acyrtosiphon pisum</i> , le puceron du pois .....	22
5. Objectifs des travaux de thèse .....	25
II. Paper 1: Fast evolution and lineage-specific gene family expansions of putative aphid effectors shed light on molecular evolution driven by aphid-plant interactions .....	27
III. Functional characterization of candidate salivary genes that are potentially involved in the pea aphid-plant interactions.....	77
1. Context.....	77
2. Paper 2: Optimization of agroinfiltration in <i>Pisum sativum</i> provides a new tool for studying the salivary protein functions in the pea aphid complex .....	81
3. Functional characterization of salivary effector candidates in the <i>A. pisum</i> -pea plant system .....	91
4. Discussion and perspectives on functional characterization of salivary effectors .....	99
5. Supplementary materials.....	103
IV. Discussion générale et perspectives .....	111
1. Effecteurs salivaires du puceron du pois .....	111
1.1 Identification des effecteurs candidats .....	111
1.2 Apports de l'analyse protéomique de la salive .....	113
2. Evolution des effecteurs salivaires et adaptation à la plante hôte .....	115
2.1 Les effecteurs aphidiens vus à une échelle macro-évolutive .....	116
2.2 Les effecteurs aphidiens vus à une échelle micro-évolutive .....	120
2.3 Adaptation à la plante hôte .....	123
3. Outils et stratégies pour une meilleure compréhension des interactions moléculaires dans le système puceron du pois-légumineuses .....	124
3.1 Différents types d'effecteurs pour différentes fonctions ? .....	124
3.2 Outils et stratégies pour caractériser la fonction des différents effecteurs dans les interactions plantes-pucerons .....	125
4. Implications évolutives et appliquées .....	129
5. Conclusion .....	130
V. Références bibliographiques.....	131

## **ABBREVIATIONS**

**ACE:** Angiotensin-converting-enzyme  
**apN:** Aminopeptidase N  
**Avr:** Avirulence  
**BEB:** Bayes empirical Bayes  
**CPM:** Count per million  
**CRP:** Cystein-rich-protein  
**CSP:** Chemosensory protein  
**DAMP:** Damage-associated-molecular-pattern  
**DIMBOA:** 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one  
**ET:** Ethylène  
**ETI:** Effector-triggered-immunity  
**ETS:** Effector-triggered-susceptibility  
**FPKM:** Fragment per kilobase million  
**GFP:** Green fluorescent protein  
**GOX:** Glucose oxidase  
**GR:** Gustatory receptor  
**GST:** Glutathione S-transferase  
**HAMP:** Herbivore-associated-molecular-pattern  
**HGT:** Horizontal gene transfer  
**HIGS:** Host-induced-gene-silencing  
**HTI:** HAMP-triggered-immunity  
**IR:** Inotropic receptor  
**JA:** Acide jasmonique  
**MAPK:** Mitogen-activated-protein-kinase  
**MS:** Mass spectrometry  
**NB-LRR:** Nucleotide binding-leucine rich repeat  
**NSP:** Nitril-specifier-protein  
**OBP:** Odorant-binding protein  
**OPDA:** 12-oxo-phytodienoic acid  
**OR:** Olfactory receptor  
**PAD4:** PHYTOALEXIN DEFICIENT 4  
**PAMP:** Pathogen-associated-molecular-pattern  
**PC:** Parthénogénèse cyclique  
**PO:** Parthénogénèse obligatoire  
**PP1:** Phloem protein 1  
**PP2:** Phloem protein 2  
**PTI:** PAMP-triggered-immunity  
**QTL:** Quantitative trait locus  
**ROS:** Reactive oxygen species  
**SA:** Acide salicylique  
**SG:** Salivary gland  
**SHP:** Sheath protein  
**SSGP:** Secreted salivary gland protein  
**T3SS:** Type III secretion system  
**VIGS:** Virus-induced-gene-silencing  
**VPS52:** Vacuolar protein sorting-associated protein 52  
**WIR:** Wound-induced-resistance

# I. Introduction générale

Comprendre les mécanismes sous-jacents à l'adaptation des organismes à leur environnement est une question centrale en biologie évolutive. L'adaptation en tant que processus peut être définie comme l'ensemble des modifications anatomiques, physiologiques ou comportementales d'une population d'individus sous l'effet de la sélection naturelle. Ainsi, les nouveaux états de ces différents caractères améliorent la survie et le succès reproductif des individus (i.e. *fitness*) qui en sont porteurs dans un environnement particulier. Les récents progrès des technologies « omiques » et les développements méthodologiques en génétique des populations offrent aujourd'hui la possibilité d'utiliser de nouvelles approches théoriques et conceptuelles pour mettre en évidence les bases génomiques de l'adaptation des organismes dans un grand nombre d'écosystèmes (Stapley et al. 2010; Savolainen et al. 2013; Bindschedler et al. 2016).

Les insectes représentent le plus large groupe de métazoaires de la planète, et il apparaît aujourd'hui évident que l'évolution de l'herbivorie a joué un rôle fondamental dans la diversification des espèces de ce groupe (Mitter et al. 1988; Mayhew 2007; Wiens et al. 2015). A ce jour, un million d'espèces d'insectes sont recensées et la moitié d'entre-elles se nourrit sur des plantes (Wu and Baldwin 2010). De plus, Wiens et al. (2015) ont montré un taux de diversification plus élevé chez les insectes herbivores, expliquant ainsi la richesse d'espèces de ce clade regroupant plus de la moitié des espèces vivantes.

Les interactions entre les plantes et les insectes herbivores durent depuis plus de 350 millions d'années (Gatehouse 2002). Au cours de leur évolution, les insectes herbivores ont établi des relations très intriquées avec leurs plantes-hôtes, qui constituent à la fois un habitat, une source d'alimentation et un lieu de reproduction. Pour cela, ils ont dû s'adapter à la phénologie, à la composition en nutriments et aux défenses physico-chimiques de leurs plantes hôtes (Fraenkel 1959). Par conséquent, les insectes herbivores sont souvent adaptés à une gamme restreinte d'hôtes (i.e. spécialistes), ainsi capables de se nourrir sur des plantes proches phylogénétiquement et/ou de compositions biochimiques similaires (Futuyma and Moreno 1988; Futuyma and Agrawal 2009; Forister et al. 2012). Cependant, des espèces généralistes existent et représentent environ 10% des espèces d'insectes herbivores (Bernays and Graham 1988). Il est à noter que ces deux stratégies, sont les deux extrêmes d'un continuum et qu'il existe bien souvent une variation du spectre d'hôte au sein des espèces, rendant cette estimation peu fiable (Loxdale and Harvey 2016). Par la suite, nous parlerons d'espèces généralistes



quand ces dernières sont capables d'exploiter de nombreuses plantes appartenant à différentes familles botaniques et qu'elles ne montrent pas de populations spécialisées.

Décrire les bases, aussi bien évolutives que fonctionnelles, de la spécialisation à la plante hôte chez les insectes herbivores représente un défi majeur dans la compréhension de l'adaptation des espèces à leur environnement et offre la possibilité d'éclaircir les processus qui ont conduit à une grande diversification des métazoaires sur la planète.

## 1. Evolution des insectes herbivores

Au cours du processus de spécialisation des insectes herbivores, des populations génétiquement distinctes et adaptées à différentes plantes-hôtes peuvent évoluer suite à une pression de sélection divergente imposée par les nouvelles niches écologiques (e.g. spécialisation et différenciation des populations en races hôtes, biotypes ou écotypes). L'établissement de barrières reproductives entre ces populations peut conduire à la formation d'espèces distinctes (Peccoud et al. 2009a; Powell et al. 2013).

Cette spécialisation des insectes herbivores à une plante hôte particulière peut se produire lors d'interactions antagonistes réciproques entre plantes et insectes (i.e. coévolution) ou lorsque les insectes colonisent de nouvelles plantes par le biais de sauts d'hôtes (Ehrlich and Raven 1964; Feder et al. 2003; Janz 2011).

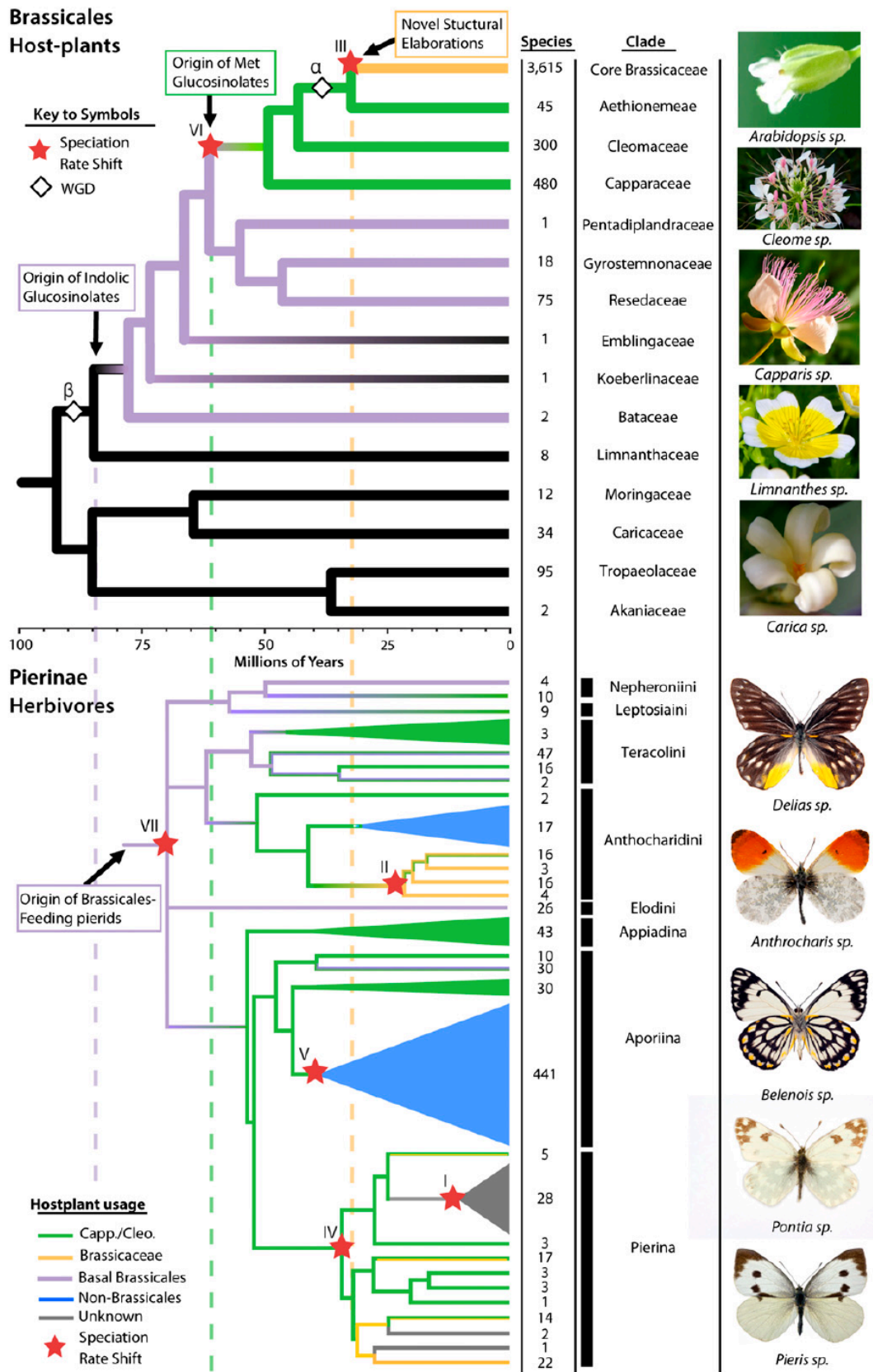
### 1.1 Coévolution antagoniste

Lors d'une interaction antagoniste (relation hôte-parasite), la sélection retient les stratégies qui permettent à chaque espèce d'augmenter sa valeur sélective, ou *fitness* (e.g. survie et reproduction), et ce, aux dépens de celle de l'autre. Le parasite développe des stratégies afin d'exploiter au mieux son hôte. A l'inverse, chez l'hôte, des stratégies de défense permettant de réduire les dommages causés par le parasite vont être sélectionnées.

C'est en examinant les relations entre les Papilionidés (Lepidoptères) et leur plantes hôtes, que Ehrlich et Raven (1964) introduisent pour la première fois le terme de coévolution. Leur théorie repose sur le fait que certains papillons phytophages ont exercé une pression de sélection sur les plantes, qui en réponse ont développé certains mécanismes de défense, entraînant en retour l'émergence, chez certains papillons, de caractères permettant de contrer ces défenses. Ehrlich et Raven émettent alors

l'hypothèse qu'une sélection réciproque a joué un rôle fondamental dans l'évolution de ces espèces, favorisant ainsi l'apparition de nouvelles lignées de plantes et de papillons. En 1977, Leigh Van Valen affirme que la persistance d'une espèce dans un environnement variable n'est possible que par un effort adaptatif permanent (la fameuse théorie de la Reine rouge inspirée de Lewis Carroll et son roman *Alice au pays des merveilles*). Dans le cas des interactions hôte-parasite, une coadaptation s'instaure dans le temps évolutif, poussant les organismes à se modifier réciproquement pour rester à la même place sur le plan adaptatif. La théorie de la Reine rouge est ainsi une métaphore symbolisant la course aux armements qui peut exister entre les espèces (Van Valen 1977). Cette course aux armements favorise alors de façon réciproque la diversification des hôtes et des parasites à travers le développement de nouveautés évolutives, *key innovations* (Simpson 1953). Finalement, en 1980 le terme de coévolution est redéfini comme étant le changement évolutif d'un trait chez une espèce en réponse à un trait d'une deuxième espèce, suivie d'une contre-adaptation évolutive chez cette dernière (Janzen 1980).

L'hypothèse proposant que la grande diversité des espèces de papillons est le résultat d'une coévolution avec leurs plantes hôtes, soutenue par Ehrlich et Raven (1964), a longtemps été remise en question du fait de l'existence de controverses dans la datation de l'origine des lépidoptères (Braby et al. 2005). En effet, pour retracer la réciprocité et l'enchaînement des adaptations/contre-adaptations supportant l'hypothèse de coévolution, la datation est primordiale. Récemment, une telle coévolution des insectes herbivores et leurs plantes hôtes a été décrite entre les papillons de la famille des Piéridés et les plantes de l'ordre des Brassicales (Wheat et al. 2007; Edger et al. 2015). Ainsi, un système de détoxification des glucosinolates (métabolites secondaires spécifiques des Brassicales, anciennement crucifères) propre aux Piéridés, les *nitrile-specifier proteins* (NSP), a été identifié comme innovation clé dans l'adaptation et la diversification de ces papillons (Wheat et al. 2007). Les NSP semblent avoir évolué peu après l'évolution du groupe des plantes hôtes et la diversification de leurs activités biochimiques suit la distribution des glucosinolates chez les Brassicales (Figure 1). De plus, les principales innovations à l'origine de cette détoxification des glucosinolates sont associées à des duplications de gènes, impliquant notamment des duplications de NSP (Edger et al. 2015). Ces études apportent les preuves d'une course aux armements graduelle entre les Piéridés et les Brassicales, et permettent d'établir un lien entre diversité des espèces, coévolution, et importance des duplications géniques en tant que sources de nouveautés biologiques.



**Figure 1.** Coévolution entre insectes herbivores et leurs plantes hôtes. Phylogénie datée des Brassicales et des Pieridés indiquant les différentes innovations et les diversifications (Edger et al. 2015).

## 1.2 Les sauts d'hôte

Le saut d'hôte, ou *host shift*, peut être décrit comme tout gain d'un nouvel hôte par une population d'insectes herbivores (Forbes et al. 2017; Vertacnik and Linnen 2017). Ce saut d'hôte peut s'accompagner d'un changement dans la gamme d'hôtes de l'insecte. L'expansion de la gamme d'hôte survient quand le changement d'hôte n'implique pas la perte d'un hôte ancestral. En revanche, une réduction de la gamme d'hôte peut apparaître quand le changement d'hôte induit la perte d'hôtes ancestraux. Les populations ont alors un degré de spécialisation plus fort que la population ancestrale (Vertacnik and Linnen 2017). Au sein d'une espèce, des populations ayant acquis un nouvel hôte peuvent diverger de celles n'ayant subi aucun changement, et un isolement reproducteur de ces populations peut aboutir à l'émergence de nouvelles espèces (e.g. spéciation écologique). Ainsi, le phénomène de changement d'hôte peut participer à la diversification des espèces d'insectes herbivores.

Un exemple de changement d'hôte est décrit chez les mouches Tephritidés appartenant au complexe d'espèces sœurs *Rhagoletis pomonella*. Les mouches *Rhagoletis* d'Amérique du Nord ont opéré un changement d'hôte de leur hôte natif *Crataegus mollis* vers un nouvel hôte, le pommier (Bush 1966). Les différentes populations de mouches ont évolué et forment maintenant un continuum comprenant des races d'hôte en isolement reproducteur partiel (McPheron et al. 1988), et des espèces sœurs lorsque l'isolement reproducteur est plus fort (Powell et al. 2013). A la base de ce changement, des flux de gènes entre 2 populations de *Rhagoletis* – Mexicaine et Nord-Américaine séparées il y a 1,57 millions d'années – affectant le polymorphisme de traits liés de la diapause, ont été identifiés. Les variations de diapause ainsi créées ont permis aux populations Nord-Américaines de s'adapter à différents hôtes avec des phénologies différentes (Feder et al. 2003).

Un phénomène de changement d'hôte a également été décrit chez le puceron vert du pêcher *Myzus persicae*. Une sous espèce de ce puceron, résistante à la nicotine, a récemment colonisé un nouvel hôte, le tabac. Le mécanisme sous-jacent à ce changement d'hôte a été identifié comme lié à une surexpression d'un cytochrome P450 permettant de détoxifier la nicotine. La surexpression de cette enzyme est causée par une très importante amplification génique (100 copies du gène CYP6CY3 codant pour le cytochrome P450) et des modifications dans les promoteurs (Puinean et al. 2010; Bass et al. 2013; Bass et al. 2014).

Bien que les insectes herbivores soient depuis longtemps des modèles d'étude de l'évolution et de la spécialisation écologique, seules les récentes avancées des techniques de séquençage haut débit ont permis d'identifier des loci qui contribuent à l'adaptation des insectes herbivores à leur plante hôte. Ainsi, de nombreux gènes candidats liés à l'adaptation ont été identifiés chez les insectes herbivores

ces dernières années, révélant certains traits déterminants de ces interactions plantes-insectes (Vertacnik and Linnen 2017).

## 2. Bases moléculaires de l'adaptation des insectes herbivores

Parmi les déterminants qui peuvent être impliqués dans l'adaptation à la plante hôte, il convient d'identifier les différents mécanismes qui contribuent à la capacité d'un insecte à localiser et exploiter ses hôtes. Ainsi, la détection de la plante-hôte, la suppression des défenses végétales, la détoxification des métabolites secondaires et les symbioses microbiennes influencent l'issue des interactions plantes-insectes herbivores (Simon et al. 2015), sans oublier les autres interactions à des niveaux trophiques supérieurs ou inférieurs (Biere and Bennett 2013; Simon et al. 2017).

### 2.1 Chimioréception

La chimioréception, qui met en jeu le système antennaire, les poils olfactifs des pièces buccales des insectes et le système nerveux central, joue un rôle majeur dans les interactions plantes-insectes en guidant les insectes vers une plante qui représente une ressource favorable. En effet, les plantes émettent un grand nombre de composés volatiles qui sont reconnus et permettent dans un premier temps aux herbivores de s'orienter vers leurs plantes hôtes. Ensuite, la perception d'autres molécules lors du contact direct avec la plante détermine les préférences des insectes herbivores. La reconnaissance des composés chimiques des plantes repose principalement sur des récepteurs olfactifs (*Olfactory receptors* ORs, *Ionotropic receptors* IRs, *Chemosensory proteins* CSP, *Odorant-binding proteins* OBPs) et des récepteurs gustatifs (*Gustatory receptors* GRs) (Missbach et al. 2014; Eyun et al. 2017). Ainsi, différentes études ont montré que les familles de gènes codants pour les ORs et les GRs évoluent rapidement, et par des mécanismes dynamiques de gain/perte de gènes chez les drosophiles, en parallèle de l'adaptation de celles-ci à de nouvelles niches écologiques (Vieira et al. 2007; Gardiner et al. 2008; Ramasamy et al. 2016). Chez les lépidoptères, l'expansion de ces deux familles multigéniques pourrait avoir contribué à l'adaptation à diverses plantes hôtes consécutivement à la radiation des angiospermes (Engsontia et al. 2014).

## 2.2 Défenses des plantes et effecteurs des insectes herbivores

### *Défenses des plantes en réponse aux insectes herbivores*

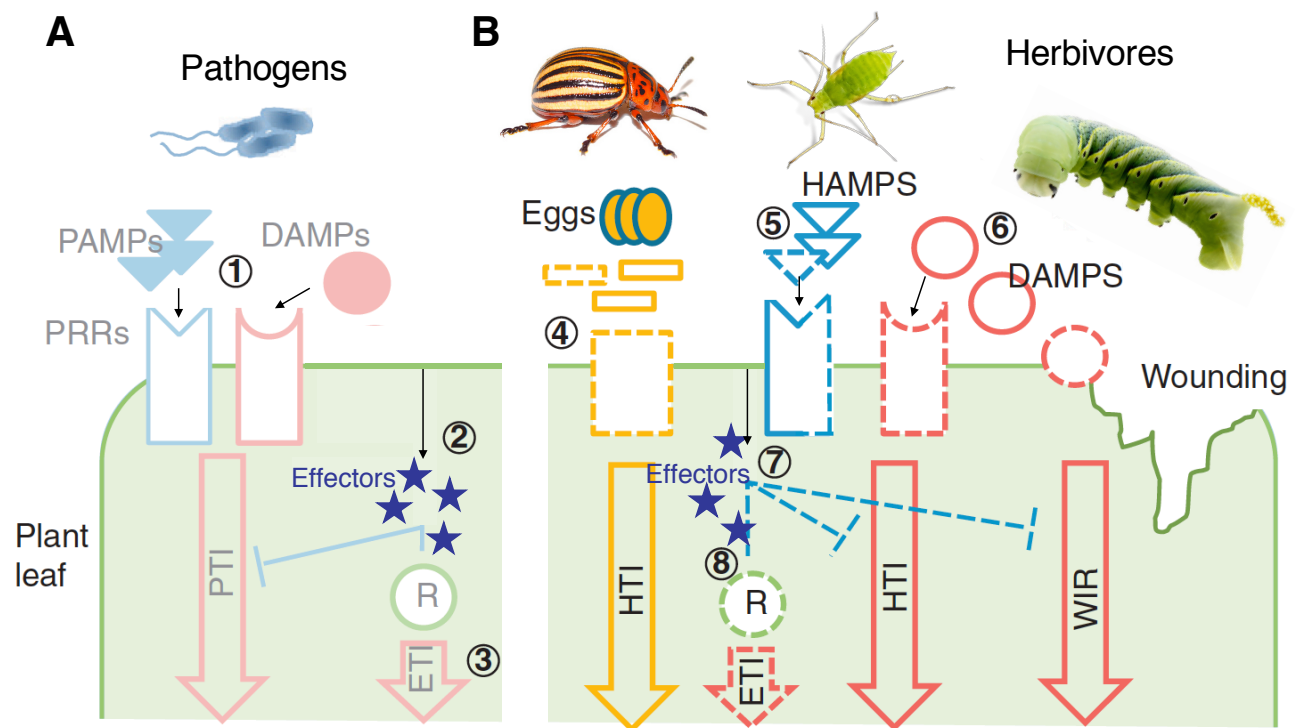
Les plantes possèdent des défenses constitutives contre les insectes herbivores, qui peuvent être des barrières physiques (comme les trichomes ou cires) ou chimiques (alcaloïdes comme la nicotine chez le tabac). Ces défenses constitutives sont utilisées de concert avec les défenses induites pour lutter efficacement contre les insectes herbivores. Différents travaux ont montré que les plantes utilisent une perception rapide et précise des herbivores ou pathogènes pour mettre en place ces défenses induites, spécifiques des parasites qui les attaquent (Howe and Jander 2008). Lorsque les insectes herbivores causent d'importantes blessures, les plantes produisent des molécules comme les DAMPs (*Damage-Associated-Molecular-Patterns*) qui vont déclencher les réactions de défenses (Figure 2). Les insectes herbivores produisent également certaines molécules qui sont reconnues par la plante et activent les réactions de défenses, les HAMPs (*Herbivore-Associated Molecular Patterns*). Ces HAMPs peuvent être présents dans les sécrétions orales, la salive, les cires cuticulaires et les sécrétions de ponte (Bonaventure et al. 2011; Acevedo et al. 2015; Hilker and Fatouros 2015; Schmelz 2015). Lorsque ces patterns sont reconnus, les plantes expriment des réactions de défenses rapides généralement non-spécifiques, qui induisent par exemple la production de composés toxiques (*reactive oxygen species*, ROS). Ces réponses sont principalement dépendante de la voie d'influx calciques dans les cellules ou encore de l'activation de la voie de signalisation de MAPK (*Mitogen activated protein kinase*) (Howe and Jander 2008).

Ensuite des réponses plus spécifiques peuvent être régulées par l'action des phytohormones, principalement l'acide jasmonique (JA), l'éthylène et l'acide salicylique (SA) (Erb et al. 2012). Ainsi, les voies JA et ET sont connues pour induire les réponses qui sont efficaces contre les insectes herbivores mais également contre les pathogènes nécrotrophes. La voie de signalisation SA joue quant-à-elle un rôle fondamental dans la défense contre les pathogènes biotrophes (Erb et al. 2012; Pieterse et al. 2012; Kazan and Lyons 2014). Une interaction antagoniste entre les voies JA et SA (*crosstalk*) semble très répandue dans le règne végétal, même si ce modèle tend à se complexifier au fur et à mesure que de nouvelles interactions sont étudiées et révèlent le rôle des autres phytohormones (Robert-Seilaniantz et al. 2011). Ainsi, les herbivores ou pathogènes peuvent manipuler ces voies pour leur propre bénéfice (Diezel et al. 2009; Chung et al. 2013; Kazan and Lyons 2014).

La voie JA est connue pour induire des défenses spécifiques directes ou indirectes en réponse à l'attaque des insectes herbivores. Ainsi, des métabolites secondaires (e.g. glucosinolates ou alcaloïdes) et des protéines de défenses – polyphénol oxydase (PPO), lipooxygénase (LOX), ou encore peptidase et phosphatases – peuvent avoir des effets anti-appétants, anti-digestifs ou toxiques directs

sur les insectes. De plus, la voie JA permet d'induire la production de composés volatils qui peuvent soit attirer des prédateurs et parasitoïdes de l'insecte herbivore, soit induire des réactions de défenses dans d'autres parties de la plante (voire chez les autres plantes alentours ; Das et al. 2013).

Cette description des défenses mises en place par les plantes pour lutter contre les herbivores étant simplifiée et non-exhaustive, plus de détails peuvent être trouvés dans les nombreuses revues qui traitent de ce sujet (Gatehouse 2002; Freeman 2008; Howe and Jander 2008; Wu and Baldwin 2010; Erb et al. 2012; War et al. 2012; Fürstenberg-Hägg et al. 2013).



**Figure 2.** (A) Modèle des réactions de défense des plantes contre les pathogènes proposé par Jones and Dangl (2006). (1) Les patterns moléculaires associés aux agents pathogènes et des dommages physiques (PAMPs et DAMPs) sont reconnus par des récepteurs des plantes ce qui déclenche le premier niveau de défense (PTI). (2) Ensuite, certains effecteurs des pathogènes suppriment la réponse PTI. (3) Des gènes de résistance peuvent alors reconnaître spécifiquement les effecteurs, et déclencher l'ETI. (B) Modèle des réactions de défense adapté aux insectes herbivores. (4-5) Les composés associés aux herbivores (HAMPs et composés des oeufs) sont reconnus par les récepteurs et entraînent le premier niveau d'immunité contre les herbivores HTI. (6) Les DAMPs, libérés lors des blessures infligées par les herbivores sont également reconnus et induisent la *wound-induced resistance* (WIR). (7) Les effecteurs sécrétés par les insectes peuvent supprimer l'HTI et le WIR. (8) A leur tour, les effecteurs d'insectes peuvent être reconnus et la réponse ETI se met en place. Adapté d'après Erb et al. (2012).

### ***Relation gène-pour-gène et effecteurs des insectes***

Jones and Dangl (2006) ont proposé un modèle théorique en zigzag décrivant différents niveaux dans l'immunité anti-pathogène des plantes, ce modèle peut par extension s'appliquer aux interactions plantes-insectes herbivores (Figure 2). Lors de la première phase, les PAMPs, (*Pathogen-Associated Molecular Patterns*), DAMPs et HAMPs reconnus déclenchent le premier niveau de défense, respectivement PTI (*PAMP-Triggered-Immunity*), WIR (*Wound-Induced Resistance*) et HTI (*HAMP-Triggered-Immunity*). Ensuite, durant la phase 2, les herbivores (Figure 2B) – comme les pathogènes (Figure 2A) – sécrètent des effecteurs qui interfèrent avec le premier niveau de défense HTI/PTI, ce qui entraîne l'Effector-Triggered-Susceptibility (ETS). Dans la troisième phase, un des effecteurs (protéine d'avirulence, Avr) de l'herbivore est reconnu par un récepteur NB-LRR (*nucleotide binding-leucine rich repeat*), codé par un gène de résistance (R) de la plante, ce qui conduit à l'*Effector-Triggered-Immunity* (ETI) (Figure 2). Cette relation gène-pour-gène Avr-R (Kerr 1987), est une composante majeure de la course aux armements entre les herbivores/pathogènes et leurs plantes hôtes.

La salive et les sécrétions orales des insectes herbivores peuvent donc contenir des molécules dont certaines déclenchent les défenses des plantes et d'autres qui permettent de manipuler la physiologie de l'hôte et de contrer les défenses, les effecteurs. L'étude des effecteurs des insectes herbivores est un domaine relativement nouveau, ils sont donc beaucoup moins connus que ceux des pathogènes bactériens ou fongiques. Néanmoins, certaines protéines comme la glucose oxidase (GOX) sécrétée par les chenilles de la famille des Noctuidés (Eichenseer et al. 2010), les SSGP-71 sécrétées par la mouche de Hesse, *Mayetiola destructor* (Zhao et al. 2015) ou encore des protéines sécrétées par les pucerons (Hogenhout and Bos 2011; Rodriguez and Bos 2012) ont été caractérisés comme des effecteurs modulant les défenses de la plante et contribuant à la *fitness* de l'insecte. Certains gènes de résistances aux insectes herbivores ont été identifiés chez les plantes, dirigés notamment contre les pucerons (Dogimont et al. 2010, voire partie I.3.2), la mouche de Hesse (Stuart 2015) et la cicadelle brune *Nilaparvata lugens* (Du et al. 2009; Hu et al. 2016). Ces gènes de résistances codent bien souvent pour des récepteurs de type NB-LRR qui sont aussi impliqués dans la résistance aux pathogènes (Jones and Dangl 2006). En revanche, l'association Avr-R n'est pleinement décrite que chez *M. destructor*, impliquant des protéines sécrétées dont deux effecteurs salivaires SSGP-71, une phosphatase et une protéine de fonction inconnue, respectivement reconnues par les gènes de résistance *H6*, *H9*, *H24* et *H13* chez le blé (Aggarwal et al. 2014; Zhao et al. 2015; Zhao et al. 2016). Les mécanismes exacts de reconnaissance restent néanmoins à éclaircir.



### ***Détoxification des métabolites secondaires***

Les insectes herbivores ont développé des activités enzymatiques leur permettant de métaboliser les métabolites secondaires qui sont constitutivement produits par les plantes ou induits lors de l'interaction plante-herbivore (Simon et al. 2015). La détoxification des composés toxiques a principalement été étudiée chez les drosophiles (Yassin et al. 2016) mais aussi chez les lépidoptères, mettant en évidence des mécanismes d'amplification génique par duplications (Edger et al. 2015). Les enzymes associées à la détoxification sont le plus souvent des cytochrome P450, des glutathione-S-transferase, ou des familles d'enzymes spécialisées dans la détoxification de certains groupes de glucosinolates (Mao et al. 2007; Celorio-Mancera et al. 2012; Edger et al. 2015; Giraud et al. 2015).

D'autres observations chez les pucerons montrent un lien entre l'expansion des familles de gènes codant pour les enzymes de détoxification et l'adaptation à la plante hôte. Ainsi, chez *M. persicae* la détoxification de la nicotine est assurée par une forme de cytochrome P450 dont le gène est fortement dupliqué dans le génome de la sous-espèce adaptée au tabac (Puinean et al. 2010; Bass et al. 2013; Bass et al. 2014).

Ces mécanismes de détoxification sont étroitement liés à l'adaptation des insectes herbivores à leurs plantes hôtes en étant associés bien souvent à des processus de coévolution antagoniste (Edger et al. 2015) ou de changement d'hôte (Bass et al. 2013) propices à l'émergence de nouvelles espèces.

### ***Influence de la communauté microbienne***

En plus des déterminants moléculaires (plantes et insectes) décrits précédemment, les interactions plantes-insectes sont grandement influencées par la communauté bactérienne/microbienne hébergée par les insectes herbivores et les plantes, qu'elle soit en association symbiotique ou non (Frago et al. 2012; Sugio et al. 2015; Giron et al. 2017; Shikano et al. 2017).

Ainsi, la communauté microbienne hébergée par les insectes peut contribuer à la diversification et à la spécialisation à l'hôte (Hansen and Moran 2014). Les interactions plantes-insectes-microbes représentent aussi un terrain propice aux transferts horizontaux de gènes (HGT) qui sont souvent à l'origine de l'acquisition de nouvelles fonctions chez les insectes, leur permettant d'exploiter de nouvelles niches écologiques (Acuña et al. 2012). La communauté microbienne qu'ils hébergent peut également interférer avec les défenses de la plante. Ainsi, le déclenchement de la signalisation SA-dépendante par la communauté microbienne peut court-circuiter (e.g. *crosstalk*) les défenses JA-dépendantes – normalement les plus efficaces contre les herbivores – ce qui profite à l'insecte et peut impacter positivement sa *fitness* sur l'hôte (Diezel et al. 2009; Chung et al. 2013). Cependant, la communauté microbienne impliquée dans les interactions plantes-insectes n'influence pas toujours les interactions en faveur de l'herbivore (Sugio et al. 2015). Des bactéries peuvent ainsi déclencher des

défenses qui sont défavorables à l'insecte herbivore, diminuant ainsi sa *fitness* (Chaudhary et al. 2014). Enfin, certaines bactéries associées aux plantes hôtes peuvent produire des toxines qui vont aussi diminuer la *fitness* des herbivores (Vachon et al. 2012).

### 3. Interactions plantes-pucerons

#### 3.1 Caractéristiques générales des pucerons

L'ordre des Hémiptères, qui comprend entre 50 000 et 80 000 espèces d'insectes piqueurs-suceurs, forme le plus large groupe d'insectes hétérométaboles (i.e. à métamorphose incomplète), dont 90% des espèces sont herbivores (Dolling 1991). Grâce à des parties buccales spécialisées, notamment les stylets, les pièces buccales des hémiptères herbivores peuvent naviguer entre les cellules végétales pour atteindre les tubes criblés du phloème et se nourrir de la sève élaborée.

Les pucerons appartiennent au sous-ordre des Aphidomorpha, et la famille des Aphididae est celle qui contient la majorité des espèces. A ce jour, plus de 4000 espèces de pucerons ont été décrites et comptent parmi les ravageurs de cultures les plus destructeurs des régions tempérées de la planète. Ainsi, 250 espèces de pucerons sont ravageurs d'espèces végétales d'intérêt agricole ou sylvicole (Blackman and Eastop 1994; Blackman and Eastop 2000). Les dommages engendrés directement par les pucerons sur les cultures sont dus au prélèvement de la sève, qui réduit les ressources disponibles pour la plante et contribue à diminuer les rendements. Les dommages les plus conséquents, sont ceux indirectement causés par la transmission de virus phytopathogènes dont les pucerons sont d'importants vecteurs. Les pucerons sont impliqués dans 50% des transmissions de virus vectorisés par les insectes aux plantes (Ng and Perry 2004; Harris and Maramorosch 2014). Ainsi, le puceron vert du pêcher *Myzus persicae*, herbivore généraliste, peut transmettre plus de 100 virus différents à sa large gamme d'hôtes (Blackman and Eastop 2000). Leur cycle de vie particulier, leur importante plasticité phénotypique, le développement de symbioses mutualistes avec des bactéries et la sécrétion d'effecteurs salivaires permettant de contourner les défenses de la plante hôte, sont autant de caractéristiques des pucerons qui en font de redoutables ravageurs.

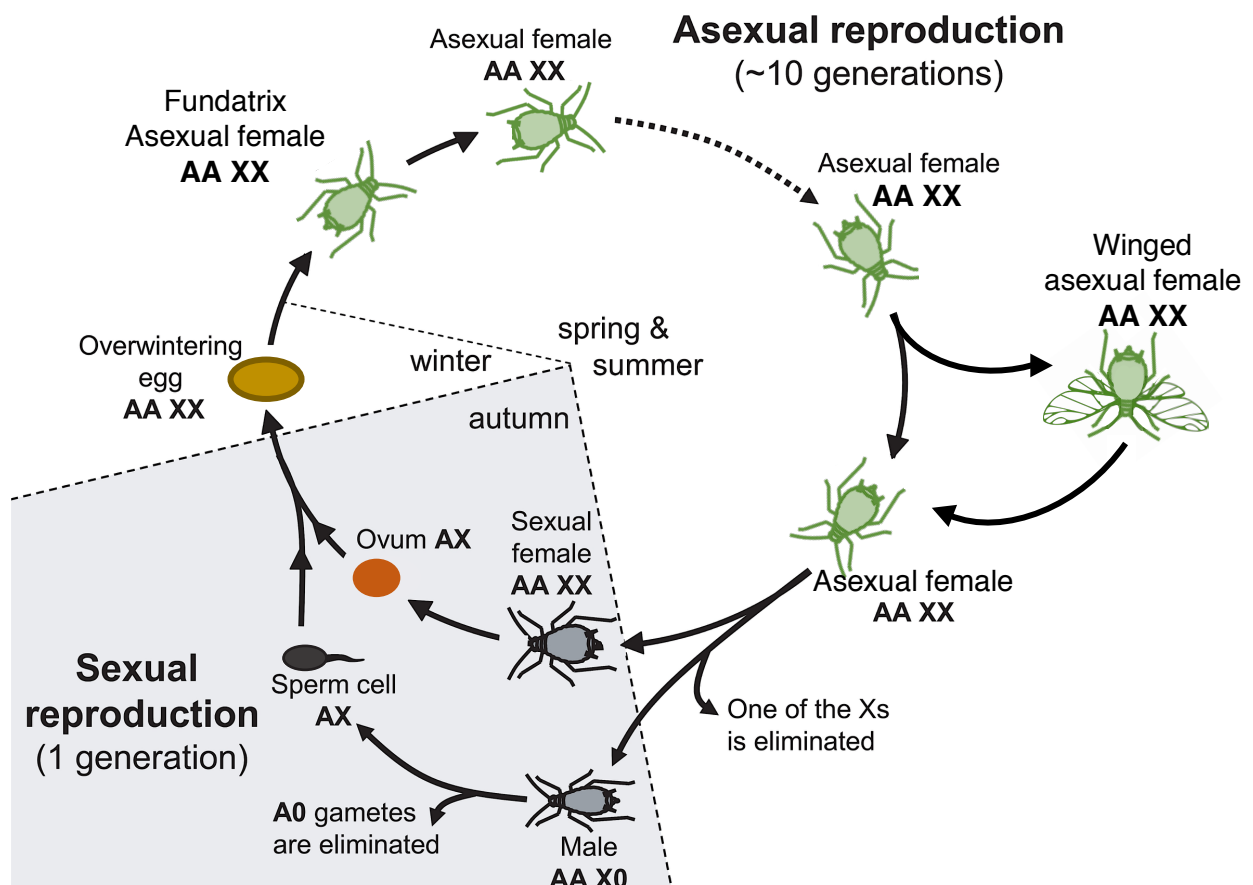
## ***Cycle de vie***

Une des caractéristiques qui contribuent au succès des pucerons en tant que ravageurs de plantes, réside dans la mise en place d'un cycle de vie complexe permis par une importante plasticité phénotypique (Hardy et al. 2015). Au cours de leur cycle de vie, ils expriment des phénotypes différents à partir d'un même matériel génétique (i.e. polyphénisme). Ainsi, le polyphénisme de dispersion permet la production de formes aptères ou ailées au sein de la même lignée clonale (Braendle et al. 2006), et le polyphénisme de reproduction permet l'alternance de formes sexuées et asexuées, ces dernières étant produites par parthénogénèse apomictique ou reproduction clonale (Simon et al. 2010).

Lorsque qu'il y a réalisation d'un cycle complet avec alternance entre reproduction sexuée et asexuée, on parle alors de parthénogénèse cyclique (PC). La phase sexuée a pu être perdue dans diverses sous-familles de pucerons, certaines espèces ne se reproduisent plus que par parthénogénèse dite obligatoire (PO). La PO est favorisée par des hivers doux, ainsi la plupart des espèces tropicales ont perdu la phase sexuée de leur cycle, mais cela existe aussi chez certains pucerons ravageurs des milieux tempérés. De plus, une variabilité de cycles peut co-exister au sein de la même espèce dont certaines populations se reproduisent par PC et d'autres par PO en fonction de la rigueur de l'hiver et de la disponibilité des plantes hôtes. Ce phénomène est par exemple observé chez *M. persicae*, le puceron du merisier à grappes *Rhopalosiphum padi* et le puceron de la pomme de terre *Macrosiphum euphorbiae*, qui sont d'importants ravageurs (Blackman and Eastop 2000).

Les espèces à PC qui exploitent la même espèce ou famille d'hôte durant les phases sexuée et asexuée sont dites monœciques. En revanche, 10% des espèces de pucerons effectuent leur phase de reproduction sexuée sur un hôte primaire – généralement une plante ligneuse – et leur phase asexuée sur un hôte secondaire le plus souvent une plante herbacée, elles sont dites diœciques. Le déterminisme du sexe des pucerons se fait par un système X0, ou les femelles possèdent 2 chromosomes sexuels (XX) et les mâles un seul (X0) alors que les deux sexes ont des chromosomes autosomes diploïdes AA (Hales et al. 2002). La Figure 3 illustre le cycle de vie à PC monœcique du puceron du pois *Acyrtosiphon pisum*. Les différentes phases décrites sont les mêmes pour les espèces de pucerons monœciques, excepté le changement d'hôte. Ainsi, les pucerons passent l'hiver sous forme d'œuf (œuf d'hiver) – sur l'hôte primaire pour les diœciques – duquel une femelle fondatrice va émerger au printemps. Celle-ci va ensuite produire, par parthénogénèse vivipare, d'autres femelles qui se reproduiront elles aussi de manière clonale pendant plusieurs générations (Figure 3). Pour les espèces diœciques, la colonie qui se développe sur l'hôte primaire au printemps va produire des femelles parthénogénétiques ailées qui vont coloniser les hôtes secondaires. Durant l'été, les générations clonales se succèdent. Lorsque la plante entre en sénescence ou qu'il y a une surpopulation de pucerons

sur celle-ci, de nouvelles femelles ailées émergent afin de disperser sur de nouvelles plantes (hôtes secondaires). C'est durant cette phase (reproduction clonale combinée à de grandes capacités de dispersion) que les populations sont les plus importantes, correspondant aux pics d'infestation. En réponse à la diminution de la longueur du jour en automne, des femelles sexupares sont produites et vont donner naissance à la fois à des femelles sexuées (XX) et à des mâles (X0 après l'élimination d'un des deux chromosomes X). Ces femelles s'accouplent alors avec les mâles – unique reproduction sexuée du cycle – et vont ensuite pondre un œuf d'hiver (sur hôte primaire pour les espèces diéciques). De cet œuf qui peut résister à des températures très basses en hiver (jusqu'à -30°C), émergera au printemps une nouvelle femelle fondatrice (XX). Le mâle ne pouvant transmettre qu'un chromosome X à la suite de la dégénérescence des gamètes 0 (Figure 3), la descendance de cette phase sexuée est obligatoirement femelle.



**Figure 3.** Cycle monœcique du puceron du pois, *Acyrthosiphon pisum* (adapté de Jaquiéry et al. 2013).

## ***Symbioses bactériennes***

Les Hémiptères constituent le seul groupe d'animaux à utiliser la sève élaborée du phloème des plantes comme ressource alimentaire principale. Ce mode d'alimentation a évolué de multiples fois chez ces derniers, et est devenu prédominant dans le sous-ordre des Sternorrhyncha, comprenant les pucerons, les aleurodes, les psylles et les cochenilles, ainsi que chez les Auchenorrhyncha regroupant les cicadelles, les fulgores, et la plupart des Hétéroptères herbivores comme les pentatomidés (Dolling 1991). Ces insectes sont adaptés à une niche alimentaire spécifique, le phloème, qui est riche en sucres mais dans lequel certains acides aminés et vitamines sont en quantité insuffisante. Ainsi, seulement 8,2% des acides aminés présents dans la sève de la féverole, *Vicia faba*, sont essentiels aux pucerons alors que 72% correspondent à l'asparagine, un acide aminé non-essentiel (Douglas 2006).

Chez les pucerons, ce déficit en apport azoté est surmonté grâce à une symbiose obligatoire avec la bactérie *Buchnera aphidicola* (Buchner 1965). Le genre *Buchnera* forme un groupe monophylétique dont la phylogénie suit parfaitement celle de ses hôtes pucerons, ce qui démontre un co-spéciation à long terme entre les deux partenaires (Moran et al. 1993). *B. aphidicola* est une  $\gamma$ -proteobactérie intracellulaire qui est confinée dans le cytoplasme de cellules spécialisées de l'insecte, les bactériocytes. Ces symbiotes comptent pour une très large partie du microbiome des pucerons et peuvent représenter jusqu'à 10% de leur biomasse totale.

*Buchnera* est transmise verticalement à la descendance soit dans l'œuf d'hiver, soit aux embryons lors de la parthénogénèse vivipare. De nombreuses expériences ont montré que les pucerons dépendent de cette bactérie endosymbiotique pour la synthèse des neuf acides aminés essentiels (Douglas 1998). Au cours de l'évolution de cette relation symbiotique établie entre 160 et 180 millions d'années, le génome de *Buchnera* s'est considérablement réduit (Gil et al. 2002; Ham et al. 2003) jusqu'à atteindre une taille entre 500 et 600 kilobases. Ainsi, les gènes codant pour les enzymes impliquées dans les voies de biosynthèse des acides aminés essentiels ont été maintenus alors que ceux des voies de synthèse des non-essentiels ont été perdus. Une relation de mutualisme s'est donc créée entre les pucerons et leur symbiote obligatoire *Buchnera*, le puceron fournit un environnement de vie protégé et des acides aminés non-essentiels alors que *Buchnera* assure la synthèse d'acides aminés essentiels et de vitamines.

En plus du symbiote obligatoire *B. aphidicola* qui est présent chez tous les Aphididae, des symbiotes facultatifs sont hébergés par les pucerons. Ces symbiotes facultatifs peuvent influencer l'écologie et la physiologie de leurs hôtes de plusieurs manières (Oliver et al. 2010). Parmi les nouveaux traits écologiques que peuvent conférer ces symbiotes aux pucerons qui les hébergent, le plus documenté est certainement la protection contre les parasitoïdes, les pathogènes ou les prédateurs.

Ainsi, la  $\gamma$ -proteobactérie *Hamiltonella defensa* confère à son hôte puceron une protection contre les guêpes parasitoïdes. Cette protection contre le parasitisme a d'abord été montrée dans le système puceron du pois, *A. pisum* et son parasitoïde, *Aphidius ervi* (Oliver et al. 2003; Oliver et al. 2006). Chez *A. pisum*, la présence d'*H. defensa* a un effet bénéfique sur la *fitness* quand il y a une pression exercée par les parasitoïdes, mais représente un coût quand ceux-ci sont absents (Oliver et al. 2006). D'autres études ont également observé ce phénomène de protection chez le puceron noir de la fève, *Aphis fabae*, ainsi que chez le puceron noir de la luzerne, *Aphis craccivora* (Schmid et al. 2012; Asplen et al. 2014). Ce phénotype n'est pas induit par la seule présence d'*H. defensa*, il est influencé par l'infection de celle-ci par un phage lambdaïde lysogénique qui produit des toxines probablement impliquées dans la résistance (Moran et al. 2005; Oliver et al. 2009). La bactérie endosymbiotique *Regiella insecticola*, confère quant à elle une résistance au champignon entomopathogène *Pandora neoaphidis* (Ferrari et al. 2004; Scarborough et al. 2005). Les symbiotes facultatifs peuvent aussi influencer les interactions entre les pucerons et leurs prédateurs. Le symbiote *Rickettsiella* provoque un changement de couleur du rose au vert chez le puceron du pois, ce qui confère au puceron un camouflage plus efficace sur sa plante hôte et est supposé le protéger des prédateurs (Tsuchida et al. 2010).

Les symbiotes facultatifs peuvent également permettre une plus grande tolérance à la chaleur chez les pucerons. Ainsi en conditions de fortes températures, de meilleures performances (survie, temps de développement, fécondité) sont observées chez les pucerons du pois qui possèdent la bactérie *Serratia symbiotica* comparés à ceux qui ne la possèdent pas (Montllor et al. 2002; Russell and Moran 2006).

Enfin, peu d'études décrivent le rôle des symbiotes secondaires dans l'adaptation à la plante hôte. Le symbiote *R. insecticola* a été montré comme étant associé à de meilleures performances d'*A. pisum* sur le trèfle, *Trifolium sp.* (Tsuchida et al. 2004). Cependant, ce phénotype semble dépendre d'une association complexe entre la souche de *R. insecticola*, le génotype du puceron hôte et la plante. En effet, dans certains cas il est impossible de démontrer une quelconque association entre *R. insecticola* et l'utilisation du trèfle comme plante hôte ou dans d'autres cas, celle-ci est négative (Leonardo 2004; Ferrari et al. 2007; McLean et al. 2011). Le symbiote *Arsenophonus*, est quant à lui associé à de meilleures performances du puceron du soja *Aphis glycines* sur sa plante hôte (Wulff and White 2015), et également à l'exploitation du robinier faux-acacia par le puceron noir de la luzerne, *Aphis craccivora* (Wagner et al. 2015).

## 3.2 Dialogue moléculaire lors des interactions plantes-pucerons

### *Préférence et acceptation de l'hôte*

Avant d'ingérer de la sève du phloème – grâce à leurs pièces buccales spécialisées, notamment les stylets (Figure 4A) – les pucerons doivent sélectionner la plante hôte sur laquelle se nourrir. Une séquence de comportements est alors mise en place à longue et courte distances pour localiser celle-ci (résumé par Powell et al. 2006).

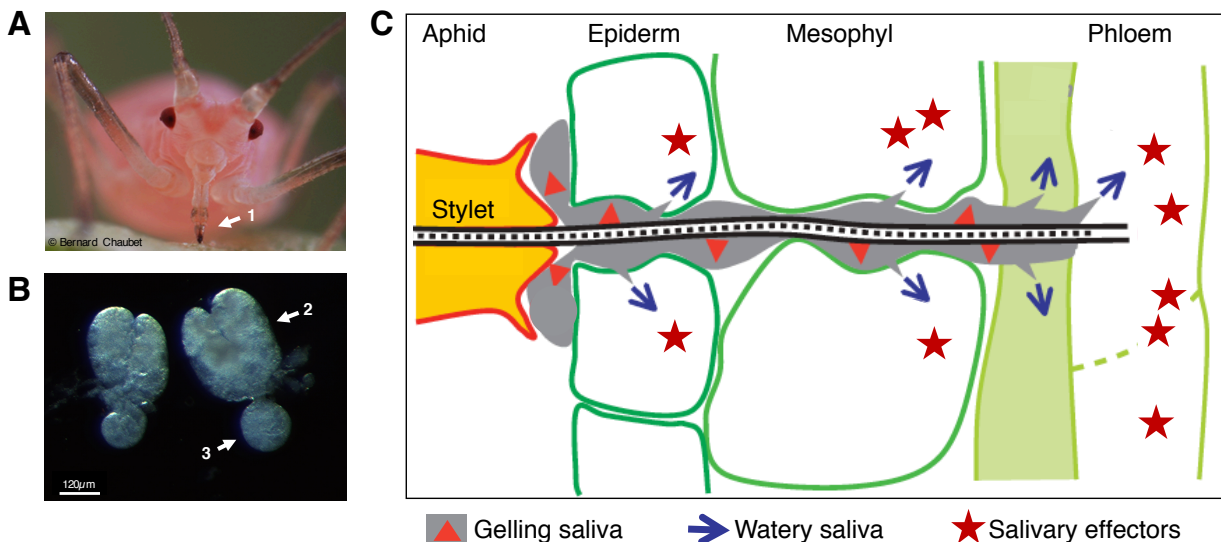
A longue distance, les pucerons vont d'abord s'orienter en fonction des longueurs d'ondes lumineuses reflétées par les plantes. Ces signaux visuels ne permettent probablement pas d'identifier spécifiquement la plante hôte, ainsi l'orientation est sûrement complétée par la perception de composés volatils émis par plantes, qui sont détectés grâce aux sensilles olfactives antennaires du puceron (Chung Park and Hardie 2003).

Une fois une plante atteinte, les pucerons vont apprécier, par des mouvements d'antennes, les caractéristiques physico-chimiques à la surface de la plante, telles que la texture, la couleur, les exsudats des trichomes ou encore les cires cuticulaires (Powell et al. 2006). Un comportement de sondage de l'épiderme de la plante par piqûres à la surface va néanmoins apparaître aussi bien dans les interactions compatibles qu'incompatibles. C'est pendant cette phase de sondage que les pucerons transmettent un grand nombre de phytovirus aux plantes hôtes et non-hôtes. La préférence pour la plante est essentiellement basée sur la phase de sondage

Ensuite le stylet, très flexible, empreinte un chemin majoritairement extracellulaire (Figure 4C). De brèves ponctions intracellulaires sont effectuées par le stylet tout au long de son trajet, ce qui permet au puceron de se diriger vers le phloème grâce à des gradients de pH et sucrose entre les différentes cellules végétales (Hewer et al. 2011). Cette phase de ponction dans les différentes couches végétales, qui précède la phase d'alimentation durable dans le phloème, semble jouer un rôle fondamental dans l'acceptation de la plante hôte (i.e. initiation de la production des descendants). Ainsi la reproduction est stimulée par les signaux chimiques des tissus périphériques avant la pénétration du stylet dans les tubes criblés du phloème (Powell et al. 2004). La chimioréception joue donc un rôle prépondérant dans les différentes phases de sélection de la plante hôte (Pickett et al. 1992). Les gènes codants pour ces chimiorécepteurs sont des déterminants majeurs de la reconnaissance des plantes hôtes et ont pu évoluer sous sélection positive suggérant leur rôle dans l'adaptation et la spécialisation des pucerons à leurs plantes hôtes (Smadja et al. 2009). Même si à ce stade le puceron a déjà accepté la plante comme un hôte, dans certains cas il ne peut établir de connexion durable avec le phloème. Ainsi, l'ingestion durable ou non de la sève par le puceron semble dépendante de la salivation qui est mise en place après la phase de sondage.

## Sécrétions salivaires des pucerons

Tout au long de son trajet dans les tissus végétaux, la salive gélifiante est sécrétée pour former autour du stylet une gaine protectrice – gaine sétale ou *sheath saliva* (Tjallingii and Esch 1993; Tjallingii 2006). Cette gaine produite par les glandes salivaires (Figure 4B), est riche en protéines, phospholipides et glucides. Elle limite le contact avec l’apoplasme et les parois cellulaires végétales et permet de protéger le stylet des composés de défenses de la plante (Hogenhout and Bos 2011). De plus, la protéine qui compose majoritairement cette gaine (SHP, *sheath protein*) est essentielle lors de l’alimentation du puceron. En effet, lorsque l’expression du gène *SHP* est réprimée par ARN interférence la formation de la gaine sétale est perturbée ce qui induit un arrêt de l’alimentation et une diminution de la fécondité des pucerons (Will and Vilcinskas 2015).



**Figure 4.** Alimentation et salivation des pucerons. (A) Puceron s’alimentant sur sa plante hôte, (B) glandes salivaires, (C) trajet du stylet dans les tissus végétaux et sécrétions des différents types de salive d’après Tjallingii 2006). 1, 2, et 3 indiquent respectivement le stylet, les glandes salivaires primaires et les glandes salivaires accessoires.

Un deuxième type de salive, la salive aqueuse, est sécrétée le long du trajet du stylet dans les cellules de l’épiderme et du mésophylle et également au niveau des tubes criblés du phloème une fois celui-ci atteint par le puceron (Figure 4C). L’ingestion durable de la sève par le puceron est alors dépendante de cette sécrétion de salive aqueuse qui contient des protéines capables de prévenir le phénomène d’occlusion du phloème. Plusieurs mécanismes peuvent alors être impliqués (Will et al. 2013). Lorsque le stylet pénètre dans le phloème, créant une blessure mécanique, un influx de calcium



déclenche l'occlusion des tubes criblés pour prévenir la perte de sève. Des complexes protéiques forment alors des bouchons, comme le forisome qui est une plaque protéique spécifique des Fabacées. Certaines protéines présentes dans la salive possédant des domaines de liaison au calcium, sont supposées bloquer l'occlusion calcium-dépendante (Will et al. 2007; Carolan et al. 2009). Ainsi, la salive de *Megoura viciae* permet le basculement d'un état relâché du forisome – bloquant l'influx de sève – à un état contracté assurant un flux continu de sève dont le puceron va s'alimenter (Will et al. 2007). Cependant cette observation, réalisée *in vitro* sur des forisomes de *Vicia faba*, reste difficile à démontrer *in vivo* (Medina-Ortega and Walker 2013). Les protéases identifiées dans la salive de différentes espèces de pucerons (Harmel et al. 2008; Carolan et al. 2009; Rao et al. 2013; Chaudhary et al. 2015) pourraient également participer à la dégradation des plaques protéiques qui occluent les tubes criblés. Les protéases de la salive de *M. euphorbiae* et d'*A. pisum* permettent ainsi de dégrader la principale protéine du phloème PP1 (*phloem protein 1*) de *Cucurbita maxima in vitro* (Furch et al. 2015). Chez cette cucurbitacée, PP1 a été identifiée comme pouvant participer à l'obstruction du phloème en se couplant avec la protéine PP2 – *phloem protein 2* (Furch et al. 2010).

Une fois le flux de sève assuré par la réversion de l'occlusion des tubes criblés du phloème, d'autres protéines contenues dans la salive aqueuse, peuvent permettre aux pucerons de moduler les autres défenses que la plante hôte met en place suite à l'infestation par les pucerons (Hogehout and Bos 2011; Rodriguez and Bos 2012).

### **Réactions de défenses des plantes**

En plus de l'occlusion du phloème, d'autres mécanismes de défense sont mis en place par les plantes pour lutter contre les pucerons (Jaouannet et al. 2014). Ainsi, plusieurs études sur le système *M. persicae*-*A. thaliana* ont permis d'explorer les voies de signalisation qui sont impliquées dans l'activation des défenses lors des interactions plantes-pucerons. Des gènes impliqués dans les voies de biosynthèse des phytohormones telles que l'acide salicylique (SA), l'éthylène (ET) et l'acide abscissique (ABA) sont induits chez *Arabidopsis* lorsque celle-ci est attaquée par *M. persicae* (Moran and Thompson 2001; De Vos et al. 2005; Kerchev et al. 2013). Les marqueurs de la voie de signalisation de l'acide jasmonique (JA) semblent en revanche réprimés (Kerchev et al. 2013). De plus, l'activation permanente de la voie JA augmente la résistance d'*Arabidopsis* à *M. persicae* (Ellis et al. 2002). En conséquence, De Vos et al. (2007) émettent l'hypothèse que la voie SA est manipulée et induite par les pucerons pour ainsi réprimer la voie JA, normalement mise en jeu lors d'attaques d'insectes herbivores et de pathogènes nécrotrophes. Bien qu'une manipulation du *crosstalk* JA-SA soit observée chez plusieurs insectes (Diezel et al. 2009; Chung et al. 2013), ce phénomène reste peu clair concernant les interactions plantes-pucerons car l'induction de la voie SA ne supprime pas

forcément la voie JA pour promouvoir les performances des pucerons (Kerchev et al. 2013). De plus, malgré des changements dans l'expression des gènes marqueurs, laissant penser à une activation des différentes voies de signalisation, il n'y a pas d'augmentation des niveaux phytohormones JA, SA, et ET lorsque *A. thaliana* est exposé à *M. persicae* (De Vos et al. 2005). Chez *A. pisum*, les clones adaptés à la luzerne *Medicago sativa* induisent de plus faibles niveaux de JA et SA que les clones non-adaptés à cette plante. Ceci est également observé pour certains autres métabolites de la voie JA comme l'OPDA (Sanchez-Arcos et al. 2016). La manipulation des voies de signalisation SA et JA permettrait aux clones d'*A. pisum* de se développer sur leur hôte. A la vue de ces différents résultats, des études complémentaires apparaissent nécessaires pour mieux comprendre la manipulation des voies signalisation de défenses des plantes par les pucerons. En effet, les pucerons n'induisant que très peu de dégâts physiques à la plante, comparés aux insectes broyeurs (e.g. larves de lépidoptères, coléoptères), l'implication de la voie JA dans le déclenchement des défenses dirigées spécifiquement contre les pucerons reste à approfondir.

Néanmoins, certains gènes ont été caractérisés comme impliqués dans les défenses mise en place par les plantes contre les pucerons. Ainsi, le gène *PAD4* qui encode une lipase contribue à la défense d'*Arabidopsis* contre *M. persicae*, les pucerons se reproduisant mieux sur les mutants *knockout* *PAD4* et moins sur les mutants sur-exprimant ce gène (Pegadaraju et al. 2007). Un cytochrome P450 encodé par le gène *PAD3*, a aussi été montré comme impliqué dans la défense contre *M. persicae*. Ce gène induit une augmentation d'un alcaloïde secondaire, la camalexine, qui réduit les performances du puceron sur *A. thaliana* (Kettles et al. 2013). Des composés comme les glucosinolates ou le DIMBOA peuvent également s'accumuler lors d'attaques par les pucerons (Kim and Jander 2007; Meihls et al. 2013), ainsi les mutants présentant de forts taux de glucosinolates sont plus résistants aux pucerons alors que ceux présentant de plus faibles taux sont sensibles (Kim and Jander 2007; J.H. Kim et al. 2008).

A ce jour, aucune molécule (HAMP) déclenchant les premières réactions de défenses des plantes (PTI) n'a été identifiée chez les pucerons contrairement aux lépidoptères (Bonaventure et al. 2011; Acevedo et al. 2015). En revanche, il a été montré que la salive de *M. persicae* ou encore la protéine GroEL produite par le symbiote *B. aphidicola* déclenchent une telle réponse chez *A. thaliana* (De Vos and Jander 2009; Chaudhary et al. 2014).

La deuxième phase de l'immunité des plantes – ETI – implique une relation gène-pour-gène entre effecteurs et gènes de résistance. Des gènes de résistance ont été identifiés dans plusieurs systèmes plantes-pucerons. Ainsi, la très grande variation observée dans la capacité qu'ont les biotypes du puceron du pois à infester différents génotypes de la plante *Medicago truncatula* est notamment dépendante du gène de résistance *RAP1* pour *Resistance to Acyrthosiphon pisum* 1 (Stewart et al. 2009;

Kanvil et al. 2014; Kanvil et al. 2015). L'émergence de biotypes virulents chez le puceron de blé *Diuraphis noxia* (Haley et al. 2004; Burd et al. 2006), le puceron de la laitue *Nasonovia ribisnigri* (ten Broeke et al. 2013) et le puceron du soja *A. glycines* (Kim et al. 2008) apparaît également dépendante de relations gène-pour-gène.

Plusieurs gènes de résistance aux pucerons, codants pour des récepteurs cytoplasmiques NB-LRR, ont été identifiés chez les plantes (Dogimont et al. 2010). Le gène *Mi-1* confère ainsi à la tomate une résistance contre le puceron *M. euphorbiae* (Rossi et al. 1998), mais également contre les nématodes, les aleurodes, et les psylles (Rossi et al. 1998; Nombela et al. 2003; Casteel et al. 2006). Le gène *Vat* confère au melon une résistance contre certains génotypes du puceron du cotonnier *Aphis gossypii* (Villada et al. 2009). Le gène *Ra* confère quant à lui une résistance contre le puceron des racines de la laitue *Pemphigus bursarius* (Wroblewski et al. 2007). D'autres résistances identifiées notamment chez le blé, le soja, l'orge le pêcher, ou encore *Medicago truncatula* pourraient être dépendantes de récepteurs NB-LRR (voir Dogimont et al. 2010). Dans tous les cas, les effecteurs de puceron qui sont reconnus par les NB-LRR ne sont pas encore caractérisés.

### ***Effecteurs salivaires des pucerons***

Les pucerons sont capables de manipuler les réponses de leur plante hôte de diverses manières. Ainsi, ils peuvent induire des changements morphologiques (e.g. galles) au niveau de la plante (Van Emden and Harrington 2007), d'impacter l'allocation des nutriments en leur faveur (Girousse et al. 2005) ou bien de supprimer les réaction de défenses des plantes (Will et al. 2007). Ces manipulations peuvent être liées à l'injection dans la plante hôte d'effecteurs comme cela est décrit chez les pathogènes de plantes (Hogenhout and Bos 2011). De récents progrès dans le séquençage des génomes de pucerons (The International Aphid Genomics Consortium 2010; Nicholson et al. 2015; Mathers et al. 2017; Wenger et al. 2017), le séquençage des transcriptomes de glandes salivaires (Ramsey et al. 2007; Bos et al. 2010; Carolan et al. 2011; Atamian et al. 2012) et l'analyse protéomique de salive récoltée sur milieu artificiel (Carolan et al. 2009; Nicholson et al. 2012; Rao et al. 2013; Chaudhary et al. 2015) ont permis d'identifier des effecteurs salivaires candidats chez plusieurs espèces.

Certains de ces effecteurs potentiels ont des fonctions connues comme les pectinases, glucanases ou amylases impliquées dans la dégradation de la paroi des cellules végétales. D'autres protéines potentiellement sécrétées ont des fonctions de détoxification comme les metalloprotéases, oxidoreductases, les peroxidases ou les phenol oxydases (Harmel et al. 2008; Carolan et al. 2009; Carolan et al. 2011; Nicholson et al. 2012; Rao et al. 2013). Cependant, la grande majorité des effecteurs candidats ne présentent aucune similarité avec des protéines/domaines connus et sont donc spécifiques des pucerons. Parmi le large répertoire d'effecteurs potentiels, seul un nombre limité a été

caractérisé pour leur rôle dans les interactions plantes-pucerons et leur rôle dans la manipulation des défenses de la plante reste encore peu décrit.

Les pucerons n'étant pas encore modifiables génétiquement, l'étude fonctionnelle des effecteurs salivaires dépend de techniques comme (i) la répression de l'expression (ARN interférence) par injection directe d'ARN double-brin dans le corps du puceron – qui induit une forte mortalité – ou (ii) la transformation des plantes permettant l'expression/répression transitoire ou stable des effecteurs *in planta* – qui nécessite que les plantes soient transformables. Les difficultés de mise en place de ces techniques peuvent expliquer le faible nombre d'effecteurs salivaires caractérisés à ce jour.

Certains de ces effecteurs impactent négativement la fécondité des pucerons lorsqu'ils sont surexprimés dans les plantes. Ceci peut refléter une activation des défenses de la plante ou résulter d'une expression trop importante de l'effecteur dans la plante hôte qui ne reflète pas la réalité du système biologique. Ainsi, les effecteurs Mp10 et Mp42 réduisent la fécondité de *M. persicae* et déclenchent les réactions de défense lorsqu'ils sont exprimés dans *Nicotiana benthamiana* (Bos et al. 2010; Rodriguez et al. 2013). De plus, Mp10 induit une chlorose de la plante, mais il est par ailleurs caractérisé pour son rôle dans la suppression du pic oxydatif (PTI) (Bos et al. 2010) déclenché par l'éliciteur bactérien flag22.

D'autres effecteurs semblent requis lors de l'alimentation et contribuent ainsi à la *fitness* des pucerons. Le premier effecteur salivaire identifié chez les pucerons est le gène *C002* du puceron du pois. Ce gène, qui ne présente aucune fonction connue, est retrouvé uniquement chez les pucerons et est l'un des gènes les plus exprimés dans les glandes salivaires. Quand l'expression du gène *C002* est réprimée par injection de double-brin ARN dans les pucerons, ceux-ci ont des difficultés à atteindre le phloème ce qui conduit à une plus grande mortalité (Mutti et al. 2006; Mutti et al. 2008). Le gène homologue de *M. persicae* *MpC002* contribue aux performances de ce puceron sur ses hôtes *Nicotiana* et *Arabidopsis* (Bos et al. 2010).

Le fait que le *M. persicae* soit capable de se développer sur les plantes modèles, pour lesquelles les techniques d'étude fonctionnelle des effecteurs sont disponibles, a facilité la caractérisation des effecteurs de ce puceron. Ainsi, plusieurs gènes salivaires tels que *Mp1*, *Mp2*, *Mp55*, et *MpMif1* contribuent à la *fitness* de *M. persicae* (Pitino et al. 2011; Pitino and Hogenhout 2012; Elzinga et al. 2014; Naessens et al. 2015). La manière dont ces gènes interfèrent avec le système immunitaire des plantes reste peu claire. Cependant, le gène *Mp55* est directement impliqué dans la suppression des défenses en diminuant le dépôt de callose, le pic oxydatif et l'accumulation de glucosinolates quand il est exprimé chez *Arabidopsis*. L'expression de *Mif1* dans les tissus foliaires inhibe l'expression de certains gènes des voies de défenses, les dépôts de callose et la réponse cellulaire hypersensible (HR) chez *Nicotiana* (Naessens et al. 2015). Enfin, la protéine Mp1 cible une protéine impliquée dans le

trafic vésiculaire (VPS52) chez *Arabidopsis*, celle-ci pouvant être impliquée dans l'immunité des plantes (Rodriguez et al. 2017; Yun and Kwon 2017).

Chez *M. euphorbiae*, trois effecteurs salivaires Me10, Me23 et Me47 semblent avoir un rôle dans l'interaction de ce puceron avec ses hôtes Solanacées (Atamian et al. 2012; Kettles and Kaloshian 2016). Me47 présente *in vitro* une activité glutathione-S-transferase (GST) spécifique des isothiocyanates, composés impliqués dans la défense contre les herbivores. Kettles and Kaloshian (2016) suggèrent alors que les GST, qui sont impliquées dans la résistance des pucerons aux insecticides, pourraient également supprimer les défenses lors des interactions plantes-pucerons.

En plus du gène *C002* précédemment décrit, le gène *Armet* code pour un effecteur modulant l'interaction entre *A. pisum* et sa plante hôte (Wang et al. 2015a). Deux autres gènes, membres d'une famille multigénique codant pour des métallopeptidases M2, les Angiotensin-converting-enzymes (ACE), modulent ensemble l'interaction *A. pisum-V. faba* (Wang et al. 2015b). Cependant, leur potentiel rôle dans la suppression des défenses de la plante reste à éclaircir.

Les différentes études ayant caractérisé des effecteurs salivaires chez diverses espèces de pucerons montrent que ceux-ci ont des effets variés, aussi bien sur la *fitness* du puceron que sur les défenses des plantes. Cependant, ces travaux révèlent quelques caractéristiques importantes de ces gènes : (i) certains présentent une expression spécifique des glandes salivaires (Guo et al. 2014; Pan et al. 2015; Wang et al. 2015a; Wang et al. 2015b), (ii) certains sont pucerons-spécifiques comme *Mp1*, *C002*, et *Mp55*, et d'autres conservés au sein des insectes/animaux comme les *ACE*, *Armet*, *Me47*, *Mif1* (Pitino et al. 2011; Pitino and Hogenhout 2012; Naessens et al. 2015; Wang et al. 2015a; Wang et al. 2015b; Kettles and Kaloshian 2016); et finalement (iii) certains augmentent les performances des pucerons et ciblent des protéines de l'hôte de manière espèce-spécifique (plantes et pucerons) et évoluent sous sélection positive suggérant un rôle de ces effecteurs dans la spécialisation à la plante hôte (Pitino and Hogenhout 2012; Rodriguez et al. 2017).

#### 4. Modèle d'étude : *Acyrtosiphon pisum*, le puceron du pois

Le puceron du pois, *Acyrtosiphon pisum*, est un Aphididae spécialiste des plantes de la famille des légumineuses (Fabacées). Il est décrit sur une cinquantaine d'espèces de cette famille botanique, certaines ayant un intérêt agronomique (Eastop 1971). La diminution des rendements engendrée par le prélèvement de sève et la transmission de différents virus de plantes (notamment Lutéovirus, Cucumovirus, et Potyvirus) font du puceron du pois un ravageur de cultures majeur en milieu

tempérés (Ighil et al. 2011). Il se reproduit par parthénogénèse cyclique (PC) mais aussi par parthénogénèse obligatoire (Jaquiéry et al. 2014) et est considéré comme monœcique (Figure 3). Comme décrit précédemment, le puceron du pois peut héberger différents symbiotes facultatifs (Oliver et al. 2010, voire partie I.3.1) qui participent à différents traits comme par exemple la protection contre les parasitoïdes, prédateurs et pathogènes. Néanmoins, le rôle que pourraient jouer ces symbiotes dans l'adaptation à la plante hôte chez cette espèce reste peu clair (Leonardo 2004; Tsuchida et al. 2004; Ferrari et al. 2007). En 2010, le génome du puceron du pois a été séquencé, constituant ainsi le premier génome de puceron disponible (The International Aphid Genomics Consortium 2010). Le génome d'*A. pisum* (Acyr\_2.0, GCA\_000142985.2) est assemblé en 23 925 *scaffolds* pour une taille de 512 Mégabases, et est comme d'autres génomes d'insectes, riche en AT (environ 30% de GC). Actuellement, 20 923 gènes sont prédits au NCBI dont 18 601 codent pour des protéines (27 971 protéines, certaines étant issues de transcrits alternatifs du même gène), c'est sur cette annotation que sont basés ces travaux de thèse (NCBI *Acyrthosiphon pisum* Annotation Release 102).

### **Complexe de biotypes**

Différents travaux ont identifié chez *A. pisum* plusieurs populations sympatriques spécialisées sur différentes légumineuses, d'intérêt agronomique ou sauvages, et formant un complexe de biotypes ou races d'hôtes (Via 1991; Simon et al. 2003; Ferrari et al. 2006; Frantz et al. 2006).

Grâce au génotypage de 14 marqueurs microsatellites sur des individus récoltés en France et Allemagne, Peccoud et al. (2009a) ont montré l'existence de 11 populations sympatriques (biotypes) spécialisées sur différentes plantes. Ajoutant 3 populations inféodées au lotier corniculé *Lotus corniculatus*, aux mélilots officinal *Melilotus officinalis* et blanc *Melilotus albus* et à la luzerne lupuline *Medicago lupulina* à celle précédemment décrites. Parmi ces populations, certaines échangent encore du matériel génétique alors que d'autres sont considérées comme des espèces cryptiques, en isolement reproducteur complet, car aucun hybride n'a été détecté. C'est le cas des populations inféodées à *O. spinosa*, *C. scoparius* et *L. pratense*. Dans ce continuum de divergence, la proportion d'hybrides est négativement corrélée à la différenciation génétique entre les biotypes, illustrant des stades plus ou moins avancés du processus de spéciation (Peccoud et al. 2009a).

A ce jour au moins 15 biotypes sont décrits dans le complexe de puceron du pois (Peccoud et al. 2015). De plus, l'étude de l'évolution moléculaire du symbiote obligatoire *B. aphidicola* a permis de dater le dernier ancêtre maternel commun à tous les biotypes d'Europe entre 8000 et 16 000 ans avec un pic de diversification du complexe situé entre 3600 et 9500 ans. Cette radiation adaptative – la plus rapide jamais décrite – coïncide avec le réchauffement climatique qui a eu lieu sur Terre après le Pléistocène et avec la domestication par l'Homme de plusieurs des légumineuses hôtes du puceron

du pois (Peccoud et al. 2009b). A cette période, l'expansion de l'aire de répartition des légumineuses augmentant ainsi leur disponibilité et abondance pour les pucerons a pu créer des opportunités écologiques favorisant la formation de biotypes (Peccoud et al. 2009b).

Le nombre de biotypes spécialisés sur diverses plantes hôtes au sein des Fabacées, et le continuum de divergence entre des populations sympatriques et des espèces cryptiques font du puceron du pois un excellent modèle dans l'étude de la divergence adaptative et de la spéciation égologique (Peccoud and Simon 2010). De plus, le séquençage du génome d'*A. pisum* en 2010 (The International Aphid Genomics Consortium 2010) a permis l'initiation d'études de génétique des populations à large échelle visant à reconstruire l'histoire évolutive de ce complexe et d'identifier les bases de l'adaptation à la plante hôte au sein des biotypes.

### ***Fonctions candidates à l'adaptation***

Comme décrit précédemment, deux phases sont déterminantes dans les interactions plantes-pucerons : la sélection de l'hôte et son exploitation au travers de l'alimentation.

Les récepteurs olfactifs et gustatifs – faisant partie du système chimiosensoriel impliqué dans la perception et l'acceptation de l'hôte – représentent donc de bons candidats impliqués dans l'adaptation du puceron du pois à ses hôtes Fabacées. Une expansion rapide des familles multigéniques codant pour ces récepteurs, ainsi qu'une évolution récente sous sélection positive de certains membres des deux familles suggèrent un rôle important de ces gènes dans l'adaptation à la plante hôte chez le puceron du pois (Smadja et al. 2009). Des études de génétique/génomique des populations, sans *a priori* par *genome scan* (Jaquiéry et al. 2012; Nouhaud et al. 2014) ou par capture de gènes ou *targeted resequencing* (Smadja et al. 2012; Duvaux et al. 2015; Eyres et al. 2017), ont également révélé que les gènes codant ces récepteurs olfactifs et gustatifs présentaient des signatures de sélection suggérant ainsi leur implication dans l'adaptation à la plante au sein du complexe de biotypes (i.e. spécialisation).

Bien que les systèmes olfactifs et gustatifs aient un rôle dans le choix de la plante hôte (Powell et al. 2006), les biotypes d'*A. pisum* essaient de s'alimenter sur des plantes pour lesquelles ils ne sont pas spécialisés (Schwarzkopf et al. 2013), montrant ainsi l'importance des sécrétions salivaires dans l'exploitation durable de la plante hôte. Ainsi, les effecteurs salivaires, qui permettent le succès des pucerons grâce à la manipulation de la physiologie et/ou des défenses de l'hôte (Hogenhout and Bos 2011), représentent également de bons candidats à l'adaptation du puceron du pois à ses hôtes Fabacées. Les plantes hôtes pouvant différer dans leurs réactions de défenses, les effecteurs salivaires ont probablement évolué en fonction des différentes pressions exercées par leurs hôtes. Avant mon travail de thèse, un répertoire des effecteurs salivaires du puceron du pois a été créé par une double approche transcriptomique et protéomique révélant 324 gènes exprimés dans les glandes salivaires et

codant pour des protéines potentiellement secrétées dans la plante hôte (Carolan et al. 2011). La sécrétion de neuf de ces protéines candidates a été confirmée par l'analyse protéomique de salive récoltée sur milieu artificiel (Carolan et al. 2009). Certains de ces gènes salivaires ont été montrés comme étant indispensables au succès d'*A. pisum* sur sa plante hôte *V. faba*, jouant alors un rôle d'effecteur dans l'interaction du puceron du pois avec son hôte (Mutti et al. 2006; Mutti et al. 2008; Naessens et al. 2015; Wang et al. 2015a; Wang et al. 2015b). L'étude par balayage génomique (*genome scan*) de Jaquiéry et al. (2012) a révélé que 3 des 11 régions génomiques sous sélection divergente au sein du complexe de biotypes étaient situées proches de gènes salivaires, suggérant également le rôle de ces effecteurs salivaires dans la spécialisation des biotypes.

En plus de la sécrétion d'effecteurs, l'exploitation de la plante hôte par les pucerons passe par la digestion de la sève ingérée. Les cathepsines B, qui sont des protéases à cystéine exprimées dans l'intestin des pucerons, présentent une large expansion chez le puceron du pois (Rispe et al. 2007). De plus l'évolution sous sélection positive de certains gènes spécifiques d'*A. pisum* suggère leur importance dans l'adaptation à l'hôte (Rispe et al. 2007). L'implication des cathepsines B et autres enzymes de détoxification dans l'adaptation à la plante hôte au sein du complexe de biotypes du puceron du pois restent néanmoins à explorer.

## 5. Objectifs des travaux de thèse

Comme illustré précédemment, les pucerons sont d'importants ravageurs de cultures à l'échelle de la planète. Un grand nombre d'espèces de pucerons ayant à ce jour développé des résistances aux insecticides communément utilisés en agriculture (Li and Han 2004; Toda et al. 2004; Bass et al. 2014; Foster et al. 2014), un des principaux enjeux du futur va résider dans le développement de méthodes de lutte durables contre ces ravageurs. La compréhension des mécanismes moléculaires qui gouvernent l'interaction avec la plante hôte et donc l'adaptation des pucerons à des gammes d'hôtes plus ou moins large paraît alors essentielle pour développer une lutte plus durable, en créant par exemple des plantes résistantes aux pucerons en remplacement des insecticides.

Parmi les fonctions candidates à l'adaptation des pucerons à leur hôte, les effecteurs salivaires – étant secrétés – peuvent interagir de manière directe avec la plante hôte. Cette interaction directe entre les effecteurs des pucerons et la plante hôte a facilité la mise en place d'outils d'analyse fonctionnelle permettant d'étudier la fonction des effecteurs, notamment par la manipulation des plantes modèles (Pitino et al. 2011). En revanche, ce n'est pas le cas pour les gènes du système



chimiosensoriel, qui peuvent être impliqués dans des interactions indirectes avec la plante, nécessitant probablement la manipulation des pucerons plutôt que de la plante. L'analyse fonctionnelle des effecteurs salivaires montre qu'ils peuvent moduler les défenses de la plante et pour certains, être essentiels au succès des pucerons sur leur hôte (Mutti et al. 2006; Mutti et al. 2008; Pitino et al. 2011; Atamian et al. 2012; Pitino and Hogenhout 2012; Elzinga et al. 2014; Naessens et al. 2015; Wang, Dai, et al. 2015). De plus, certains effecteurs ont été caractérisés pour leur action plante-hôte spécifique (Pitino and Hogenhout 2012). Nous posons donc l'hypothèse que les gènes codants pour ces effecteurs ont été exposés à des pressions de sélection induites par les plantes au cours de l'évolution des pucerons et qu'ils ont joué un rôle clé dans l'adaptation du puceron de pois à ses hôtes Fabacées.

Mes travaux de thèse portent sur l'étude du rôle des effecteurs salivaires dans l'adaptation du puceron du pois à ses hôtes. Concernant l'adaptation d'*A. pisum* aux légumineuses, il convient de distinguer deux événements majeurs ayant eu lieu à des échelles de temps très différentes.

Ainsi comme le genre *Acyrtosiphon* comprend des espèces se nourrissant sur des familles de plantes variées (Eastop 1971), nous supposons que la spécialisation sur les légumineuses a dans un premier temps eu lieu au moment de la divergence d'*A. pisum* avec des espèces proches non-inféodées aux légumineuses il y a plusieurs millions d'années (Jaquiery et al. 2017 en révision).

Ensuite plus récemment (16 000 à 3000 ans), des populations d'*A. pisum* se sont spécialisées sur une ou quelques espèces au sein de la famille des légumineuses, formant ainsi le complexe de biotypes (Peccoud et al. 2009a; Peccoud et al. 2009b).

A l'origine, ces travaux de thèse visaient à explorer les deux échelles évolutives afin de caractériser les effecteurs salivaires candidats à l'adaptation du puceron du pois dans un contexte de spécialisation de l'espèce aux légumineuses (échelle macro-évolutive) et de spécialisation des biotypes à leurs hôtes associées (échelle micro-évolutive). En raison des contraintes de temps inhérentes à la réalisation d'une thèse en 3 ans, seule l'échelle macro-évolutive a été abordée par ces travaux. Cependant, les futures pistes d'exploration de l'échelle micro-évolutive apportées par ce travail seront exposées dans la discussion.

Mon travail a consisté dans un premier temps à créer un « catalogue » plus complet des effecteurs salivaires du puceron du pois, basé sur des approches « omiques » et un traitement bioinformatique (chapitre II).

Dans un deuxième temps, les caractéristiques évolutives des effecteurs candidats ont été analysées par des approches de génomique comparative au sein des Aphididae (chapitre II).

Enfin, l'étude fonctionnelle *in vivo* de certains effecteurs candidats a été initiée dans le système *A. pisum*-plante hôte (chapitre III).

## II. Paper 1

### FAST EVOLUTION AND LINEAGE-SPECIFIC GENE FAMILY EXPANSIONS OF PUTATIVE APHID EFFECTORS SHED LIGHT ON MOLECULAR EVOLUTION DRIVEN BY APHID-PLANT INTERACTIONS

Hélène Boulain<sup>1</sup>, Fabrice Legeai<sup>1,3</sup>, Endrick Guy<sup>1</sup>, Stéphanie Morlière<sup>1</sup>, Nadine Douglas<sup>2</sup>, Michael Smith<sup>4</sup>, Jonghee Oh<sup>5</sup>, Marimuthu Murugan<sup>6</sup>, John Reese<sup>4</sup>, Frank White<sup>7</sup>, James Carolan<sup>2</sup>, Jean-Christophe Simon<sup>1#</sup> and Akiko Sugio<sup>1#</sup>

#### AFFILIATIONS

<sup>1</sup> INRA, UMR 1349 IGEPP, Le Rheu, France

<sup>2</sup> Department of Biology, Maynooth University, Maynooth, Co. Kildare, Ireland

<sup>3</sup> Inria/IRISA GenScale, Campus de Beaulieu, Rennes, France

<sup>4</sup> Department of Entomology, Kansas State University, Manhattan, KS USA

<sup>5</sup> Department of Plant Pathology, Kansas State University, Manhattan, KS USA

<sup>6</sup> Department of Plant Molecular Biology and Biotechnology, Centre for Plant Molecular Biology, Tamil Nadu Agricultural University; Coimbatore; Tamil Nadu, India

<sup>7</sup> Department of Plant Pathology, University of Florida, Gainesville, FL USA

# Equal contribution

#### ABSTRACT

Effector molecules are playing crucial roles in plant-pathogen interactions by suppressing plant defenses, allowing parasites to invade host and hijack plant nutrients for growth. More recently, effector proteins have been characterized in insect herbivores but their contribution to the evolution of plant specialization in insects is largely ignored. Aphids feed on phloem sap and inject into the host plant saliva proteins which function as effectors. Gaining insight into the evolutionary patterns of aphid effectors is an important step toward unveiling the mechanisms of host plant specialization. The pea aphid, *Acyrtosiphon pisum*, is specialized to legumes and has experienced multiple events of

specialization resulting in a large complex of biotypes. Here, we first built a catalogue of putative effectors in the pea aphid. A transcriptomic approach identified 3603 salivary glands (SG) expressed effectors in the pea aphid genome, and tissue comparison between SG and gut revealed that 740 of them are up-regulated in SG. Comparison of both SG-expressed and SG-up-regulated effectors revealed that SG-up-regulated effectors are enriched in Aphididae genes. Moreover, single copy SG-up-regulated effectors evolve in average faster than SG-expressed effectors. *A. pisum* salivary gene family expansion, allowing diversifying selection on certain gene copies, appears as an important mechanism to generate novelties in *A. pisum* effectors. By creating a gene set of putative effectors and using comparative genomics on newly acquired aphid genomes, our study represents the more comprehensive evolutionary analysis to date of aphid salivary effectors and provides new candidates for functional characterization and further identification of effectors involved in plant specialization in the pea aphid complex.

## INTRODUCTION

Insects compose the largest group of metazoans on the planet and evidences indicate that the evolution of herbivory has played a fundamental role in species diversification (Mitter et al. 1988; Mayhew 2007; Wiens et al. 2015). Half of the currently known insect species feed on plants (Wu and Baldwin 2010), and herbivorous insect groups exhibit faster rates of diversification compared to non-herbivorous (Mitter et al. 1988, Wiens et al. 2015). A major hypothesis accounting for higher species richness in herbivorous insects relies on an evolutionary arms race between defense mechanisms of plants (e.g. structural traits and chemicals) and plant exploitation strategies of insects (Fraenkel 1959; Ehrlich and Raven 1964; Van Valen 1977; Janz 2011). Such arms races would drive the evolution of herbivores towards specialization which in turn would initiate reproductive isolation. This could happen during the course of reciprocal antagonistic interactions between plants and insects (Wheat et al. 2007; Edger et al. 2015), or when insects colonize new plants through host shifts (Forbes et al. 2017; Vertacnik and Linnen 2017). During the continuous process of specialization, genetically distinct populations that are adapted to different host plants (e.g. host races) may evolve as a result of divergent selection and potentially could form distinct species through the establishment of reproductive barriers (Drès and Mallet 2002; Peccoud et al. 2009a; Powell et al. 2013; Soria-Carrasco et al. 2014). Therefore, unveiling the evolutionary and mechanistic bases of plant specialization in insect herbivores provides opportunities to understand the processes that have led to the great diversification of metazoan life.

Aphids form a group of about 5000 species (Blackman and Eastop 1994; Blackman and Eastop 2000), most of them being considered as plant specialists, and generalists are rare (Peccoud et al. 2010). They have developed a particular feeding style relying on a dedicated mouthpart, the stylets, which navigate between plant cells to reach the phloem. Aphids are devastating pests of wild and cultivated plants as they directly reduce plant nutrients by ingesting phloem sap and indirectly cause diseases by transmitting plant pathogens (Blackman and Eastop 2000; Harris and Maramorosch 2014). Functional studies have highlighted the role of salivary proteins as aphid effectors (virulence/avirulence factors) secreted into the host plant and with the potential to interfere with plant biology (Hogenhout and Bos 2011; Elzinga and Jander 2013; Rodriguez et al. 2017). Since such effector proteins are thought to interact with plant defense systems, we hypothesize that they have been exposed to plant-induced pressures over the course of aphid evolution and have played a key role in the adaptation of aphids to their host plants.

Indeed, expression of certain salivary effectors as *C002*, *Mp1* and *Mp2* from the green peach aphid, *Myzus persicae*, were shown to promote its fecundity on *Arabidopsis thaliana* or *Nicotiana benthamiana*, which are the host plants of *M. persicae*, whereas expression of *A. pisum* homologues had no effects (Pitino and Hogenhout 2012). The positive effect of salivary effectors in a species-specific manner indicates a specialization of effectors during host plant adaptation of aphids. In addition, host specificity was also demonstrated in the interaction of *Mp1* with plant target (Rodriguez et al. 2017). As other important characteristics, the aphid effectors that contribute to maximum aphid performance may be up-regulated in salivary glands (SG) compared to other organs (Pan et al. 2015; Wang et al. 2015a; Wang et al. 2015b) and some show signatures of positive selection (Pitino and Hogenhout 2012; Thorpe et al. 2016). Although some works studied evolutionary patterns of several salivary genes at a time, a global and comprehensive evolutionary analysis of salivary genes has not been reported to the best of our knowledge.

The first large catalogue of aphid putative salivary effectors was created on *Acyrtosiphon pisum*, the first species with full genome sequence (The International Aphid Genomics Consortium 2010), using a dual transcriptomic and proteomic approach and constituted by secreted 324 proteins encoded by SG expressed genes and the proteins detected in salivary gland proteomics (Carolan et al. 2011). The catalogue was created by assembling two sets of SG ESTs, and SG enriched transcripts were identified by comparing with other organs' ESTs. Some effectors predicted in this catalogue are characterized and shown to be involved in plant-aphid interactions (Mutti et al. 2006; Mutti et al. 2008; Wang et al. 2015a; Wang et al. 2015b), but there is a possibility that several salivary effector genes were missed in this catalogue. Considering the significant development in RNA-seq technology and mass-spectrometry (MS), re-creation of an *A. pisum* salivary gene catalogue using these technologies

offers an opportunity to create a comprehensive salivary gene catalogue and to identify previously undescribed candidates. Therefore, in this study, we first aimed to improve the catalogue of *A. pisum* salivary genes and to define a set of candidate salivary effector genes for further evolutionary analyses.

The pea aphid, *Acyrtosiphon pisum* (Hemiptera, Aphididae, Macrosiphini), is specialized on legumes (Fabaceae) and can feed on more than 20 genera of this plant family (Peccoud and Simon 2010). This aphid actually forms a large complex of plant-adapted populations (also referred as biotypes or host races), each specialized to one or very few legume species. This complex of biotypes represents a continuum of divergence, from host races to presumable species (Peccoud et al. 2009a; Peccoud et al. 2014). As aphid species from the *Acyrtosiphon* genus are associated to various plant families like Fabaceae, Rosaceae, Euphorbiaceae or Poaceae (Eastop 1971), *A. pisum* has probably adapted to legume family several millions years ago, before the diversification of the biotype complex has diversified 5,000-10,000 years ago, coinciding with the onset of plant domestication for agriculture (Peccoud et al. 2009b) In addition, population genetic analyses on multiple biotypes allowed characterizing genomic regions linked with adaptive divergence induced by plant specialization, and highlighting salivary and chemosensory systems as candidate functions responsible for host adaptation in the pea aphid complex (Jaquiéry et al. 2012; Smadja et al. 2012; Duvaux et al. 2015; Eyres et al. 2017).

For a while, *A. pisum* has remained the only aphid to have its genome fully sequenced (The International Aphid Genomics Consortium 2010). The recent genome sequencing of two specialist aphids, the Russian wheat aphid, *Diuraphis noxia* (Nicholson et al. 2015), and the soybean aphid, *Aphis glycines* (Wenger et al. 2017), in addition to the generalist green peach aphid *M. persicae* (Mathers et al. 2017) enables comparative genomic approaches to explore the evolution of salivary effector genes among aphids with very different host range and to identify those genes that could have allowed pea aphid adaptation to legumes. Therefore, by using recently completed insect genome sequences, we examined the evolutionary features of the candidate salivary effector genes to test whether they could show specific patterns associated with their potential involvement in the molecular arms race between aphids and plants. Finally, we restricted our evolutionary analysis to the unigenes that have been functionally characterized in earlier studies and the large gene families that hold many salivary genes as their members. From these analyses, we revealed that the pea aphid contains a large repertoire of putative salivary effectors, which could serve as a source of novelties in the arms race with plant defenses. We also showed that some of these salivary effector genes are fast-evolving or part of lineage-specific gene families, suggesting their particular involvement in pea aphid specialization on legumes and representing valuable candidates for functional analyses.

## RESULTS

### Creation and evaluation of the *A. pisum* salivary effector candidate gene sets

We aimed to improve salivary gene catalogues and conducted RNA-seq of dissected SGs and guts to highlight effector candidates. Based on previous works assessing the tissue specificity of salivary effector expression (Pan et al. 2015; Wang et al. 2015a; Wang et al. 2015b), we hypothesized that a large majority of salivary effector genes are up-regulated in *A. pisum* SG tissue. However, some characterized salivary effector genes are not specifically up-regulated in SG compared to other tissues (Naessens et al. 2015). Hence, we created two sets of candidate salivary genes (SG-expressed and SG-up-regulated gene sets) and examined them further. We choose gut as an organ to compare the gene expression level of salivary gene candidates because it is easy to dissect and to obtain high amount of RNAs.

Our analysis of the transcriptome of *A. pisum* LSR1 salivary glands identified 12,040 genes expressed in this organ (FPKM cutoff 0.5) out of the 18,601 protein coding genes constituting the *A. pisum* gene set used in this study (fig. 1a). A protein secretion prediction pipeline was applied to the global gene set; thus, we selected the genes encoding the proteins with a signal peptide predicted by SignalP v3 or v4.1 and the proteins with non-canonical secretion signals predicted by SecretomeP. Then, the genes encoding transmembrane domains (predicted by TMHMM 2.0) or GPI-anchors (predicted with PredGPI) were removed from the previous secreted protein set since they are likely embedded in a membrane and not secreted. Among the 12,040 SG expressed genes, 3603 genes encoded proteins predicted to be secreted, constituting our SG-expressed effector candidate gene set (fig. 1a). The SG RNA libraries were also analyzed in comparison with gut libraries to identify up-regulated effector candidate genes in SG compared to gut. After filtering (CPM cutoff 1) and normalization steps, 12,378 protein coding genes were retained for the differential expression (DE) analysis. Among these genes, the linear model revealed that 1989 genes were up-regulated in SG tissues. Out of which, 740 genes were predicted to encode secreted proteins and designated as *A. pisum* SG-up-regulated effector candidate genes (fig. 1b). Out of the 740 SG-up-regulated effector candidates, 739 genes were found in the 3603 SG-expressed effector gene set, and one gene (LOC100574271) was missing as its FPKM was lower than 0.5. The comparison with the previous salivary gland secretome of *A. pisum* (Carolan et al. 2011), which identified 324 candidate effectors, placed 95 of them in the SG-up-regulated effector gene set, 93 in the SG-expressed effector gene set, 111 were SG-expressed genes encoding non-secreted proteins and 9 were genes not expressed in our study. All the effector candidate genes identified in both sets, their characteristics and the secretion prediction process are listed in supplementary table S1, Supplementary Material online.

In the second step, we conducted saliva and SG proteomics studies to evaluate the two effector candidate gene sets created by transcriptomic analysis. Direct detection of proteins in saliva is considered as a confirmation of secretion of salivary effector candidates, while SG proteomics confirms translation of effector candidate genes in SGs. Fifty-one proteins were supported by more than one peptide detected by MS in at least two out of the three replicates or by one peptide in the three replicates of saliva samples (supplementary table S2, Supplementary Material online). Of these, 35 proteins (68.62%) were included in the SG-up-regulated effector candidate set, two proteins belonged to the SG-expressed effector set, and 14 proteins were not predicted as secreted proteins and were encoded by genes up-regulated in SG tissue (11 genes) or undifferentiated genes (3 genes). Five out of the 11 SG up-regulated genes seemed to be incorrectly annotated (possible truncation of N-terminal part or missing nucleotides in scaffold) and predicted to be not secreted. The dissected SG samples detected 1837 *A. pisum* proteins (supported by more than one peptide in at least two of the three replicates or by one peptide in the three replicates) encoded by 1809 unique genes (supplementary table S3, Supplementary Material online). This analysis confirmed the presence of 205 proteins from the SG-up-regulated effector candidate set and of 447 proteins from the SG-expressed effector candidate set. Next, we analyzed the link between signal intensity of MS analysis and SG gene expression level of the genes detected by two proteomics experiments (supplementary fig. S1, Supplementary Material online). As for the genes encoding proteins detected in saliva, the correlation between gene expression level and signal intensity was not significant; however, a significant positive correlation was identified between expression level and MS signal intensity of the SG tissue ( $r=0.5446$ ,  $p < 0.001$ ).

### **Orthology analysis of pea aphid gene content revealed enrichment of Aphididae specific genes in SG-up-regulated effector candidate set**

In order to assess the orthology levels of the genes constituting the SG-up-regulated and SG-expressed effector candidate sets, a comparative genomic analysis was conducted between *A. pisum* and 16 other full genome sequenced arthropods. Based on the phylogeny of these 17 species, six levels of orthology were determined. Thus, for each *A. pisum*-gene-containing orthologue group, Arthropoda, Insecta, Hemiptera, Aphidomorpha or Aphididae levels were assigned in function of the presence-absence of at least one orthologue in the other parts of the phylogeny. Then, *A. pisum* specific category was imputed to the *A. pisum* genes that were not grouped with other species' gene (fig. 2a). Comparison of the two salivary effector sets with the *A. pisum* whole gene set revealed differences in level of conservation (fig. 2b). Relative to the global gene set, the SG-expressed effector gene set was significantly enriched in highly conserved Arthropod and Insecta genes. The Aphidomorpha level was

also slightly enriched, while the proportion of *A. pisum* specific genes was significantly reduced. In contrast, in the SG-up-regulated effector set, the proportion of genes belonging to the Aphididae category (twice the number of genes compared to whole gene set) were significantly increased whereas highly conserved genes within the Arthropod category were reduced (fig. 2b).

### **Evolutionary rates vary among of *A. pisum* salivary candidate effectors**

The presence of specific gene functions and Aphididae specific genes in the SG-up-regulated effector candidate set indicates an enrichment of aphid specific effectors in this gene set. Since parasite effectors are under the evolutionary pressure induced by plant defense systems, it is commonly believed that effectors evolve rapidly compared to the rest of genes (Jiang et al. 2008; Remenant et al. 2010).

To examine the hypothesis that SG-up-regulated gene set differs in rate of evolution, we estimated the ratio of nonsynonymous to synonymous substitutions, dN/dS or  $\omega$  (Yang 2007). To simplify the analysis, we examined only the single copy gene pairs of orthologues between *A. pisum* and the three Aphididae species included in the orthology search (*A. pisum/A. glycines*, *A. pisum/D. noxia*, *A. pisum/M. persicae*). For each pairwise comparison, the genes were classified into three categories: SG-up-regulated effector candidate set, SG-expressed but not up-regulated effector candidate set and the rest of gene set (fig. 3). In every case, the distribution of evolutionary rates of the SG-up-regulated effector candidate genes was significantly different from the two other gene sets: SG-up-regulated effector candidates tend to have higher dN/dS. The SG-up-regulated candidate effectors with dN/dS above 95<sup>th</sup> percentile of each pairwise comparison (genes having dN/dS in the 5% highest values) were then considered as fast-evolving candidate salivary effectors and listed in supplementary table S1, Supplementary Material online.

We then examined the evolutionary patterns of the pea aphid orthologues of the 11 single copy aphid salivary effector genes that were previously functionally characterized and shown to be involved in plant-aphid interactions (table 1). All these single copy salivary effectors or their *A. pisum* orthologues were found in the SG-up-regulated effector candidate gene set with the exception of *Mif1* gene (Naessens et al. 2015), which was not predicted as secreted in our study (table 1). Most of them were highly expressed and up-regulated in SG compared to gut. The genes *C002*, *Mp1*, *Ap25*, *Me23*, *Me10*, *Mp55*, *Shp* and *Mp2* were Aphididae or Aphidomorpha specific genes as no orthologues were identified in other insect species and most of them (except *Mp55*) were considered as fast-evolving genes (top 5% of dN/dS values obtained from pairwise comparisons). In contrast, *Mif1*, *Armet*, and



*Mp10* appeared more conserved as their dN/dS ratios never exceeded 0.12. To deeply examine evolutionary rates of these characterized salivary effectors, we searched for other aphid species sequences in the public databases (supplementary table S4, Supplementary Material online), and estimated the dN/dS under different site models of selection. The site-model M0 (one ratio), providing a global dN/dS value for each gene, revealed a range from 0.06298 (*Armet*) to 0.64653 (*Mp1*) (table 1; supplementary table S5, Supplementary Material online). This analysis revealed clear differences between the two categories of salivary genes; insect or Arthropod conserved effector genes like *Armet*, *Mif1* and *Mp10* evolve more likely under purifying selection (with lowest dN/dS), and aphid specific effector genes evolve faster as their dN/dS values are above 0.40. Although some genes appeared to evolve under different selective regimes, the examination of null and alternative models of codon substitutions of M1 versus M2 and M7 versus M8 (Nielsen and Yang 1998; Yang et al. 2000; Yang and Bielawski 2000) revealed signature of positive selection only in *Mp1* and *Me10* effector genes (supplementary table S5, Supplementary Material online). In *Mp1* and *Me10*, a Bayes empirical Bayes (BEB) approach (Yang et al. 2005) detected sites under positive selection mainly in the region coding mature protein injected in the host plant (supplementary fig. S2, Supplementary Material online). The positively selected sites in *Mp1* gene were consistent with those previously identified by Pitino and Hogenhout (2012), and four sites with dN/dS above 2.5 were supported by BEB probability greater than 0.75 in the *Me10* gene.

These evolutionary analyses of 11 previously functionally characterized salivary genes revealed that there are two types of salivary genes: genes that are widely conserved in Arthropods or Insect orders and under purifying selection, and genes that are Aphididae or Aphidomorpha specific and under faster evolution.

### **Episodic diversifying selection occurred in *A. pisum* expanded salivary gene families**

Evolutionary novelties are usually brought by gene duplication followed by diversification in functions (reviewed by Zhang 2003; Conant and Wolfe 2008; and Kondrashov 2012). Strikingly, one third of the SG-up-regulated effector candidate set was composed of duplicated genes that could be grouped into 125 families (each family contains at least two homologous genes). Here, we focused on the three peculiar multigene families that were strongly supported by detection of some members in the saliva proteomics (with at least one specific peptide discriminating the different copies). Indeed, a cystein-rich protein family (CRP), the Angiotensin-Converting-Enzyme-like family (ACE) and the Aminopeptidase-N (apN) family represented 46.9% of the proteins detected in *A. pisum* saliva (supplementary table S6, Supplementary Material online); thus, they were relevant effector candidates.

The *ACE* and *apN* gene families belonged to the metallopeptidase family (GO:0008237) involved in proteolysis (GO:0006508), two functions that were enriched in the SG-up-regulated effector candidate set of the pea aphid (supplementary fig. S3, Supplementary Material online).

### The Cysteine-Rich Protein (CRP) family

The *CRP* Aphididae specific gene family was previously described by Guo et al. (2014) as a family of 14 proteins encoded by 12 genes in *A. pisum*. By thorough investigation of the orthologue groups, we refined this analysis and identified 15 members encoding small proteins of less than 200 amino acids with highly conserved 14 cysteine residues, in the *A. pisum* genome. This family was clearly expanded in *A. pisum* (15 copies) compared to the other aphid species: *D. noxia*, *A. glycines* and *M. persicae* presented 6, 4 and 1 copies, respectively, of the CRP family members in their genomes. Among these 15 *A. pisum* gene copies, four (*CRP-3*, *CRP-7*, *CRP-9*, and *CRP-10*) were not expressed in either SG or gut tissues. Eleven genes were up-regulated in SG with nine (*CRP-1*, *CRP-2*, *CRP-4*, *CRP-5*, *CRP-6*, *CRP-8*, *CRP-11*, *CRP-12*, *CRP-13*, *CRP-14* and *CRP-15*) of them encoding proteins predicted to be secreted and included in the SG-up-regulated effector set (supplementary table S6, Supplementary Material online). The pea aphid CRP-12 and 14 proteins were found in our saliva proteomic analysis, which is the first study detecting the *CRP* gene family products in aphid saliva. We searched for signatures of diversifying selection among the different branches along the Aphididae CRP family tree (fig. 4a). The branch-site (BS) model of codon substitutions (Yang and Nielsen 2002) was used to compare each branch (foreground) against the others (background) within the Aphididae species. The alternative model allowing a proportion of sites under positive selection in the foreground branch was compared to the null model where dN/dS was fixed to 1 (neutral evolution on foreground branch). After multiple comparison adjustments of p-values provided by the Likelihood Ratio Test (LRT), traces of episodic diversifying selection were detected on five branches in the CRP tree (under neutral or purifying selection in background branches). Thus, signatures of positive selection were found in one copy of *D. noxia* CRP (*CRP-5*) and on four branches after the duplication event occurred in *A. pisum* genome (fig. 4a). Thus ancestral branches #2 and #4 and the *A. pisum* *CRP-2* experienced positive selection whereas other branches in the tree evolved under purifying or neutral selection. The proportions of sites within each class of the BS model and their average dN/dS values are reported in fig. 4b. The BEB posterior probability of each site to be under positive selection among the 5 retained branches is provided in the supplementary fig. S4a, Supplementary Material online, and the CRP codon-based alignment of the analyzed sites (supplementary fig. S4b, Supplementary Material online) indicated the different amino acids encoded by the sites with BEB posterior probability above 0.75. Due to the absence of known domains within these CRP proteins, we were unable to check whether

these sites are involved in the functions of the proteins. Although positive selection occurred in some *A. pisum* *CRP* copies, no positively selected sites were detected in the *CRP-13*, the most highly expressed *A. pisum* *CRP*. However, this gene seems to be involved in *A. pisum*-plant interactions as the aphids express more *CRP-13* when feeding on plants compared to an artificial diet (Guo et al. 2014).

### The Angiotensin-Converting-Enzyme-like (*ACE*) gene family

In contrast to *CRP* gene family which is Aphididae specific, the *ACE* gene family was conserved within arthropods and genes were found in most insect genomes used in our orthology analysis (fig. 5 and supplementary table S6, Supplementary Material online). Wang et al. (2015b) identified three *ACE* in the *A. pisum* genome and characterized the *ACE1* and *ACE2* genes as modulators of plant-*A. pisum* interaction as they are required to feed on plant but not on artificial diet. Searching in created orthologue groups yielded 10 *ACE* copies in *A. pisum* genome (supplementary table S6, Supplementary Material online) distributed in four clades with their orthologues from other insect species (fig. 5). Seven of these genes were up-regulated in *A. pisum* SG tissue including three genes encoding predicted secreted ACE that were listed in the SG-up-regulated effector candidate set (*ACE1*, *ACE2* and *ACE5*). The result that ACE10 protein, in addition to ACE1 and 5 were detected by saliva proteomics (supplementary table S6, Supplementary Material online) suggests a likely problem of the secretion prediction caused by incorrect annotation of *ACE10*. The clade 4 showed a clear expansion of *A. pisum* *ACE* (4 copies) compared to other aphid species having only one copy (fig. 6a). All these clade 4 ACE proteins seemed to be functional peptidases, containing a M2 peptidase domain (IPR001548) with presence of the active site HEXXH (fig. 6a), apart from ACE10 which is probably partially annotated (in the edge of the scaffold). The BS model of codon substitution analysis was performed on the four ACE clades separately between the four closely related Aphididae species and the alternative model proved to be significant for the clade 4 only. Three branches along the Aphididae ACE clade 4 were significant beyond the correction of LRT p-values (fig. 6b and c), revealing episodic diversifying selection in those branches. Thus, the ancestor branch (#1) of *A. pisum* *ACE 1* and *ACE5*, the *A. pisum* *ACE1* gene and the ancestral branch #3 presented sites under positive selection whereas these sites evolved under purifying or neutral selection in other branches of the phylogenetic tree. The BEB posterior probability (above 0.75) revealed the position of sites under positive selection in each significant branch (supplementary fig. S5a, Supplementary Material online) and the codon-based alignment indicated that the involved amino acid residues were scattered across the protein (supplementary fig. S5b, Supplementary Material online).

### The Aminopeptidase-N (*apN*) gene family

The proteins derived from the Aminopeptidase-N (*apN*) gene family was the most represented in *A. pisum* saliva proteomics. Indeed, 18 *apN* proteins were detected in our study accounting for 36.7% of the 49 proteins detected in saliva, and one more *apN* protein was identified by Carolan et al. (2009) (supplementary table S6, Supplementary Material online). Forty-seven *apN* genes were identified in the pea aphid genome, 27 of them belonged to orthologous groups conserved among insect species and 20 were *A. pisum* specific and too divergent to be clustered with any orthologues (supplementary table S6, Supplementary Material online). All of these 20 *A. pisum* specific *apN* genes were up-regulated in SG tissue. Half of them were SG-up-regulated candidate effectors and four were detected in saliva proteomics. The other 10 *A. pisum* specific *apN* proteins were not included in effector candidate set as they were not predicted as secreted. However, four out of the 10 were detected in saliva proteomics. Therefore, there is a possibility that some of these genes were not properly annotated (e.g. missing 5' end), leading to their classification as non-secreted.

The 27 *apN* genes with orthologues in other aphid and insect species were clustered in eight distinct clades (fig. 7). The pea aphid genes from clades 5, 6 and 8 were gut up-regulated or undifferentiated genes, whereas those from clades 1, 2, 3, 4, and 7 were mainly SG up-regulated genes (few copies were undifferentiated between analyzed tissues). Apart from *A. pisum* gene LOC100169117 (encoding the protein XP\_016656878.1) which belongs to the clade 1, the other SG-up-regulated candidate effector genes were all found in the *apN* clade 4 (fig. 7).

In the *apN* clade 4, orthologous genes were identified in all studied insect species apart from the Hymenoptera *Nasonia vitripennis* and *Apis mellifera* and the soybean aphid *A. glycines*. This *apN* clade was remarkably expanded in *A. pisum* genome (14 gene copies) compared to all other insect species and especially to *D. noxia* (3 gene copies) and *M. persicae* (single copy) aphids. All of the 11 pea aphid gene copies that were SG up-regulated encode proteins that were detected in saliva proteomics. However, two of them were not included in the SG-up-regulated effector candidate set, as they are not predicted as secreted (fig. 7). The genes classified as Aphididae clade 4 *apN* encode proteins constituted by a signal peptide followed by a M1 peptidase domain (IPR014782) and an ERAP1-C domain (IPR024571); however some of them had lost the region encoding one of the two domains, and the large majority (apart from *D. noxia apN-3*, *A. pisum apN-1* and *A. pisum apN-5*) possessed the HEXXH active site ensuring the peptidase function (fig. 8a). The multiple duplication events observed in the pea aphid might have increased the chance to accumulate mutations conferring selective advantage in the environment. To test this hypothesis, the BS model of codon substitution was used to examine episodic diversifying selection in specific lineage along the Aphididae phylogenetic tree of *apN* clade 4. Traces of positive selection were detected in seven branches of the

tree (fig. 8b and c). Sites were under positive selection in *D. noxia* gene *apN-1* (branch #1), the ancestral branches #3, #5, #7 and *A. pisum* *apN-3*, *apN-8* and *apN-12* genes whereas they were under purifying or neutral selection in other branches. The BEB posterior probability for each sites to be under positive selection was computed for these seven branches and revealed that significant sites (BEB > 0.75) were mainly located in the M1 peptidase and ERAP1-C domains and in the unknown part between signal peptide sequence and M1 peptidase domain for the branches #1, #2, and #3 (supplementary fig. S6, Supplementary Material online). The amino acids encoded by these sites were highlighted in the codon-based alignment used to compute dN/dS in supplementary fig. S7, Supplementary Material online.

## DISCUSSION

Our study identified 3603 genes that are expressed in salivary glands and potentially secreted candidate effector genes. Out of them, 740 genes are up-regulated in salivary glands, representing the SG-up-regulated candidate effector gene set. Comparative genomics with other aphid and insect genomes revealed that SG-up-regulated candidate effector set is enriched in Aphididae conserved genes, whereas SG-expressed set is enriched in genes with higher degree of conservation among insects/arthropods. Single copy genes that are classified as SG-up-regulated candidate effectors tend to evolve faster than the rest of the genes classified as SG-expressed candidate effectors. However, functional analyses of previously described effectors highlighted that both fast and slow evolving effector genes are involved in plant-aphid interactions. Three pea aphid expanded gene families were identified in SG-up-regulated effector set. In all three families, multiple branches containing SG-up-regulated effector candidates were shown to be under diversifying selection; thus, gene duplication and diversification appear to contribute to form the battery of *A. pisum* salivary proteins required for host specialization.

### Improvement of salivary candidate effector gene sets

Our study based on differential expression analysis identified 740 SG-up-regulated candidate effector genes in the *A. pisum* genome. The tissue specificity of gene expression assessed by comparing SGs and guts was fundamental to define a restricted list of candidates among the 3603 SG-expressed candidate effector genes. Moreover, SG-up-regulated candidate effectors show different evolutionary patterns (e.g. faster evolution) than SG-expressed candidates, which is a convincing evidence to focus

primarily on this set when we look for the genes that are involved in the interaction with specific host plant. The newly identified SG-up-regulated candidate effector gene set represents an outstanding increase compared to the previous catalogue of 324 *A. pisum* effector genes by Carolan et al. (2011). This difference was mainly due to the use of RNA seq approach that represented an important improvement of gene expression detection and quantification compared to EST sequencing previously used. Further sequencing of RNAs from other tissues in addition to the SGs and guts is however needed to confirm tissue-specificity of SG-up-regulated candidate effector genes. Even if several of studied aphid effectors are up-regulated in SG tissues (Atamian et al. 2012; Guo et al. 2014; Pan et al. 2015; Wang et al. 2015a; Wang et al. 2015b) we cannot completely ignore the salivary genes that are not up-regulated in SG. As we see an example with *Mif1*, some important salivary effectors may not be encoded by genes that are up-regulated in SG. (Naessens et al. 2015). In addition, we have to consider that some effector proteins could be produced in elsewhere than salivary glands. Indeed, it has been suggested that salivary glands function as a combined excretory and secretory organ in aphids, as they do not have Malpighian tubules, which normally ensures this role in most of the insect species (Miles 1972). Therefore, it has been proposed that proteins can move from the hemocoel of aphids to salivary glands to be then secreted through saliva (Chaudhary et al. 2014). However, we have no such evidence from literature, except for the chaperonin GroEL produced by the obligate endosymbiont *Buchnera aphidicola* (which lives in specialized organs, the bacteriocytes), and injected by its aphid host the potato aphid *Macrosiphum euphorbiae* into the plant via saliva (Chaudhary et al. 2014; Chaudhary et al. 2015).

Although proteomics dataset was size-limited compared to transcriptomics dataset, it allowed confirming the secretion of 35 SG-up-regulated candidate effectors. Moreover, our hypothesis of SG up-regulation of *A. pisum* effectors is strongly supported by the proteomic analysis as 46 proteins (90%) out of the 51 detected in saliva were encoded by SG up-regulated genes, the remaining 10% being encoded by non-differentially expressed genes. That we did not find a link between proteins secreted in saliva and their gene expression in SG suggests the existence of regulation processes to control which effectors are secreted by the pea aphid. Plant cues (reviewed in Powell et al. 2006) may be essential to trigger aphid responses as we did not detect some of the highly expressed salivary genes in our saliva proteomics. One of these is C002, which is known to be secreted *in planta* and required for aphid feeding (Mutti et al. 2008). The comparison of both transcriptomic and proteomic approaches also revealed some limits in the secretion pipeline. Indeed, 14 proteins that we detected in the proteomic study of *A. pisum* saliva were not predicted to be secreted. Five of them may be miss-annotated but the rest of proteins might have been secreted by pathways which we did not consider for the secretion prediction

We revealed marked differences between SG-up-regulated and SG-expressed effector gene sets both at functional and evolutionary levels. Indeed, compared to SG-expressed effector gene set, the SG-up-regulated candidate effector gene set was enriched in proteins with proteolysis and metallopeptidase activities. Metallopeptidase M1 (Aminopeptidase N) and M2 (Angiotensin-converting-enzyme) represent nearly half of saliva detected proteins suggesting a key role in *A. pisum*-plant interactions. The comparison of gene conservation levels revealed a clear contrast between the SG-up-regulated gene set where 50% of the candidate effectors are strictly found in Aphidomorpha, Aphididae or *A. pisum* and the SG-expressed gene set where 64.2% of the genes are highly conserved candidate effectors. To conclude this part, our work shows that SG-up-regulated candidate effector gene set presents evolutionary properties that are probably linked to the specialization of SG tissue during establishment of aphid-feeding style and suggests that these effector candidates are more likely involved in legume adaptation in *A. pisum* than those which are only SG expressed.

### **Faster evolutionary rates in candidate effectors gene set**

The global analysis of pairwise dN/dS ratio revealed average higher evolutionary rates of the single copy genes classified as SG-up-regulated effectors. In comparison, SG-expressed effector genes tend to be exposed to a stronger purifying selection. This apparent globally faster evolution of SG-up-regulated effectors suggests their involvement in aphid specialization to particular host plants. However, analysis of the single copy genes previously characterized for their role in plant-aphid interactions showed heterogeneous selective regimes. We hypothesize that the more conserved salivary effectors (e.g. *Mif1*, *Armet* and *Mp10*) could be involved in fundamental functions required for plant feeding (e.g. prevention of phloem clogging, repression/manipulation of general plant defenses, Will et al. 2009; Bos et al. 2010; Furch et al. 2015; Naessens et al. 2015), explaining why diversifying mutations are removed to conserve fundamental role of the effectors. By contrast, eight of the characterized aphid effectors present evolutionary rates above 0.40, much higher than the conserved effectors with dN/dS not exceeding 0.12. Seven of these genes were among the fastest-evolving genes in pairwise analysis between *A. pisum* and the three other fully-genome-sequenced aphid species. As they evolve more rapidly and have been characterized as modulators of plant-aphid interactions in functional studies, these effectors are likely involved in host-plant adaptation of aphids. Indeed, some of them (*C002*, *Mp1* and *Mp2*) are known to act in a species-specific manner to promote aphid performances on host plant (Pitino and Hogenhout 2012). Moreover, positively selected sites detected in *Mp1* (Pitino and Hogenhout 2012; this study) and *Me10* salivary effectors suggest that these sites may be involved in the arms race with plant defenses.

## Large salivary gene family expansions in *A. pisum*

In addition to fast-evolving single copy genes, three multigenic families, which are either aphid-lineage specific (*CRP*) or conserved in insect species (*ACE* and *apN*), were examined. Analysis revealed that SG-up-regulated effector candidates were often created by multiple gene duplication events that occurred after the split of *A. pisum* from other aphids, and that episodic diversifying selection drove their evolution. So, gene duplications appear as a fuel for evolutionary novelties in salivary effectors of the pea aphid, which coincide with the evolutionary patterns of other parasites' effectors (Stergiopoulos et al. 2012; Oliva et al. 2015; Zhao et al. 2015). This potential source of novel effectors may have been recruited during biotype formation of the pea aphid complex, allowing a finer specialization on specific legume host species of each biotype. Further works, aiming to identify biotype-specific polymorphisms and/or biotype-specific expression patterns across these effector families, and functional analysis of effectors in different biotype-host systems would help to characterize their role in biotypes specialization.

Effector genes are often aggregated in large clusters at several genomic loci of plant pathogens (e.g. CRN and RXLR effectors of the oomycete *Phytophthora infestans*) and are thought to result from non-allelic homologous recombination and tandem duplications (facilitated in repeat-rich genomic regions), generally associated with rapid birth and death evolution (Win et al. 2007; Jiang et al. 2008; Haas et al. 2009). This was also observed in herbivore insects such as the gall midge *M. destructor* that shows massive expansion of effectors in genome clusters thought to provide a reservoir for diversification of new functions (Zhao et al. 2015). Although a few members of each pea aphid salivary family *CRP*, *ACE* and *apN*, are clustered on the same scaffolds in the pea aphid genome (supplementary table S6, Supplementary Material online), we were not able to identify clear gene clusters resulting from tandem duplication events due to the highly fragmented assembly of the pea aphid genome (The International Aphid Genomics Consortium 2010).

The functions of these three gene families in insects and especially in aphids are largely unknown. The Aphididae specific *CRP* gene family was first described by Guo et al. (2014) who demonstrated in particular that *A. pisum CRP-13* is induced when aphids feed on plant in comparison with artificial diet. Other cysteine-rich proteins were characterized for their antifungal activities in different insect species (Banzet et al. 2002) or for their role in inhibition of muscle contraction as in snake venom (Yamazaki and Morita 2004). Interestingly, some small secreted cysteine-rich proteins were characterized as effectors in the interaction between the Asian soybean rust fungus, *Phakopsora pachyrhizi*, and its host plant (Qi et al. 2016). One particular effector was shown to suppress plant immunity by interacting with a soybean transcription factor essential for negative regulation of immunity. Thus, the various cysteine-rich proteins appear to be active in different systems including



plant-pathogen interactions, so this aphid-specific CRP family may play a key role for aphid success once injected into the host plant. The ACE proteins are known to be involved in many different processes such as gut and heart muscle contraction, reproduction and blood pressure homeostasis in mammals (Harrison and Acharya 2014). In all cases, their functions rely on their peptide processing activity ensured by the M2 peptidase domain. The *A. pisum* *ACE1* and *ACE2* genes, were previously reported to contribute to aphid full performance on plant (Wang et al. 2015b), suggesting their possible involvement in cleavage of certain plant defense proteins during aphid feeding. Aminopeptidases have been found in aphid guts and reported as involved in digestion (Rahbé et al. 1995; Cristofolletti et al. 2003) or in virus binding (Linz et al. 2015) but little is known about their function in aphid saliva. However, Furch et al. (2015) demonstrated that *A. pisum* salivary metalloproteases are active *in vitro* and able to degrade major phloem protein 1 (PP1), which is involved in the protein deposition on sieve plates after severe metabolic disturbance (Gaupels et al. 2008), suggesting that these proteases may help to prevent sieve-element occlusion triggered by aphid feeding. Despite the diversifying selection acting on several *A. pisum* M1 and M2 peptidase genes, the active site HEXXH is conserved in most of the proteins indicating a functional cleaving activity. Nevertheless, some copies may have lost active sites or domains, particularly in the *A. pisum* ap-N family members that were too divergent to be clustered, reflecting possible neofunctionalization after high diversification.

Comparative analysis involving the four aphid genomes showed that the three gene families were significantly expanded in an *A. pisum* specific manner. Indeed, whether they are aphid lineage-specific as *CRP* or conserved among insects as *ACE* and *apN*, *A. pisum* shows the highest number of copies in each family with 15, 10 and 47 genes respectively. Gene family expansion was previously described in *A. pisum*, particularly concerning cathepsin B gut proteases (Rispe et al. 2007), chemoreceptors (Smadja et al. 2009), small RNAs machinery (Jaubert-Possamai et al. 2010; Lu et al. 2011), amino acid transporters (Price et al. 2011) and cuticular proteins (The International Aphid Genomics Consortium 2010). In some cases, fast evolution and positive selection occurred on recent duplicated genes (Rispe et al. 2007; Smadja et al. 2009). In line with previous works, the expansion of salivary gene families reported here in *A. pisum* and subsequent diversifying selection strongly highlight the importance of gene duplication as a mechanism of genomic adaptation to changing environment in the pea aphid; it also suggests an involvement of these three families during the specialization of the pea aphid on legumes. Nevertheless, we noticed the presence of Aphididae specific duplications in smaller number in the *D. noxia* genome for *CRP* and *apN* family, as well as in *A. glycines* for the *CRP*. Interestingly, *A. glycines* gene copy was absent from the clade 4 of *apN* family which contains a large majority of *A. pisum* salivary genes. This could result from a gene loss during the course of specialization to soybean. Further analyses are necessary to test whether the loss

happened in an ancestor of the Aphidini subtribe or only in the soybean aphid. Positive selection also acted on specific duplicated copies in *D. noxia* but a tissue specific expression analysis or saliva proteomics is required to know whether they are SG specialized potential effectors.

Within the Macrosiphini subtribe the most notable contrast concerning salivary gene families is certainly between *A. pisum* and *M. persicae*, as we found only one copy of each *CRP* and *apN* (clade 4) family in the genome of this generalist aphid. Recent genome sequencing of *M. persicae*, which is able to colonize many plant species across 40 plant families (Blackman and Eastop 2000), highlights a lower rate of gene expansions in its genome compared to *A. pisum* (Mathers et al. 2017). Although less species-specific duplicated genes were found in *M. persicae*, some clusters of aphid duplicated genes (e.g. genes resulting from before-speciation duplications) were differentially regulated within two days after host switch experiments. Moreover, knock-down of the differentially expressed genes reduced aphid fitness only on the plants that induced over-expression. (Mathers et al. 2017). Thus, the genuine generalist lifestyle of *M. persicae* is thought to be enabled by a rapid transcriptional plasticity of duplicated genes (Mathers et al. 2017). By contrast, very little changes in gene expression have been found in *A. pisum* biotypes when aphids were shifted from the universal host (*Vicia faba*) to the specific host of a given biotype (Eyres et al. 2016). Hence, we suggest that species-specific expansion of salivary gene families, among others, is one of the key innovations enabling the specialization of the pea aphid to legumes. These two mechanisms of gene duplication and transcriptional plasticity appear to drive evolution of aphids. However, further investigations are needed to test whether specialism in aphids is more governed by copy number variation in salivary effector genes and generalism by transcriptional plasticity.

## CONCLUSION

By creating a gene set of putative effectors and applying comparative genomics to newly acquired aphid genomes, our study represents the more comprehensive evolutionary analysis to date of aphid salivary effectors. We revealed that the pea aphid possesses a large repertoire of salivary effectors, which constitutes a source of novelties to cope with antagonistic interactions exerted by plant defenses. Some of these effectors seem to be conserved across aphid species and may be involved in fundamental functions required for plant feeding. Others are fast-evolving or members of lineage-specific gene families and may contribute to specialization of pea aphids to legumes. To gain more insights into the involvement of effectors in plant specialization, more functional studies are needed and our work provides a selection of valuable salivary candidate genes for such analyses. By

combining comparative genomics at different evolutionary levels and functional approaches on aphid effectors and their plant targets, we hope our understanding on how effector diversity and evolution shaped plant adaptation in aphids would be largely improved.

## **MATERIALS AND METHODS**

### ***Aphids, Plants and Growth Conditions***

*A. pisum* lineage LSR1 was maintained in a growth chamber at 18°C with a 16 h day/8 h night photoperiod on the broad bean, *Vicia faba* (Castel), at low density to avoid the production of winged individuals. *V. faba* was grown in a growth chamber at 18°C with a 16 h day/8 h night photoperiod for 10 days before installation of the aphids.

### ***RNA sequencing***

To prepare RNA samples from aphid salivary glands and guts, nine-day-old *A. pisum* LSR1 reared on *V. faba* at the density of 10-15 aphids/plant were rapidly dissected with fine forceps in saline solution. The dissected organs were soaked in RNA later (QIAGEN) immediately after the dissection to avoid RNA degradation. On average 200 pairs of salivary glands and 20 guts (the maximum numbers of organs we could dissect in one day) were pooled for one replicate of RNA extraction by NucleoSpin RNA XS (Macherey-Nagel). Four replicates of each sample were prepared.

rRNA depletion, stranded-RNA library preparation, multiplexing, and sequencing were performed by Genewiz (New Jersey, USA). Sequencing was performed on the Illumina HiSeq2500 platform, in a 2x125bp paired-end (PE) configuration in High Output mode (V4 chemistry). Each sample was divided in four and run in four different lanes to avoid lane effect. In total, 471,933,074 reads were obtained for eight samples.

### ***Mapping and differential expression analysis***

Gene expression of *A. pisum* salivary glands and guts was analyzed using the Acyr\_2.0 (GCF\_000142985.2) reference genome assembly and the NCBI *Acyrtosiphon pisum* Annotation Release 102, both available at <ftp://ftp.ncbi.nlm.nih.gov/genomes>. The paired-end libraries were mapped using STAR v2.5.2 (Dobin et al. 2013) with the following parameters: `outFilterMultimapNmax=5`, `outFilterMismatchNmax=3`, `alignIntronMin=10`, `alignIntronMax=50000`, `alignMatesGapMax=50000`. Fragment counts per genes were estimated by Subread featureCounts (Liao et al. 2014) using default parameters. These raw fragment counts were converted to counts per

million (CPM) based on Law et al. (2016) using the edgeR (Robinson et al. 2010) and limma (Ritchie et al. 2015) R-implemented packages (R core Team 2017). Expressed genes were filtered by removing genes with a CPM<1 in at least three libraries on the eight analyzed libraries and CPMs have been normalized by using edgeR TMM method for Normalization Factor calculation (Robinson and Oshlack 2010). Heteroscedasticity from count data was removed using precision weights calculated by the voom function (Law et al. 2014). Differentially expressed genes analyses have been performed by fitting linear models (Smyth 2004; McCarthy and Smyth 2009) to compare salivary gland vs gut tissues. Graphics supporting the validation of the described differential expression analysis are provided in supplementary fig. S8, Supplementary Material online. In addition, normalized fragments per kilobase million (FPKM) were calculated with edgeR on the four salivary glands samples to obtain the whole transcriptome and rank the genes by level of expression in this organ.

### ***Saliva and salivary gland proteomics***

#### Saliva collection

LSR1 aphids of mixed age were reared on *V. faba* and approximately 2000 aphids were installed on 12 plastic rings, each contains 5 ml of artificial diet in between two stretched parafilms set at the top of a ring (radius 4.5 cm, height 5 cm). The aphids were contained within the ring and the bottom of the ring was sealed by a stretched parafilm. The diet was collected every day and replaced by new diet. In the same time, new aphids were added from *V. faba* and old aphids at the bottom of the ring were removed. The diet was collected three successive days (in total about 4000 aphids were used) and kept at -80°C. The diet collected in three days (approximately 150 ml) were pooled and the proteins were concentrated by VIVACELL 250 with VIVACELL ultrafiltration concentrator membrane insert MWCO 5kD (Sartorius) and by VIVASPIN 6 MWCO 5 kD (Sartorius) to 300 µl. Proteins from diet concentrates were purified further using a 2D Clean-up Kit (GE HealthCare) following the manufacturer's instructions. The resulting protein pellet was resuspended in 6M urea, 2M thiourea, 0.1 M Tris-HCl, pH 8.0 and quantified using the Qubit™ protein quantification system (Invitrogen), following the manufacturer's instructions.

#### Salivary glands

Adult LSR1 aphids (14 to 16-days old) were dissected in ice-cold saline and salivary glands were immediately transferred to PBS (phosphate-buffered saline) containing Roche cOmplete™ protease inhibitor cocktail (final EDTA concentration: 1mM). About 30 pairs of salivary glands in 60 µl of PBS containing cOmplete were prepared for one replicate of protein extraction and 60 µl of 12M

urea, 4M thiourea and Complete was added before purification by the 2D Clean-up Kit (GE HealthCare). Resuspended protein lysates were quantified using the Qubit™ protein quantification system (Invitrogen), following the manufacturer's instructions and 20 µg and 5 µg was removed from each sample for protein digestion and SDS PAGE respectively. Salivary gland protein quality was assessed using SDS PAGE. Samples were diluted with an equal volume of 2X Laemmli sample buffer (Sigma Aldrich; 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris HCl, pH 6.8), loaded onto a 12% resolving, 4% stacking SDS-PAGE gel and electrophoresed at 110 V for two hours and thirty minutes. Gels were stained using the PloS One Silver Stain Kit (GE HealthCare) following the manufacturer's instructions.

### Sample preparation for mass spectrometry

Three independent biological replicates were analyzed in this study. 50mM ammonium bicarbonate was added to each sample and proteins were reduced with 0.5M dithiothreitol (DTT) at 56°C for 20 min and alkylated with 0.55M iodoacetamide (IAA) at room temperature for 15 min, in the dark. 1 µl of a 1% w/v solution of Protease Max Surfactant Trypsin Enhancer (Promega) and 0.5 µg of Sequence Grade Trypsin (Promega) was added to give a protein:trypsin ratio of 40:1 and 80:1 for saliva and salivary glands respectively. The protein/trypsin mixture was incubated at 37°C for 18 hours. Digestion was terminated by adding 1 µl of 100% trifluoroacetic acid (Sigma Aldrich) and incubation at room temperature for 5 min. Samples were centrifuged for 10 min at 13,000 × g and purified for mass spectrometry using either the ZipTip pipette procedure (Millipore) for saliva and C18 Spin Columns (Pierce) for salivary glands, following the manufacturer's instructions. The eluted peptides were dried using a SpeedyVac concentrator (Thermo Scientific Savant DNA120) and resuspended in 2% v/v acetonitrile and 0.05% v/v trifluoroacetic acid (TFA). Samples were sonicated for 5 min to aid peptide resuspension followed by centrifugation for 5 min at 13,000 × g. The supernatant was removed and used for mass spectrometry.

### Mass spectrometry

1 µg of each digested sample was loaded onto a QExactive (ThermoFisher Scientific) high-resolution accurate mass spectrometer connected to a Dionex Ultimate 3000 (RSLCnano) chromatography system. The peptides were separated by a 2% to 40% gradient of acetonitrile on a Biobasic C18 Picofrit™ column (100mm length, 75mm ID), using a 120 minutes reverse-phase gradient at a flow rate of 250nL min<sup>-1</sup> with a runtime of 50 and 130 minutes for saliva and salivary glands respectively. All data were acquired with the mass spectrometer operating in automatic data dependent switching mode. A full MS scan at 70,000 resolution and a range of 400-1600 m/z was

followed by an MS/MS scan, resolution 17,500 and a range of 200-2000 m/z, selecting the 15 most intense ions prior to MS/MS.

Protein identification of MS/MS data was performed using MaxQuant v1.5.6.5 (<http://www.maxquant.org>) following the general procedures and settings outlined in Hubner et al. (2010). The Andromeda search algorithm (Cox et al. 2011) incorporated in the MaxQuant software was used to correlate MS/MS data against the protein reference sequence set of *A. pisum* obtained from the NCBI (27984 entries, may 2016) and a contaminant sequence set provided by MaxQuant. The following search parameters were used: first search peptide tolerance of 20 ppm, second search peptide tolerance 4.5 ppm with cysteine carbamidomethylation as a fixed modification and N-acetylation of protein and oxidation of methionine as variable modifications and a maximum of two missed cleavage sites allowed. False Discovery Rates (FDR) were set to 1% for both peptides and proteins and the FDR was estimated following searches against a target-decoy database. Peptides with minimum length of seven amino acids were considered for identification.

### Data analysis

Perseus v.1.5.5.3 ([www.maxquant.org](http://www.maxquant.org)) was used for data analysis, processing and visualization. The data matrix was first filtered for the removal of contaminants and peptides identified by site. LFQ intensity values were log<sub>2</sub> transformed and proteins not found all three replicates in at least one group were omitted from the analysis. A data-imputation step was conducted to replace missing values with values that simulate signals of low abundant proteins chosen randomly from a distribution specified by a downshift of 1.8 times the mean standard deviation (SD) of all measured values and a width of 0.3 times this SD.

### ***Gene ontology and secretion prediction***

The GO annotation was performed on the whole *A. pisum* proteome available on NCBI using Blast2GO v2.5.0 (Götz et al. 2008; [www.blast2go.org/](http://www.blast2go.org/)). Associations were realized using a blastp (BLAST+ v2.5.0) against nr database (release 2017-2-4) with the following parameters: `evalue=1e-8`, `max_target_seqs=20`, `soft_making=false`, and Interproscan v5.13.52.0. Assigned GO terms for gene of interest groups were categorized by molecular function (MF), biological process (BP) and cellular component (CC) and GO enrichment was investigated using hypergeometric tests (stats R package, R Core Team, 2017) and p-values were adjusted for multiple comparisons using Bonferroni correction. *A. pisum* whole proteome was analyzed with SignalP v3.0 and v4.1 ([www.cbs.dtu.dk/services/SignalP](http://www.cbs.dtu.dk/services/SignalP)) and SecretomeP v2.0 ([www.cbs.dtu.dk/services/SecretomeP](http://www.cbs.dtu.dk/services/SecretomeP); mammalian mode; NN-score $\geq$ 0.6) to

characterize the presence of signal peptide or non-classical secretion signal, respectively. Then, membrane inserted domains were predicted using TMHMM v2.0 ([www.cbs.dtu.dk/services/TMHMM](http://www.cbs.dtu.dk/services/TMHMM)) for transmembrane domains and PredGPI ([gpcr2.biocomp.unibo.it/predgpi](http://gpcr2.biocomp.unibo.it/predgpi)) for GPI anchors. Finally, the results were combined to define the list of *A. pisum* secreted proteins that have secretion signals and no membrane insertion domains.

### ***Orthology analysis and species phylogeny***

A comparative genomic analysis by OrthoDB (Kriventseva et al. 2015) was performed between *A. pisum* genome and 16 arthropod genomes (15 insects and one arachnid; supplementary table S7, Supplementary Material online) to assess groups of orthologues and define orthology level of *A. pisum* proteins. The longest isoform of each gene from different genomes was isolated with a home-made perl script, and OrthoDB\_soft\_1.6 (ORTHOPIPE v5.4.5, SWIPE v2.0.11, cd-hit v4.6, BLAST+ 2.3.0, [www.orthodb.org/](http://www.orthodb.org/)) was run to generate groups of orthologs. 478 groups of conserved single copy orthologues were extracted and their protein sequences were aligned with MAFFT (Katoh et al. 2005). Alignments were concatenated by species and used to generate a maximum likelihood phylogenetic tree with RAxML (Stamatakis 2006; Stamatakis et al. 2008) with default parameters and 1000 replicates considering the arachnid species (*Tetranychus urticae*) as outgroup. The phylogeny was then used to establish the different level of orthology we would assess to each *A. pisum* gene. The enrichments of orthologous categories among dataset were analyzed using hypergeometric test implemented in the stats R package (R Core Team 2017) followed by and p-values adjustments for multiple comparisons using Bonferroni correction.

### ***Evolution analyses of salivary genes***

To examine the evolutionary patterns of *A. pisum* salivary genes, we estimated the ratio of nonsynonymous (dN) to synonymous substitutions (dS), dN/dS or  $\omega$ , across closely related species in Aphididae. The dN/dS is an indicator of evolutionary pressures: low ratios reflect highly constrained sequences under strong purifying selection, ratios close to the unity signify relaxed selection and values above unity result from positive selection promoting accelerated changes at the amino acid level in response to environment pressure (Yang and Bielawski 2000; Yang and Bielawski 2000).

### **Single copy genes**

In order to describe the evolutionary patterns of aphid single copy genes and compare the different categories of salivary effectors, the orthologue groups containing a pea aphid single copy

gene and a single copy orthologue in at least one other Aphididae species were extracted. When alternative transcripts of a gene were present in the different aphid species datasets, the longest CDS was used for the analysis. CDS from orthologous gene pairs (*A. pisum/A. glycines*, *A. pisum/D. noxia*, *A. pisum/M. persicae*) were aligned using MAFFT (Kato et al. 2005) within the Hot algorithm from GUIDANCE2 (Sela et al. 2015) to produce reliable codon-based alignments by removing poorly aligned regions. From these codon alignments, pairwise dN/dS values were calculated with PAML v4.8 using the YN00 program (Yang and Nielsen 2000; Yang 2007). To limit the presence of inaccurate dN/dS estimations, different filters were applied before comparison of evolutionary rates. We removed orthologue pairs with a dS > 2 as they probably suffer from saturation, and insufficiently diverged pairs were removed whether they showed dN or dS < 0.01 and less than 50 synonymous sites as described in Mathers et al. (2017). Then, dN/dS distributions of different gene categories within each pairwise analysis were compared using a non-parametric Kruskal-Wallis test after assessment of the non-normality and heteroscedasticity of the data and Nemenyi-Tests for multiple comparisons were realized with the R package PMCMR (Pohlert and Pohlert 2016).

In addition to the four aphid species which are included in the OrthoDB analysis, orthologues of characterized salivary effectors were searched using blastn (BLAST+ v2.5.0) in all Aphididae sequences available in Genbank database and in 454 contigs of *Rhopalosiphon padi* and *Schizaphis graminum*, created for this study (see supplementary material and methods S1, Supplementary Material online) in order to compute more accurate evolutionary rates and test selection models. CDS sequences (from start to stop) were used to produce codon-based alignments with PRANK (Löytynoja and Goldman 2005). The reliability of alignments was assessed using GUIDANCE2 (Sela et al. 2015) and poorly aligned regions with score under 0.93 were removed. Final alignments were used to generate maximum likelihood phylogenetic trees with RAxML and both served to perform an estimation of dN/dS ratio by maximum likelihood analysis of codon substitutions with codeml implemented in PAML v4.8 (Yang 2007) under different models (Yang et al. 2000; Yang and Nielsen 2000). The null models (M0=one ratio, M1=neutral, M7= $\beta$ ) were compared to alternative models (M2=selection, M8= $\beta + \omega$ ) using the likelihood ratio test (LRT), which compares twice the difference in log likelihood to a  $\chi^2$  distribution with required number of degrees of freedom. Finally, when selection models were significant, the Bayes Empirical Bayes (BEB) procedure was used to compute the posterior probability that a site belong to the class of positive selection (Yang et al. 2005).

### Gene families

*A. pisum* salivary gland expressed gene families and their orthologues in arthropod species were extracted from OrthoDB groups. In a first step, amino acid sequences were aligned with PRANK



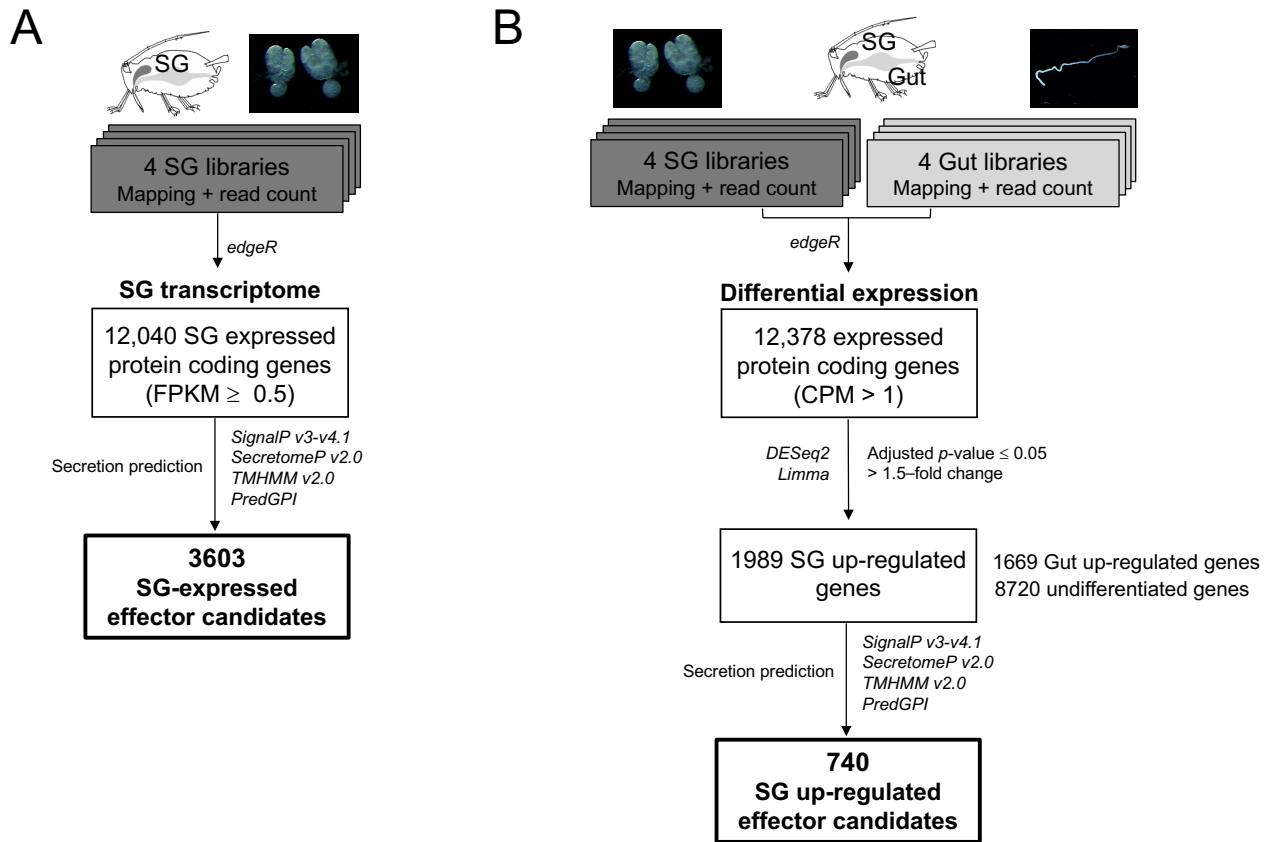
(Löytynoja and Goldman 2005), before the construction of phylogenetic trees with RAxML (Stamatakis 2006; Stamatakis et al. 2008) (1000 replicates). Based on phylogeny relationships, subclades were defined for each gene family. Starting from Aphididae CDS, the combination of PRANK aligner GUIDANCE2 (Sela et al. 2015) and RAxML (1000 replicates) provided reliable codon-based alignments and phylogenetic trees for further analyses of codon substitutions. The codeml branch-site (BS) model (Yang and Nielsen 2002) was performed to test whether positive selection affected certain sites on a gene copy in a particular species (selected foreground branch compared to background branches). The BS model classified the sites in 4 categories, the class 0 where codons were under purifying selection ( $\omega < 1$ ) on both foreground and background branches, the class 1 where codons evolved under neutral evolution ( $\omega = 1$ ) on both foreground and background branches, the class 2a where the sites were under positive selection ( $\omega \geq 1$ ) on the foreground branch and under purifying selection ( $\omega < 1$ ) on background branches and the class 2b where the codons evolved under positive selection ( $\omega \geq 1$ ) on the foreground branch and under neutral evolution ( $\omega = 1$ ) on background branches. The null model in which the foreground branch may have different proportions of the sites under neutral evolution (dN/dS fixed to 1), and the alternative model, in which the foreground branch may have a proportion of the sites under positive selection, were applied to all branches of each gene families. To evaluate the significance of the selection model, the LRT calculation and comparison to a  $\chi^2$  distribution were performed and provided p-values which were adjusted for multiple testing using the Hommel's method as described in Anisimova and Yang (2007). When the selection model was significant, the BEB procedure was realized to sites belonging to the class of positive selection in particular branch of the tree.

## FIGURES AND TABLES

**Table 1.** Aphid single copy salivary effector genes characterized for their role in the interaction with the plant.

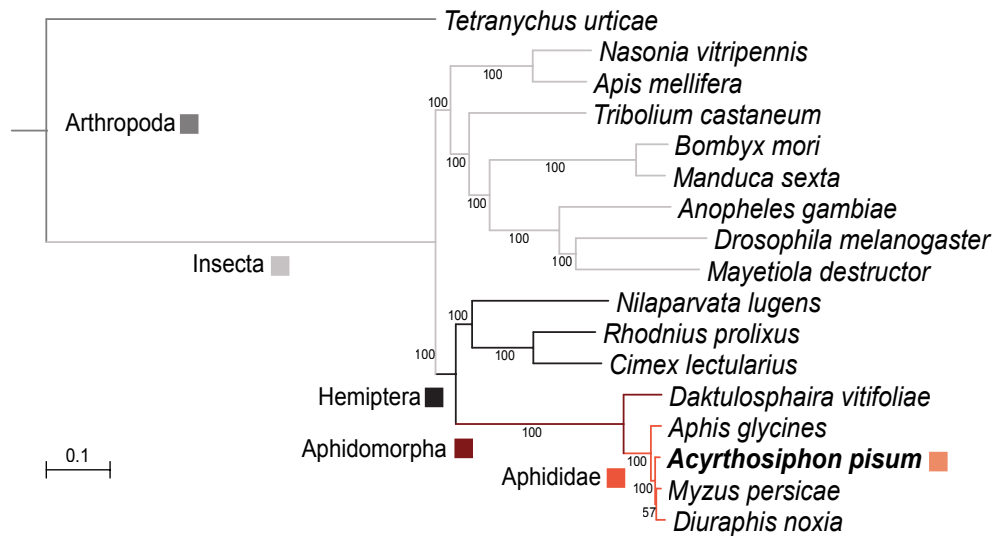
Gene	<i>A. pisum</i> gene	Effector set	Expression rank	Orthology level	$\omega$ dN/dS	Aphid phenotype	References*
<i>Mp1</i>	LOC100165393 ACYPI006346	SG-up-regulated	19/740	Aphididae	0.64653	Expression promotes fecundity	[3,4,6,10,15]
<i>C002</i>	LOC100167863 ACYPI008617	SG-up-regulated	12/740	Aphididae	0.62307	Expression promotes fecundity Silencing reduces survival/fecundity	[1,2,3,4,6,13]
<i>Ap25</i>	LOC100169287 ACYPI009919	SG-up-regulated	57/740	Aphididae	0.56219	Expression promotes fecundity	[14]
<i>Me23</i>	LOC100161198 ACYPI002439	SG-up-regulated	109/740	Aphididae	0.53025	Expression promotes fecundity	[5]
<i>Me10</i>	LOC100167427 ACYPI008224	SG-up-regulated	5/740	Aphididae	0.51787	Expression promotes fecundity	[5]
<i>Shp</i>	LOC100169243 ACYPI009881	SG-up-regulated	1/740	Aphidomorpha	0.48863	Silencing reduces survival/fecundity	[8,12]
<i>Mp55</i>	LOC100569515 ACYPI33755	SG-up-regulated	18/740	Aphididae	0.42972	Expression promotes fecundity Silencing reduces survival/fecundity	[7]
<i>Mp2</i>	LOC100160479 ACYPI001774	SG-up-regulated	107/740	Aphididae	0.40948	Expression promotes fecundity Silencing reduces survival/fecundity	[6]
<i>Mp10</i>	LOC100145855 ACYPI000097	SG-up-regulated	303/740	Insect	0.12212	Expression reduces fecundity	[3]
<i>Mif1</i>	LOC100161225 ACYPI002465	SG-expressed	4700/12040	Arthropod	0.09454	Silencing reduces fecundity	[9]
<i>Armet</i>	LOC100167188 ACYPI008001	SG-up-regulated	122/740	Arthropod	0.06298	Silencing reduces survival	[11]

Expression ranks show the rank of effector SG expression level (based on Log[FPKM]) within the SG-up-regulated effector set except for *Mif1* that does not belong to effector sets (global SG expression scale).  $\omega$  values represent the evolutionary rates within closely related Aphididae species. \*<sup>1</sup>Mutti et al. 2006; <sup>2</sup>Mutti et al. 2006; <sup>3</sup>Bos et al. 2010; <sup>4</sup>Pitino et al. 2011; <sup>5</sup>Atamian et al. 2012; <sup>6</sup>Pitino and Hogenhout 2012; <sup>7</sup>Elzinga et al. 2014; <sup>8</sup>Abdellatef et al. 2015; <sup>9</sup>Naessens et al. 2015; <sup>10</sup>Pan et al. 2015; <sup>11</sup>Wang et al. 2015; <sup>12</sup>Will and Vilcinskis 2015; <sup>13</sup>Zhang et al. 2015; <sup>14</sup>Guy et al. 2016; <sup>15</sup>Rodriguez et al. 2017

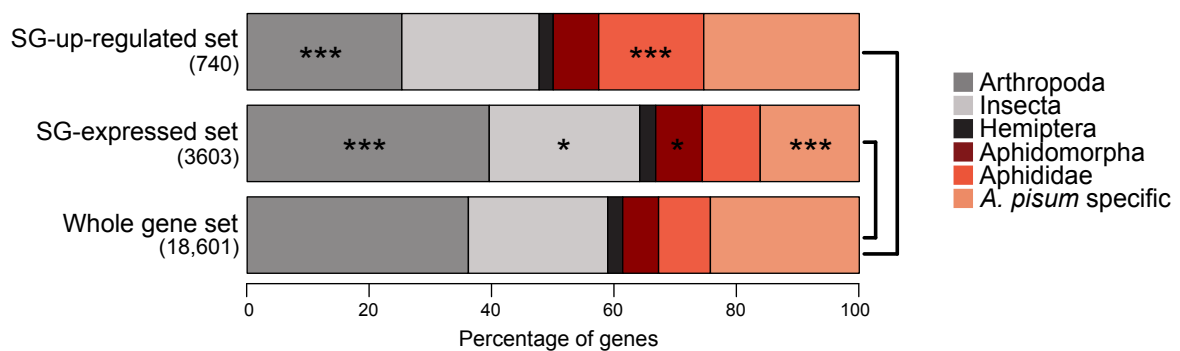


**Fig. 1.** Overview of pipelines to identify *A. pisum* salivary effector candidate gene sets. (A) Identification of the SG-expressed effector candidates. The SG transcriptome revealed 12,040 genes expressed in SG tissue and 3603 of them were predicted to be secreted. (B) Identification of the SG up-regulated effector candidates. The differential expression analysis between SG and gut samples identified 1989 genes up-regulated in SG and of these, 740 were predicted to be secreted.

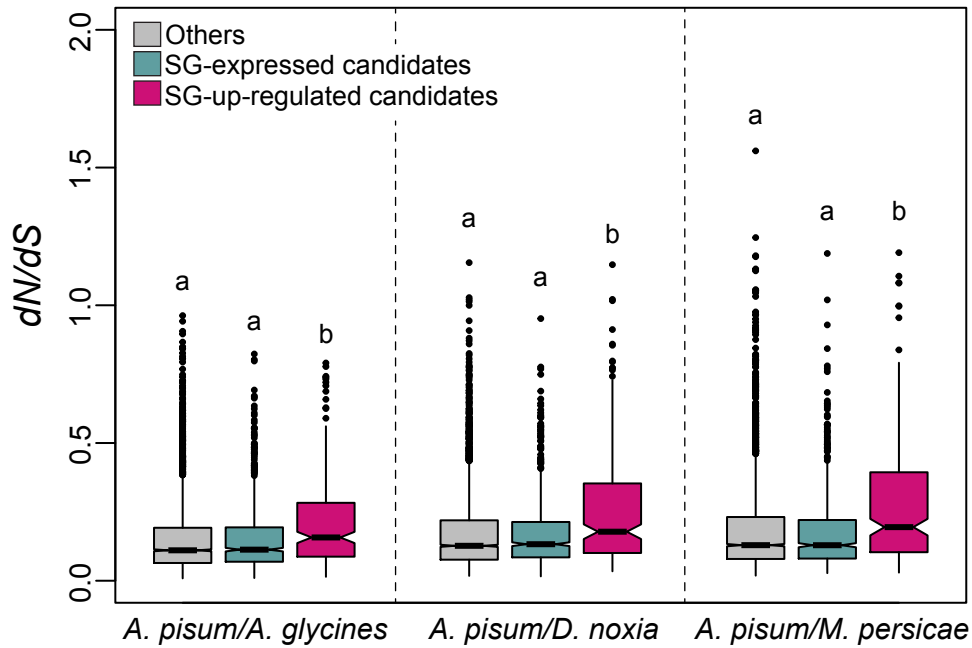
A



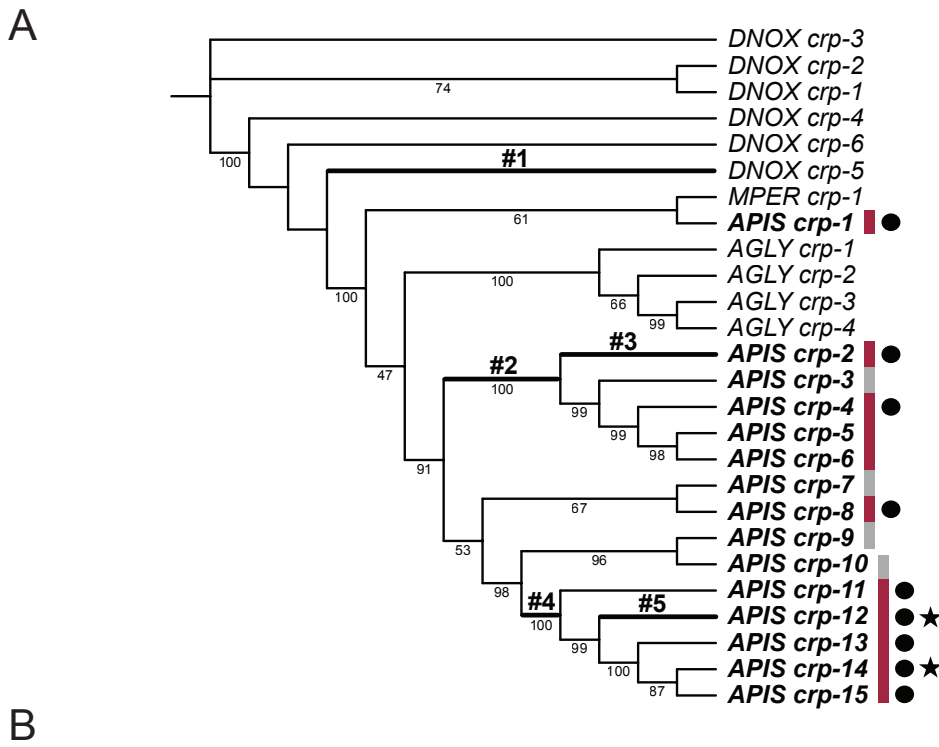
B



**Fig. 2.** Orthology profile of *A. pisum* SG effector candidate gene sets. (A) Phylogeny of arthropod and insect species analyzed to determine orthologue groups. 478 single copy orthologous gene groups were aligned with MAFFT (Kato et al. 2005) and a protein maximum likelihood phylogenetic tree was built by RAxML (Stamatakis 2006; Stamatakis et al. 2008) with 1000 replicates. The color squares on the tree indicate the 6 levels of orthology determined and used to examine the conservation of *A. pisum* genes. The tree was designed in iTOL (Letunic and Bork 2007). (B) Proportions of the different orthology levels in SG effector sets and *A. pisum* genome. Proportion of each orthology group of SG-expressed or SG-up-regulated effector sets were compared with that of the global gene set using hypergeometric tests. p-values were adjusted by Bonferroni correction. The stars indicate significant differences between SG-expressed or SG-up-regulated effector sets and the global gene set. Significance degree: \* $p < 0.05$ , \*\*\*  $p < 0.001$ .

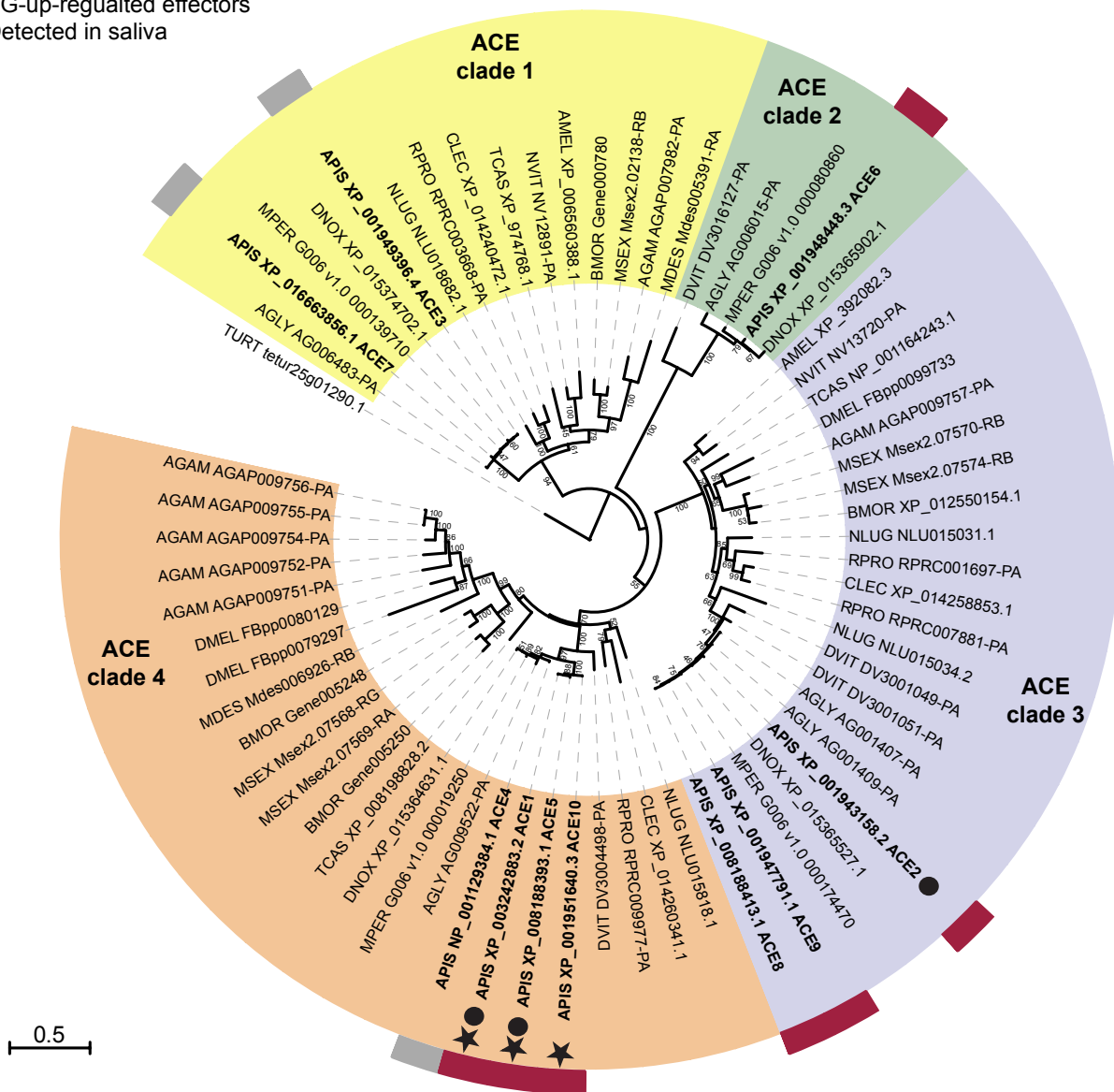


**Fig. 3.** Evolutionary rates of single copy Aphididae genes orthologous to *A. pisum* effector candidates. Note that SG-expressed candidates do not contain SG-up-regulated candidates. Pairwise dN/dS values were calculated using the YN00 model in PAML (Yang 1997; Yang and Nielsen 2000). For the three pairwise comparison between *A. pisum* and *A. glycines*, *D. noxia* and *M. persicae*, 4781, 4696 and 4491 single copy orthologue pairs were retained respectively. a and b indicate groups determined by multiple comparisons after Kruskal-Wallis test within each pairwise comparison (*A. pisum/A. glycines*:  $H=37.322$ , 2 d.f.,  $P>0.001$ ; *A. pisum/D. noxia*:  $H=42.498$ , 2 d.f.,  $P>0.001$ ; *A. pisum/M. persicae*:  $H=47.962$ , 2 d.f.,  $p>0.001$ ).

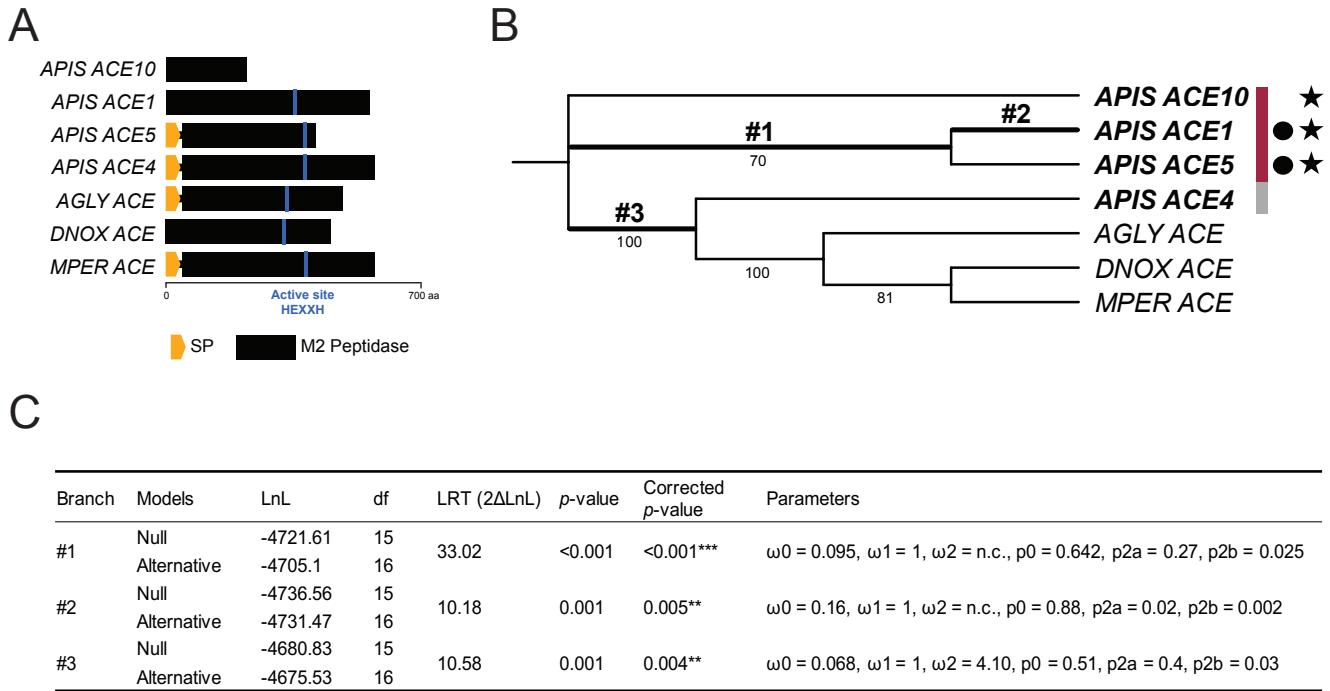


**Fig. 4.** Evolutionary analysis of the Aphididae *CRP* gene family. (A) Phylogenetic tree used to compute dN/dS under the branch-site (BS) model of codon substitution. The maximum likelihood phylogenetic tree was based on the codon based-alignment of Aphididae sequences and supporting bootstrap scores of 1000 replicates are indicated on the tree whether they were above 40. DNOX, MPER, AGLY and APIS indicate sequences from *D. noxia*, *M. persicae*, *A. glycines* and *A. pisum* respectively. The *A. pisum* genes that were up-regulated in SG tissue are indicated in magenta and undifferentiated or unexpressed ones in grey. The black circles show the *A. pisum* genes referenced in the SG up-regulated effector gene set and the stars highlight genes encoding for proteins detected by saliva proteomics. The branches where positive selection affected certain site specifically are in bold. (B) Estimated parameters of the 5 branches for which the alternative condition of the branch-site model analysis proved to be significant.  $\omega_0$ ,  $\omega_1$  and  $\omega_2$  indicate the average dN/dS ratio for sites assigned in the class 0 ( $\omega < 1$ , purifying selection), class 1 ( $\omega = 1$ , neutral evolution) and class 2 ( $\omega > 1$ , positive selection) respectively in the foreground branch, and n.c. indicates a fail in calculation of dN/dS due to a low or null dS (synonymous substitutions).  $p_0$ ,  $p_{2a}$  and  $p_{2b}$  show proportions of sites in classes 0, 2a and 2b respectively for each foreground branch (Yang and Nielsen 2002).

- A. *pisum* gene expression
- SG up-regulated
  - Undifferentiated
  - SG-up-regulated effectors
  - ★ Detected in saliva



**Fig. 5.** Phylogeny of the *ACE* gene family among insect species. The maximum likelihood phylogenetic tree was based on the alignment of proteins from orthologue groups. The supporting bootstrap score from 1000 replicates is indicated on each branch of the tree when above 40. Species abbreviations are AGAM, *Anopheles gambiae*; AGLY, *Aphis glycines*; AMEL, *Apis mellifera*; APIS, *Acyrtosiphon pisum*; BMOR, *Bombyx mori*; CLEC, *Cimex lectularius*; DMEL, *Drosophila melanogaster*; DNOX, *Diuraphis noxia*; DVIT, *Daktulosphaira vitifoliae*; MDES, *Mayetiola destructor*; MPER, *Myzus persicae*; MSEX, *Manduca sexta*; NLUG, *Nilaparvata lugens*; NVIT, *Nasonia vitripennis*; RPRO, *Rhodnius prolixus*; TCAS, *Tribolium castaneum*; TURT, *Tetranychus urticae*. The tree was designed in iTOL (Letunic and Bork 2007).

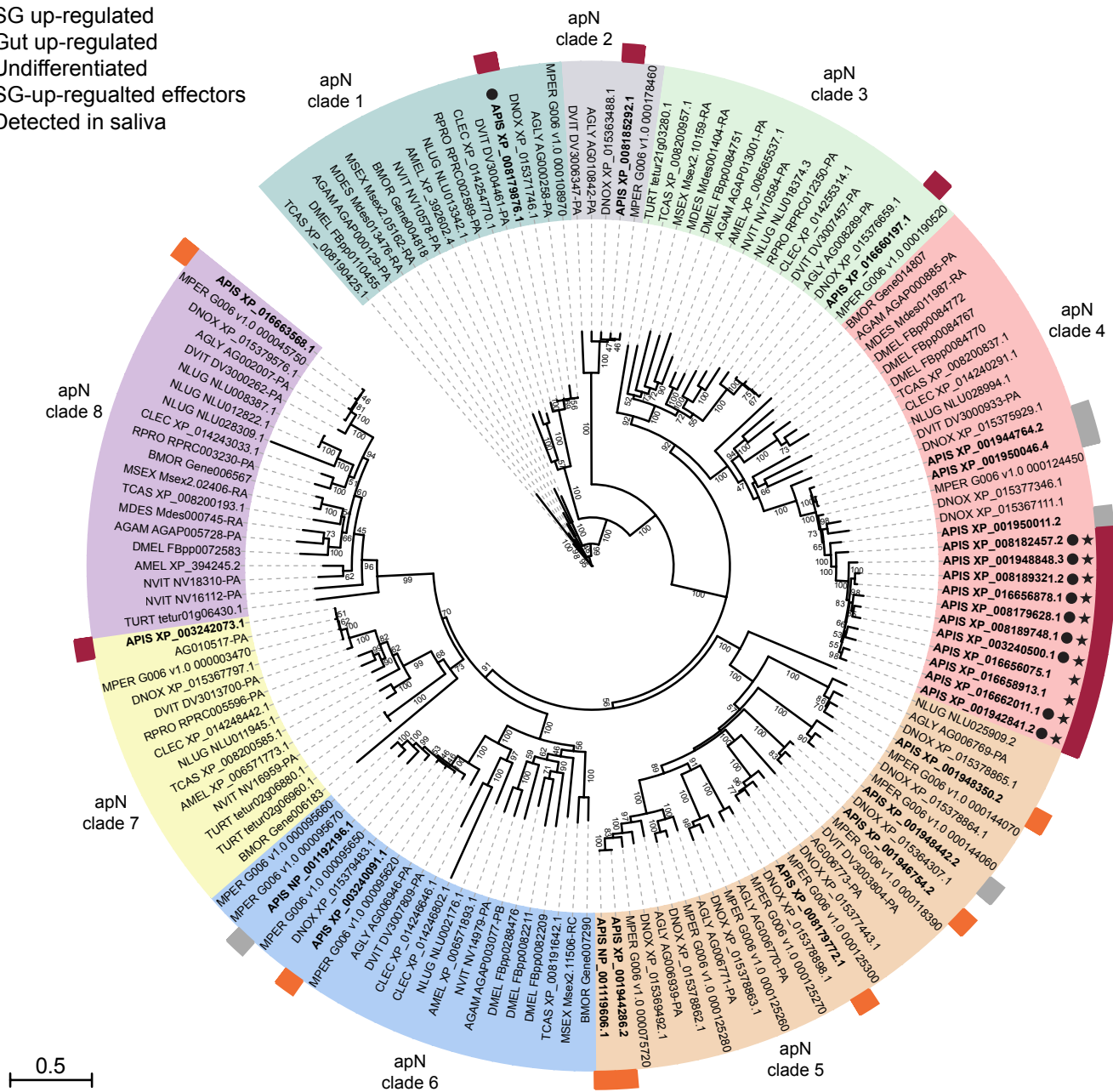


**Fig. 6.** Evolutionary analysis of the Aphididae clade 4 of the *ACE* gene family. (A) Protein structure of the Aphididae ACE from the clade 4. Protease active sites are shown in blue lines. (B) Phylogenetic tree used to compute dN/dS under the branch-site (BS) model of codon substitution. The maximum likelihood phylogenetic tree was based on the codon based-alignment of Aphididae sequences and supporting bootstrap scores of 1000 replicates are indicated on the tree whether they were above 40. The tree was designed in iTOL (Letunic and Bork 2007). DNOX, MPER, AGLY and APIS indicate sequences from *D. noxia*, *M. persicae*, *A. glycines* and *A. pisum* respectively. The *A. pisum* genes that were up-regulated in SG tissue are indicated in magenta and undifferentiated or unexpressed ones in grey. The black circles show the *A. pisum* genes categorized as the SG up-regulated effector genes and the stars highlight the genes encoding for proteins detected by saliva proteomics. The branches where positive selection affected specific site are shown in bold. (C) Estimated parameters of the 3 branches for which the alternative condition of the branch-site model analysis proved to be significant.  $\omega_0$ ,  $\omega_1$  and  $\omega_2$  indicate the average dN/dS ratio for sites assigned in the class 0 ( $\omega < 1$ , purifying selection), class 1 ( $\omega = 1$ , neutral evolution) and class 2 ( $\omega > 1$ , positive selection) respectively in the foreground branch, and n.c. indicates a fail in calculation of dN/dS due to a low or null dS (synonymous substitutions).  $p_0$ ,  $p_{2a}$  and  $p_{2b}$  show proportions of sites in classes 0, 2a and 2b respectively for each foreground branch (Yang and Nielsen 2002).

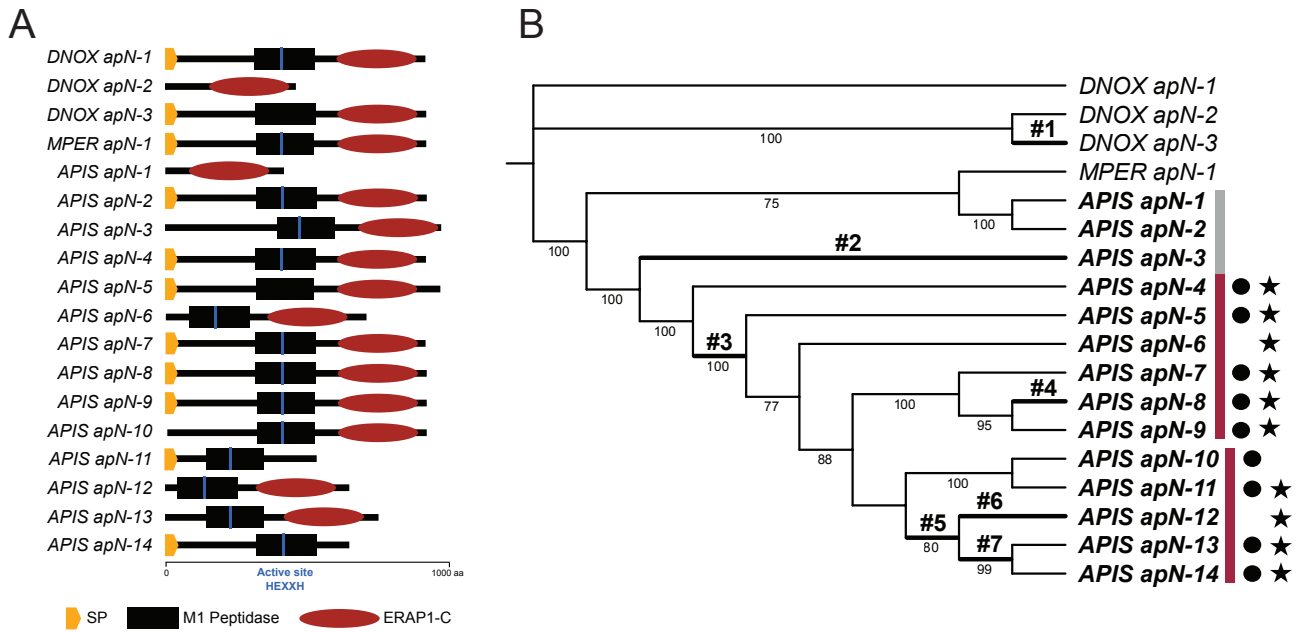


A. *pisum* gene expression

- SG up-regulated
- Gut up-regulated
- Undifferentiated
- SG-up-regulated effectors
- ★ Detected in saliva



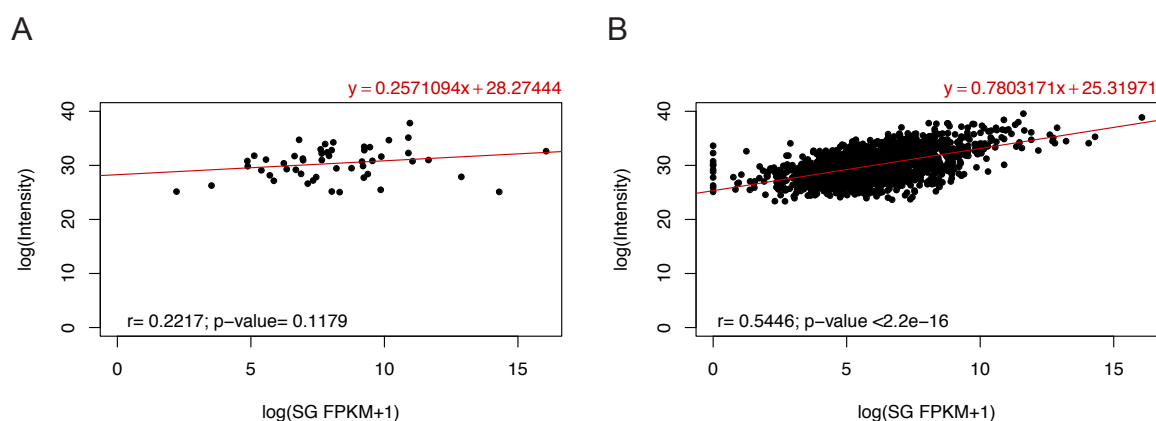
**Fig. 7.** Phylogeny of the *apN* gene family among insect species. The maximum likelihood phylogenetic tree was based on the alignment of *apN* proteins from orthologue groups. The supporting bootstrap score from 1000 replicates is indicated on each branch of the tree when above 40. Species abbreviations are AGAM, *Anopheles gambiae*; AGLY, *Aphis glycines*; AMEL, *Apis mellifera*; APIS, *Acyrtosiphon pisum*; BMOR, *Bombyx mori*; CLEC, *Cimex lectularius*; DMEL, *Drosophila melanogaster*; DNOX, *Diuraphis noxia*; DVIT, *Daktulosphaira vitifoliae*; MDES, *Mayetiola destructor*; MPER, *Myzus persicae*; MSEX, *Manduca sexta*; NLUG, *Nilaparvata lugens*; NVIT, *Nasonia vitripennis*; RPRO, *Rhodnius prolixus*; TCAS, *Tribolium castaneum*; TURT, *Tetranychus urticae*. The tree was designed in iTOL (Letunic and Bork 2007).



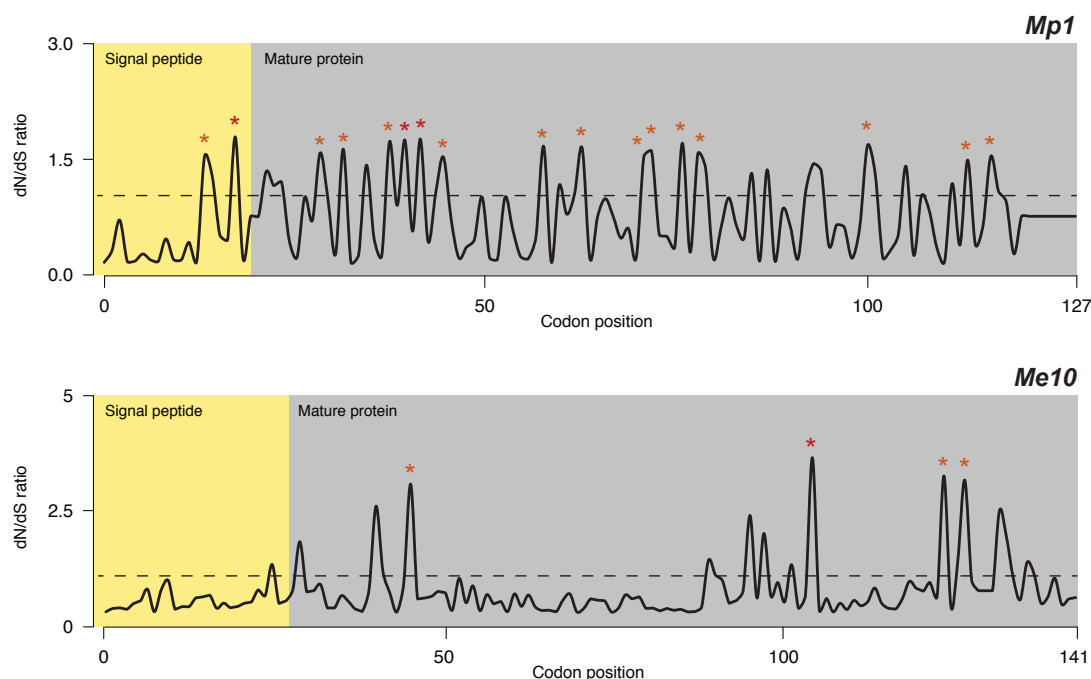
Branch	Models	LnL	df	LRT (2ΔLnL)	p-value	Corrected p-value	Parameters
#1	Null	-14944.26	37	6.68	0.009	0.048*	$\omega_0 = 0.14, \omega_1 = 1, \omega_2 = 3.57, p_0 = 0.57, p_{2a} = 0.09, p_{2b} = 0.04$
	Alternative	-14940.92	38				
#2	Null	-14944.09	37	6.97	0.008	0.039*	$\omega_0 = 0.15, \omega_1 = 1, \omega_2 = \text{n.c.}, p_0 = 0.65, p_{2a} = 0.005, p_{2b} = 0.002$
	Alternative	-14940.6	38				
#3	Null	-14945.9	37	12.94	<0.001	0.001**	$\omega_0 = 0.15, \omega_1 = 1, \omega_2 = 8.24, p_0 = 0.61, p_{2a} = 0.05, p_{2b} = 0.025$
	Alternative	-14939.42	38				
#4	Null	-14949.01	37	7.31	0.006	0.031*	$\omega_0 = 0.15, \omega_1 = 1, \omega_2 = 24.8, p_0 = 0.65, p_{2a} = 0.007, p_{2b} = 0.004$
	Alternative	-14945.34	38				
#5	Null	-14948.09	37	15.11	<0.001	<0.001***	$\omega_0 = 0.15, \omega_1 = 1, \omega_2 = \text{n.c.}, p_0 = 0.64, p_{2a} = 0.005, p_{2b} = 0.003$
	Alternative	-14940.53	38				
#6	Null	-14949.42	37	24.19	<0.001	<0.001***	$\omega_0 = 0.15, \omega_1 = 1, \omega_2 = 27.7, p_0 = 0.62, p_{2a} = 0.03, p_{2b} = 0.01$
	Alternative	-14937.32	38				
#7	Null	-14949.43	37	10.08	0.001	0.006**	$\omega_0 = 0.15, \omega_1 = 1, \omega_2 = \text{n.c.}, p_0 = 0.65, p_{2a} = 0.006, p_{2b} = 0.003$
	Alternative	-14949.43	38				

**Fig. 8.** Evolutionary analysis of the Aphididae clade 4 of the *apN* gene family. (A) Protein structure of the Aphididae *apN* from the clade 4. (B) Phylogenetic tree used to compute dN/dS under the branch-site (BS) model of codon substitution. The maximum likelihood phylogenetic tree by RAxML was based on the codon based-alignment of Aphididae *apN* sequences and supporting bootstrap scores of 1000 replicates are indicated on the tree whether they were above 40. The tree was designed in iTOL (Letunic and Bork 2007). DNOX, MPER, AGLY and APIS indicate sequences from *D. noxia*, *M. persicae*, *A. glycines* and *A. pisum* respectively. The *A. pisum* genes that were up-regulated in SG tissue are indicated in magenta and undifferentiated or unexpressed ones in grey. The black circles show the *A. pisum* genes referenced in the SG up-regulated effector gene set and the stars highlight genes encoding for proteins detected by saliva proteomics. The branches where positive selection affected certain site specifically are in bold. (C) Estimated parameters of the 7 branches for which the alternative condition of the branch-site model analysis proved to be significant.  $\omega_0$ ,  $\omega_1$  and  $\omega_2$  indicate the average dN/dS ratio for sites assigned in the class 0 ( $\omega < 1$ , purifying selection), class 1 ( $\omega = 1$ , neutral evolution) and class 2 ( $\omega > 1$ , positive selection) respectively in the foreground branch, and n.c. indicates a fail in calculation of dN/dS due to a low or null dS (synonymous substitutions).  $p_0$ ,  $p_{2a}$  and  $p_{2b}$  show proportions of sites in classes 0, 2a and 2b respectively for each foreground branch (Yang and Nielsen 2002).

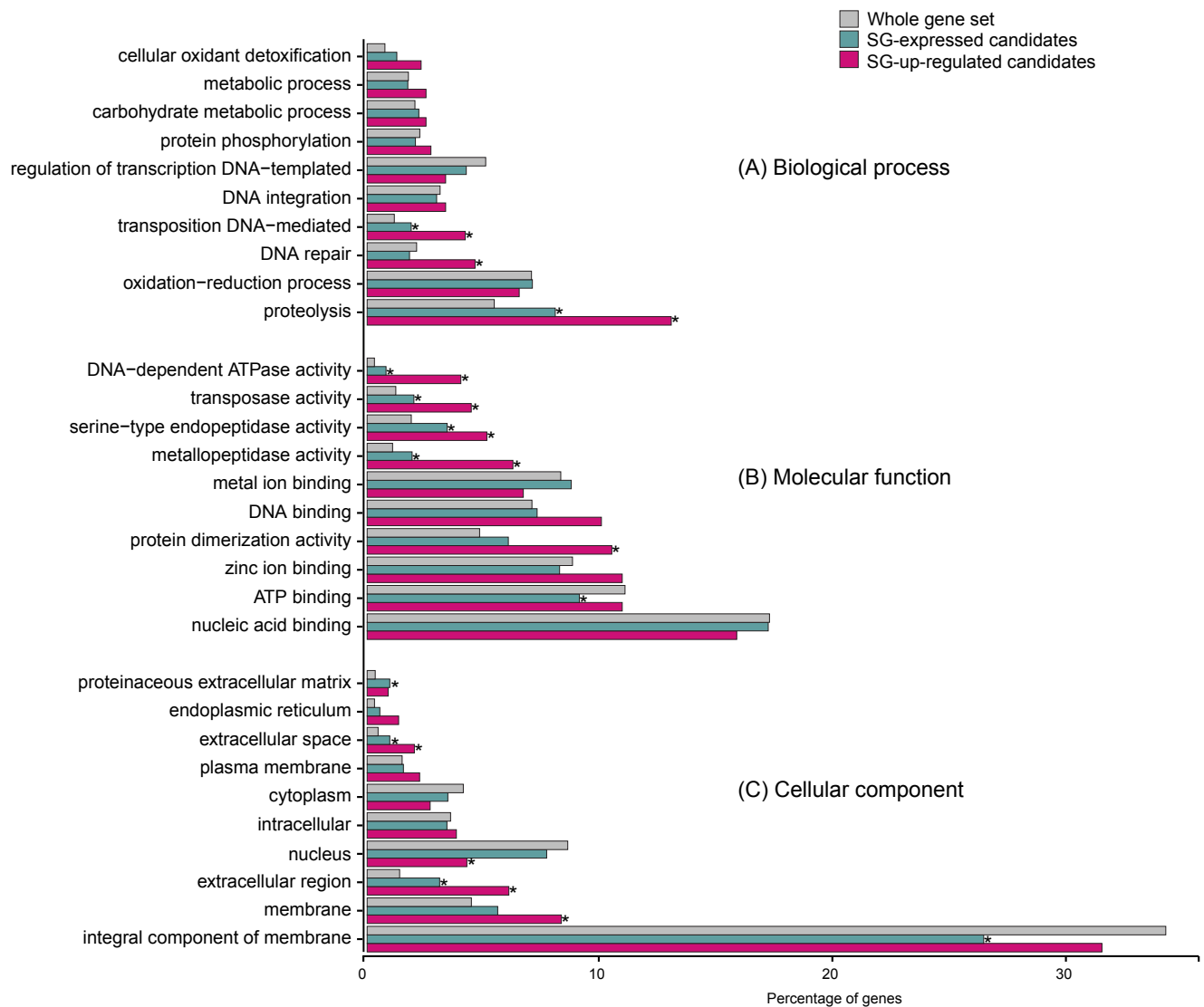
## SUPPLEMENTARY MATERIAL



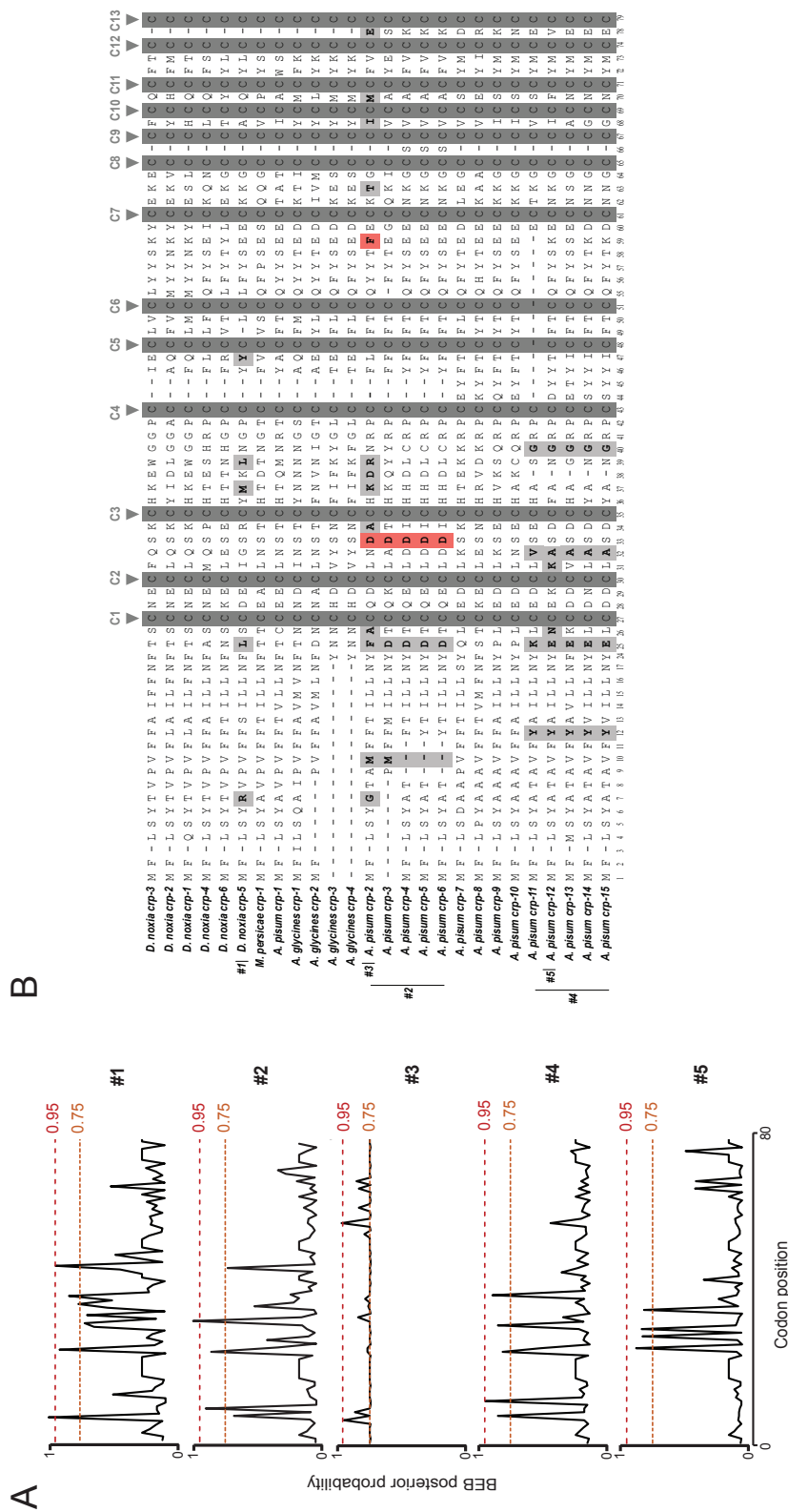
**Supplementary fig. S1.** Correlation between gene expression and protein intensity. (A) Correlation plot of the 51 saliva detected proteins and their corresponding SG gene expression levels. (B) Correlation plot of the 1809 SG detected proteins and their corresponding SG gene expression levels. The correlation coefficients were calculated with Pearson method after the assessment of the distributions normality by Shapiro-Wilk normality test and QQ-plot viewing.



**Supplementary fig. S2.** Sites under positive selection in the Aphididae genes *Mp1* and *Me10*. Plots of the dN/dS values calculated for each site under the M8 model of codon substitutions with selection. The dashed line indicates a dN/dS equal to 1 meaning neutral evolution. The significance of each dN/dS value is shown by stars, red indicates codons with BEB posterior probability to be under selection ( $dN/dS > 1$ ) above 0.95 and orange BEB posterior probability above 0.75.

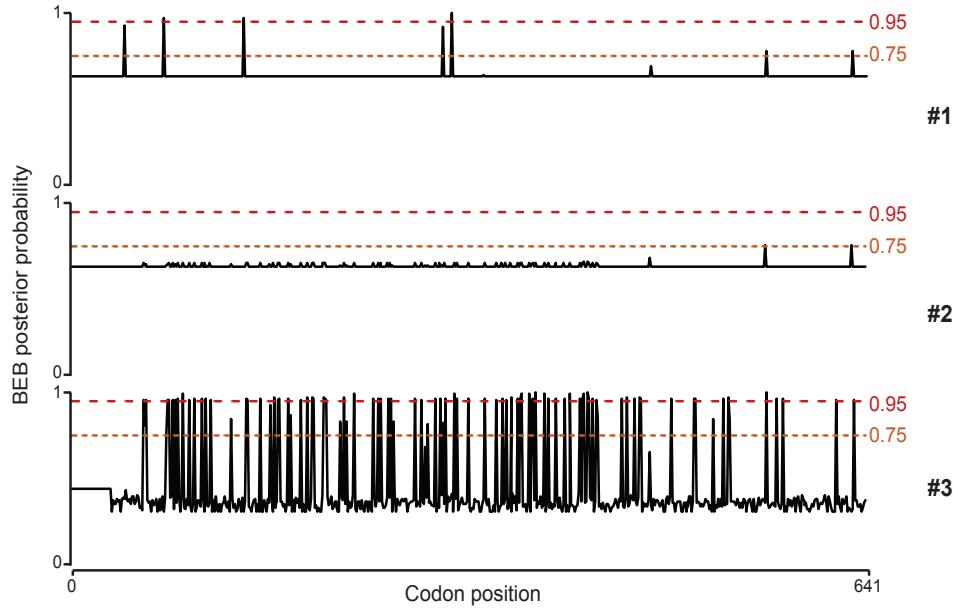


**Supplementary fig. S3.** Gene ontology analysis of salivary effector candidate gene sets. The proportions of genes for each GO term (biological process, BP; molecular function, MF and cellular component, CC) were compared between SG-expressed or SG-up-regulated effector sets and the whole gene set. Within these 3 categories, the 10 most abundant GO terms identified in SG-up-regulated effector candidate set were kept for enrichment analyses between gene sets. Differences in proportions were assessed by hypergeometric test, and p-values adjusted by Bonferroni correction. The asterisks indicate proportions in SG-expressed or SG-up-regulated effector sets that were significantly different from those in whole gene set. \* $p < 0.05$ . Blast2GO analysis conducted with the *A. pisum* global gene set revealed 13,017 genes with at least one GO term annotation (69.98%) and 5584 genes without any known function (30.02%). In the SG-up-regulated effector candidate set, 60.14% of the genes (445 genes) were annotated by at least one GO term and 39.86% (295 genes) remained unknown, showing significantly lower annotation rate compared to the whole gene set. The SG-expressed effector candidate set had similar annotation rate as the global gene set with 71.5% of genes with annotated functions. Most of the enriched GO terms were the same for the 2 effector sets compared to whole gene set. However, enrichments were stronger for the SG-up-regulated effector candidate set. Among the functions that stand out, proteolysis, metallopeptidase activity and DNA-dependent ATPase activity represented respectively 13.93%, 6.29% and 4.04% of SG-up-regulated candidate effectors having GO terms annotations.



**Supplementary fig. S4.** Sites under positive selection in specific branches of the *CRP* gene family tree. (A) Plots of the BEB posterior probability of each codon to be under positive selection in each significant branch. (B) Translation of the codon-based alignment used to compute dN/dS among Aphididae *CRP* sequences. The conserved cysteine residues are highlighted in dark grey, the light grey and red indicate amino acids encoded by codons with BEB probability to be under positive selection above 0.75 and 0.95 respectively.

A



B

```

APIS ACE10 -----
APIS ACE1  MKSLIHRGCSHWVWYRRLKLSRF-----KLYLHLCVILCLSATVVRPVLSSAGIEDSSHSSEIASEASKYLENANEMMAQWANQADNNWATLTDENAA
APIS ACE5  -----KLYLHLCVILCLSATVVRPVLSSAGIEDSSHSSEIASEASKYLENANEMMAQWANQADNNWATLTDENAA
APIS ACE4  -----TMKLYHLCVILCLLATVVRPVLSSAGIKNDSHSSEIASEASKYLENANFALQWTRNANNNMTLNTLTDENAE
AGLY ACE   -----MIKLYCLYAILWLSVTVVRPVLSSAGTKK--SEIASEASKYLDNANFALQWTRNANNNMTLNTLTDENAE
DNOX ACE   -----
MPER ACE   -----TMKLYHLCVILCLSSSTVWQPVLSSAGIKKCTYSEAEASKYLEKANVLELQWTRNANNNMTLNTLTDENAE

APIS ACE10 -----
APIS ACE1  KKLATSLSLEYSKFSKCMWEETVKYPWTTYKDPNKRQPKLMSDITGTDALPEDKVKKLEETIDAMQSLYDKRATIPYCGDTSNRRLLSFPEDITDILANSTDV
APIS ACE5  KKLATSLSLEYSKFSKCMWEETVKYPWTTYKDPNKRQPKLMSVIGTDALPEDKVKKLEETIDAMQSLYDKRATIPYCGDTSNRRLLSFPEDITDILANSTDV
APIS ACE4  KKLAIANLLEFSKFKCMWEETVKYPWTTYKDPNKRQPKLMSVIGTDALPEDKVKKLEETIDAMQSLYDKRATIPYCGDTSNRRLLSFPEDITDILANSTDV
AGLY ACE   KKLAIANLLEFSKFKLMSWVETVKYPWTTYKDPNKRQPKLMSVIGTDALPEDKVKKLEETIDAMQSLYDKRATIPYCGDTSNRRLLSFPEDITDILANSTDV
DNOX ACE   -----MWEETVKYPWSTYRKN-DIKRQPKLMSVIGTDALPEDKVKKLEETIDAMQSLYDKRATIPYCGDTSNRRLLSFPEDITDILANSTDV
MPER ACE   KKLAIANLLEFSKFKRMWEETVKYPWSTYRKN-DIKRQPKLMSVIGTDALPEDKVKKLEETIDAMQSLYDKRATIPYCGDTSNRRLLSFPEDITDILANSTDV

APIS ACE10 -----
APIS ACE1  NELTYMVMQWREATGKKIRPLYPEYVKLNNEAARLNNTDTEAYWIRDYEVDDFRQRIECLWQIKFPLYLQIHAYVRSKLRREIYGESLITRKRGLIPAQLL
APIS ACE5  NELTYMVMQWREATGKKIRPLYPEYVKLNNEAARLNNTDTEAYWIRDYEVDDFRQRIECLWQIKFPLYLQIHAYVRSKLRREIYGESLITRKRGLIPAQLL
APIS ACE4  NELKRVVMQWREATGKKIRPLYPEYVKLNNEAARLNNTDTEAYWIRDYEVDDFRQRIECLWQIKFPLYLQIHAYVRSKLRREIYGESLITRKRGLIPAQLL
AGLY ACE   NELKRVVMQWREATGKKIRPLYPEYVKLNNEAARLNNTDTEAYWIRDYEVDDFRQRIECLWQIKFPLYLQIHAYVRSKLRREIYGESLITRKRGLIPAQLL
DNOX ACE   NELKRVVMQWREATGKKIRPLYPEYVKLNNEAARLNNTDTEAYWIRDYEVDDFRQRIECLWQIKFPLYLQIHAYVRSKLRREIYGESLITRKRGLIPAQLL
MPER ACE   NELKRVVMQWREATGKKIRPLYPEYVKLNNEAARLNNTDTEAYWIRDYEVDDFRQRIECLWQIKFPLYLQIHAYVRSKLRREIYGESLITRKRGLIPAQLL

APIS ACE10 -----
APIS ACE1  GNMWAQSWENLNDTRPYPTDDVNPPTSAMINQYTPKMKFMAEEFYTSNLSMPTFWERSVIEKPNGVDMCECFASAWDFSDSKDFRIKECTSVNFDIF
APIS ACE5  GNMWAQSWENLNDTRPYPTDDVNPPTSAMINQYTPKMKFMAEEFYTSNLSMPTFWERSVIEKPNGVDMCECFASAWDFSDSKDFRIKECTSVNFDIF
APIS ACE4  GNMWAQSWERLDDFTRPYPTDDVNPPTSAMINQYTPKMKFMAEEFYTSNLSMPTFWERSVIEKPNGRDLVCHASAWDFSDSKDFRIKECTSVNFDIF
AGLY ACE   GNMWAQSWERLDDFTRPYPTDDVNPPTSAMINQYTPKMKFMAEEFYTSNLSMPTFWERSVIEKPNGRDLVCHASAWDFSDSKDFRIKECTSVNFDIF
DNOX ACE   GNMWAQSWERLDDFTRPYPTDDVNPPTSAMINQYTPKMKFMAEEFYTSNLSMPTFWERSVIEKPNGRDLVCHASAWDFSDSKDFRIKECTSVNFDIF
MPER ACE   GNMWAQSWERLDDFTRPYPTDDVNPPTSAMINQYTPKMKFMAEEFYTSNLSMPTFWERSVIEKPNGRDLVCHASAWDFSDSKDFRIKECTSVNFDIF

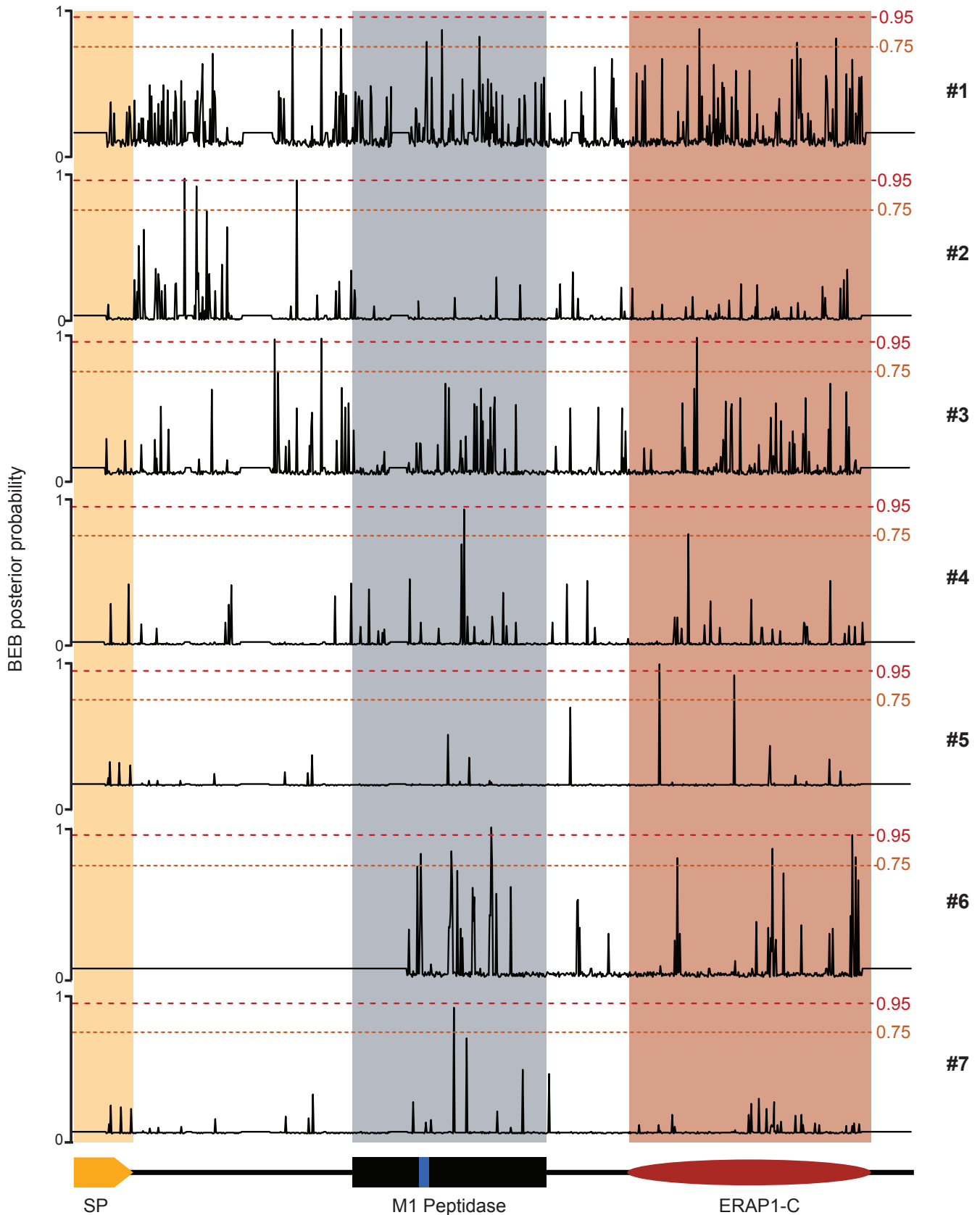
APIS ACE10 -----
APIS ACE1  ITLHEHMGQIEEYMQYANQPYIYHGGANFHEAIGDITALSVSSTKHLNKGILGFLPNTSRSCAEKIGLQKIAFLPFGYLMDLWRNVDVFKGITTEQYNCMDM
APIS ACE5  ITLHEHMGQIEEYMQYANQPYIYHGGAN-----
APIS ACE4  ITLHEHMGQIQYELQYKQPFIRYREGANFHEAIGDITALSVSSTKHLNKGILGFLPNTSRSCAEKIGLQKIAFLPFGYLMDLWRNVDVFKGITTEQYNCMDM
AGLY ACE   ITLHEHMGQIQYELQYKQPFIRYREGANFHEAIGDITALSVSSTKHLNKGILGFLPNTSRSCAEKIGLQKIAFLPFGYLMDLWRNVDVFKGITTEQYNCMDM
DNOX ACE   ITLHEHMGQIQYELQYKQPFIRYREGANFHEAIGDITALSVSSTKHLNKGILGFLPNTSRSCAEKIGLQKIAFLPFGYLMDLWRNVDVFKGITTEQYNCMDM
MPER ACE   ITLHEHMGQIQYELQYKQPFIRYREGANFHEAIGDITALSVSSTKHLNKGILGFLPNTSRSCAEKIGLQKIAFLPFGYLMDLWRNVDVFKGITTEQYNCMDM

APIS ACE10 -----
APIS ACE1  KLRCSYQGIIEPPVTRTEEDFDPGSKFHIIVADVPYQFQHQALCEKAGQDFDKNPSQPLHLCDDIYQSTDAGNALKDMCLKGSSKPFDFLLTGQREMDAGP
APIS ACE5  KLRCSYQGIIEPPVTRTEEDFDPGSKFHIIVADVPYQFQHQALCEKAGQDFDKNPSQPLHLCDDIYQSTDAGNALKDMCLKGSSKPFDFLLTGQREMDAGP
APIS ACE4  KLRCSYQGIIEPPVTRTEEDFDPGSKFHIIVGPNVPYQFQHQALCEKAGQDFDKNPSQPLHLCDDIYQSTDAGNALKDMCLKGSSKPFDFLLTGQREMDAGP
AGLY ACE   KLRCSYQGIIEPPVTRTEEDFDPGSKFHIIVGPNVPYQFQHQALCEKAGQDFDKNPSQPLHLCDDIYQSTDAGNALKDMCLKGSSKPFDFLLTGQREMDAGP
DNOX ACE   KLRCSYQGIIEPPVTRTEEDFDPGSKFHIIVGPNVPYQFQHQALCEKAGQDFDKNPSQPLHLCDDIYQSTDAGNALKDMCLKGSSKPFDFLLTGQREMDAGP
MPER ACE   KLRCSYQGIIEPPVTRTEEDFDPGSKFHIIVGPNVPYQFQHQALCEKAGQDFDKNPSQPLHLCDDIYQSTDAGNALKDMCLKGSSKPFDFLLTGQREMDAGP

APIS ACE10 -----
APIS ACE1  LLNYFNPLYEWLKKNENERTGEYIGWETNKKVCLKKGETSQP
APIS ACE5  LLNYFNPLYEWLKKNENERTGEYIGWETNKKVCLKKGETSQP
APIS ACE4  LLNYFNPLYEWLKKNENERTGEYIGWETNKKVCLKKGETSQP
AGLY ACE   --NY-----
DNOX ACE   LLNYFNPLYEWLQENKERTGERLGMWETNKKVCLKKGETSQP
MPER ACE   LLNYFNPLYEWLQENKERTGERLGMWETNKKVCLKKGETSQP

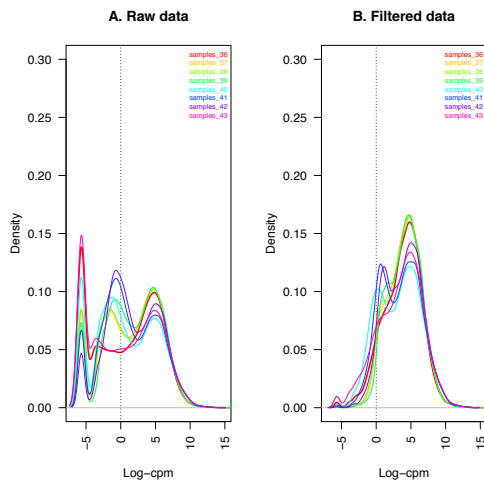
```

**Supplementary fig. S5.** Sites under positive selection in specific branches of the clade 4 of the *ACE* gene family tree. (A) Plots of the BEB posterior probability of each codon to be under positive selection in each significant branch. (B) Translation of the codon-based alignment used to compute dN/dS among Aphididae *ACE* sequences. The peptidase active site HEXXH is highlighted in blue, grey and red indicate amino acids encoded by codons with BEB probability to be under positive selection above 0.75 and 0.95 respectively.

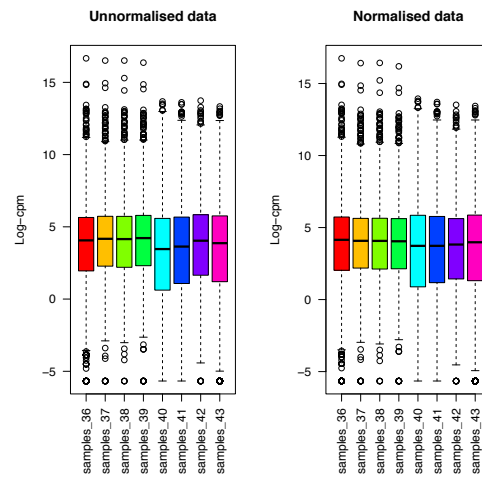


**Supplementary fig. S7.** Sites under positive selection in specific branches of the clade 4 of the *apN* gene family tree. Plots of the BEB posterior probability of each codon to be under positive selection in each significant branch. Numbers on the right indicate the branch numbers relative to the fig. 8b. The peptidase active site HEXXH is highlighted in blue. Grey and red squares indicate the amino acids encoded by the codons to be under positive selection according to the BEB probability (above 0.75 and 0.95, respectively).

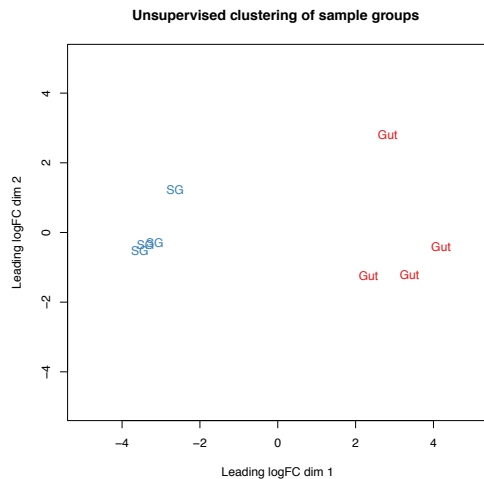
A



B



C



**Supplementary fig. S8.** (A) Effect of expressed gene filtering on genes density. Density plot of log-cpm of all genes (Raw data) and after removing genes with a CPM < 1 in at least 3 libraries on the 8 analyzed libraries (Filtered data). The essential part of the unexpressed genes was removed after applying the filter, ensuring efficient differentially expressed gene analyses. (B) Effect of cpm normalization on gene expression profiles. Boxplots representing the expression (log-cpm) distribution of the expressed genes (filtered) before and after cpm normalization using TMM method for Normalization Factor calculation. After normalization, the distribution of gene expression of the 8 samples was similar. Samples 36 to 39 were the SG libraries and samples 40 to 43 corresponded to gut libraries. (C) Assessment of the biological replicates reproducibility. MDS plot of log-cpm values over dimensions 1 and 2 with in SG samples in blue and gut samples in red. The unsupervised clustering of sample groups has allowed to verify the repeatability of the biological replicates, no outlier was detected and thus, all the samples were kept for differentially expressed genes analysis.



For better viewing, the following tables and figures are available online at:  
[https://www.dropbox.com/sh/cft2z2z8btywmpe/AACbMn-ppgVl4aUu66\\_Ptd\\_pa?dl=0](https://www.dropbox.com/sh/cft2z2z8btywmpe/AACbMn-ppgVl4aUu66_Ptd_pa?dl=0)

**Supplementary table S1.** Characteristics, expression levels, secretion predictions, proteomic detection and evolutionary rates of the 18601 *A. pisum* genes analyzed in this study.

**Supplementary table S2.** The 51 proteins identified with high confidence in *A. pisum* LSR1 saliva. Protein abundances in individual replicates are represented by log<sub>2</sub> transformed intensities. Summed intensities, scores, overall number of peptides and mass spectrum count were combined and given for the three replicates. NaN: Not a number, indicating that peptides were not detected in a specific sample. The combined label free quantitative log<sub>2</sub> normalized values for all replicates were determined and used to represent relative abundances for comparison to transcript counts

**Supplementary table S3.** The 1837 proteins identified with high confidence in *A. pisum* LSR1 salivary glands. Protein abundances in individual replicates are represented by log<sub>2</sub> transformed intensities. Summed intensities, scores, overall number of peptides and mass spectrum count were combined and given for the three replicates. NaN: Not a number, indicating that peptides were not detected in a specific sample. The combined label free quantitative log<sub>2</sub> normalized values for all replicates were determined and used to represent relative abundances for comparison to transcript counts.

**Supplementary table S4.** Details of the Aphididae single copy effector genes used to compute dN/dS.

**Supplementary table S5.** Estimated evolutionary rates (dN/dS or  $\omega$ ) for aphid characterized effector genes in Aphididae species comparisons. The null models (M0=one ratio, M1=neutral, M7= $\beta$ ) were compared to alternative models (M2=selection, M8= $\beta + \omega$ ) using the likelihood ratio test (LRT). The stars indicate significance of the p-values obtained by confrontation of LRT with  $\chi^2$  distribution.

**Supplementary table S6.** CRP, ACE and apN *A. pisum* gene family characteristics.

**Supplementary table S7.** Arthropod and insect genomes used to compute orthologue groups.

**Supplementary fig. S6.** Sites under positive selection in specific branches of the clade 4 of the *Aminopeptidase-N* gene family tree. Plots of the BEB posterior probability of each codon to be under positive selection in each significant branch.

## Supplementary material and methods S1.

### 454 sequencing of *Rhopalosiphon padi* and *Schizaphis graminum* salivary gland transcriptomes.

*R. padi* and *S. graminum* were reared in Kansas State University (USA). Approximately 300 salivary glands from adult aphids were dissected in saline, and total RNA was extracted. 500ng of total RNA was used for the first-strand cDNA synthesis using SMARTer PCR cDNA synthesis kit (Clontech) according to Clontech protocol except using the primer 5'-AAG CAG TGG TAT CAA CGC AGA GTA CTT TTT CTT TTT CTT TTT GTT TTT VN-3'. Double-stranded cDNA was amplified by long-distance (LD) PCR using Advantage 2 PCR kit (Clontech). Double stranded full length cDNA was used to prepare 454 sequencing libraries using the GS FLX Titanium general Library Preparation kit (Roche). The obtained libraries were sequenced with the Titanium chemistry on the GS FLX System followed standard Roche protocols. The 454 sequencing was conducted at the KSU Integrated Genomics Facility. Sequencing runs produced 541513 and 644071 reads for *R. padi* and *S. graminum* libraries, respectively. The reads were cleaned and assembled by Newbler version 2.6 resulting in 6743 and 8529 isotigs, respectively. Consensus contigs were created from all the isotigs that belong to the same isogroup using CAP3 program (Huang and Madan 1999). *S. graminum* library was filtered for the pea aphid contamination. As a result, 6463 and 7981 isoconsensus were created from *R. padi* and *S. graminum* libraries, respectively. The trimmed reads of each aphid sample were deposited to NCBI with the accession numbers SRR6118147 for *R. padi* and SRR6118148 for *S. graminis*. The consensus isogroups are available at:

[http://bipaa.genouest.org/sp/rhopalosiphum\\_padi/download/transcriptome/SG/BCO.fa](http://bipaa.genouest.org/sp/rhopalosiphum_padi/download/transcriptome/SG/BCO.fa) and  
[http://bipaa.genouest.org/data/public/s\\_graminum/Transcriptome/SG/GB.fa](http://bipaa.genouest.org/data/public/s_graminum/Transcriptome/SG/GB.fa)

## REFERENCES

- Abdellatef E, Will T, Koch A, Imani J, Vilcinskas A, Kogel K-H. 2015. Silencing the expression of the salivary sheath protein causes transgenerational feeding suppression in the aphid *Sitobion avenae*. *Plant Biotechnol. J.* 13:849–857.
- Anisimova M, Yang Z. 2007. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Mol. Biol. Evol.* 24:1219–1228.
- Atamian HS, Chaudhary R, Cin VD, Bao E, Girke T, Kaloshian I. 2012. *In planta* expression or delivery of potato aphid *Macrosiphum euphorbiae* effectors Me10 and Me23 enhances Aphid fecundity. *Mol. Plant. Microbe Interact.* 26:67–74.
- Banzet N, Latorse M-P, Bulet P, François E, Derpierre C, Dubald M. 2002. Expression of insect cystein-rich antifungal peptides in transgenic tobacco enhances resistance to a fungal disease. *Plant Sci.* 162:995–1006.
- Blackman RL, Eastop VF. 1994. *Aphids on the world's trees: an identification and information guide.* Cab International
- Blackman RL, Eastop VF. 2000. *Aphids on the world's crops: an identification and information guide.* John Wiley & Sons Ltd
- Bos JIB, Prince D, Pitino M, Maffei ME, Win J, Hogenhout SA. 2010. A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (Green Peach Aphid). *PLOS Genet.* 6:e1001216.
- Carolan JC, Caragea D, Reardon KT, Mutti NS, Dittmer N, Pappan K, Cui F, Castaneto M, Poulain J, Dossat C, et al. 2011. Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrtosiphon pisum*): a dual transcriptomic/proteomic approach. *J. Proteome Res.* 10:1505–1518.
- Carolan JC, Fitzroy CIJ, Ashton PD, Douglas AE, Wilkinson TL. 2009. The secreted salivary proteome of the pea aphid *Acyrtosiphon pisum* characterised by mass spectrometry. *PROTEOMICS* 9:2457–2467.
- Chaudhary R, Atamian HS, Shen Z, Briggs SP, Kaloshian I. 2014. GroEL from the endosymbiont *Buchnera aphidicola* betrays the aphid by triggering plant defense. *Proc. Natl. Acad. Sci.* 111:8919–8924.
- Chaudhary R, Atamian HS, Shen Z, Briggs SP, Kaloshian I. 2015. Potato aphid salivary proteome: enhanced salivation using resorcinol and identification of aphid phosphoproteins. *J. Proteome Res.* 14:1762–1778.
- Conant GC, Wolfe KH. 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet.* 9:938–950.
- Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. 2011. Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10:1794–1805.

- Cristofolletti PT, Ribeiro AF, Deraison C, Rahbé Y, Terra WR. 2003. Midgut adaptation and digestive enzyme distribution in a phloem feeding insect, the pea aphid *Acyrtosiphon pisum*. *J. Insect Physiol.* 49:11–24.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
- Drès M, Mallet J. 2002. Host races in plant–feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357:471–492.
- Duvaux L, Geissmann Q, Gharbi K, Zhou J-J, Ferrari J, Smadja CM, Butlin RK. 2015. Dynamics of copy number variation in host races of the pea aphid. *Mol. Biol. Evol.* 32:63–80.
- Eastop V. 1971. Keys for the identification of *Acyrtosiphon* (Hemiptera: Aphididae)
- Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci.* 112:8362–8366.
- Ehrlich PR, Raven PH. 1964. Butterflies and plants: a study in coevolution. *Evolution* 18:586–608.
- Elzinga DA, De Vos M, Jander G. 2014. Suppression of plant defenses by a *Myzus persicae* (Green Peach Aphid) salivary effector protein. *Mol. Plant. Microbe Interact.* 27:747–756.
- Elzinga DA, Jander G. 2013. The role of protein effectors in plant–aphid interactions. *Curr. Opin. Plant Biol.* 16:451–456.
- Eyres I, Duvaux L, Gharbi K, Tucker R, Hopkins D, Simon J-C, Ferrari J, Smadja CM, Butlin RK. 2017. Targeted re-sequencing confirms the importance of chemosensory genes in aphid host race differentiation. *Mol. Ecol.* 26:43–58.
- Eyres I, Jaquiéry J, Sugio A, Duvaux L, Gharbi K, Zhou J-J, Legeai F, Nelson M, Simon J-C, Smadja CM, et al. 2016. Differential gene expression according to race and host plant in the pea aphid. *Mol. Ecol.* 25:4197–4215.
- Forbes AA, Devine SN, Hippee AC, Tvedte ES, Ward AK, Widmayer HA, Wilson CJ. 2017. Revisiting the particular role of host shifts in initiating insect speciation. *Evolution* 71:1126–1137.
- Fraenkel GS. 1959. The Raison d’Être of secondary plant substances. *Science* 129:1466–1470.
- Furch ACU, Bel V, E AJ, Will T. 2015. Aphid salivary proteases are capable of degrading sieve-tube proteins. *J. Exp. Bot.* 66:533–539.
- Gaupels F, Furch ACU, Will T, Mur LAJ, Kogel K-H, Van Bel AJE. 2008. Nitric oxide generation in *Vicia faba* phloem cells reveals them to be sensitive detectors as well as possible systemic transducers of stress signals. *New Phytol.* 178:634–646.
- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36:3420–3435.

- Guo K, Wang W, Luo L, Chen J, Guo Y, Cui F. 2014. Characterization of an aphid-specific, cysteine-rich protein enriched in salivary glands. *Biophys. Chem.* 189:25–32.
- Guy E, Boulain H, Aigu Y, Le Pennec C, Chawki K, Morlière S, Schädel K, Kunert G, Simon J-C, Sugio A. 2016. Optimization of agroinfiltration in *Pisum sativum* provides a new tool for studying the salivary protein functions in the pea aphid complex. *Front. Plant Sci.* 7.
- Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461:393–398.
- Harris KF, Maramorosch K. 2014. *Aphids as virus vectors*. Elsevier
- Harrison C, Acharya KR. 2014. ACE for all – a molecular perspective. *J. Cell Commun. Signal.* 8:195–210.
- Hogenhout SA, Bos JJ. 2011. Effector proteins that modulate plant–insect interactions. *Curr. Opin. Plant Biol.* 14:422–428.
- Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9:868–877.
- Hubner NC, Bird AW, Cox J, Spletstoesser B, Bandilla P, Poser I, Hyman A, Mann M. 2010. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J. Cell Biol.* 189:739–754.
- Janz N. 2011. Ehrlich and Raven Revisited: Mechanisms underlying codiversification of plants and enemies. *Annu. Rev. Ecol. Evol. Syst.* 42:71–89.
- Jaquiéry J, Stoeckel S, Nouhaud P, Mieuze L, Mahéo F, Legeai F, Bernard N, Bonvoisin A, Vitalis R, Simon J-C. 2012. Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex. *Mol. Ecol.* 21:5251–5264.
- Jaubert-Possamai S, Rispe C, Tanguy S, Gordon K, Walsh T, Edwards O, Tagu D. 2010. Expansion of the miRNA pathway in the Hemipteran insect *Acyrtosiphon pisum*. *Mol. Biol. Evol.* 27:979–987.
- Jiang RHY, Tripathy S, Govers F, Tyler BM. 2008. RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proc. Natl. Acad. Sci.* 105:4874–4879.
- Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33:511–518.
- Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc B* 279:5048–5057.
- Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM. 2015. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.* 43:D250–D256.
- Law CW, Alhamdoosh M, Su S, Smyth GK, Ritchie ME. 2016. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* 5:1408.

- Law CW, Chen Y, Shi W, Smyth GK. 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15:R29.
- Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128.
- Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930.
- Linz LB, Liu S, Chougule NP, Bonning BC. 2015. In vitro evidence supports membrane alanyl aminopeptidase N as a receptor for a plant virus in the pea aphid vector. *J. Virol.* 89:11203–11212.
- Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. U. S. A.* 102:10557–10562.
- Lu H, Tanguy S, Rispe C, Gauthier J-P, Walsh T, Gordon K, Edwards O, Tagu D, Chang C, Jaubert-Possamai S. 2011. Expansion of genes encoding piRNA-associated argonaute proteins in the pea aphid: diversification of expression profiles in different plastic morphs. *PLOS ONE* 6:e28051.
- Mathers TC, Chen Y, Kaithakottil G, Legeai F, Mugford ST, Baa-Puyoulet P, Bretaudeau A, Clavijo B, Colella S, Collin O, et al. 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol.* 18.
- Mayhew PJ. 2007. Why are there so many insect species? Perspectives from fossils and phylogenies. *Biol. Rev.* 82:425–454.
- McCarthy DJ, Smyth GK. 2009. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* 25:765–771.
- Miles PW. 1972. The saliva of Hemiptera. *Adv. Insect Physiol.* 9:183–255.
- Mitter C, Farrell B, Wiegmann B. 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *Am. Nat.* 132:107–128.
- Mutti NS, Louis J, Pappan LK, Pappan K, Begum K, Chen M-S, Park Y, Dittmer N, Marshall J, Reese JC, et al. 2008. A protein from the salivary glands of the pea aphid, *Acyrtosiphon pisum*, is essential in feeding on a host plant. *Proc. Natl. Acad. Sci.* 105:9965–9969.
- Mutti NS, Park Y, Reese JC, Reeck GR. 2006. RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrtosiphon pisum*. *J. Insect Sci.* 6:1–7.
- Naessens E, Dubreuil G, Giordanengo P, Baron OL, Minet-Kebdani N, Keller H, Coustau C. 2015. A secreted MIF cytokine enables aphid feeding and represses plant immune responses. *Curr. Biol. CB* 25:1898–1903.
- Nicholson SJ, Nickerson ML, Dean M, Song Y, Hoyt PR, Rhee H, Kim C, Puterka GJ. 2015. The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC Genomics* 16:429.
- Nielsen R, Yang Z. 1998. Likelihood Models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936.

- Oliva RF, Cano LM, Raffaele S, Win J, Bozkurt TO, Belhaj K, Oh S-K, Thines M, Kamoun S. 2015. A recent expansion of the RXLR effector gene Avrblb2 is maintained in global populations of *Phytophthora infestans* indicating different contributions to virulence. *Mol. Plant-Microbe Interact.* 28:901–912.
- Pan Y, Zhu J, Luo L, Kang L, Cui F. 2015. High expression of a unique aphid protein in the salivary glands of *Acyrtosiphon pisum*. *Physiol. Mol. Plant Pathol.* 92:175–180.
- Peccoud J, Bonhomme J, Mahéo F, de la Huerta M, Cosson O, Simon J-C. 2014. Inheritance patterns of secondary symbionts during sexual reproduction of pea aphid biotypes. *Insect Sci.* 21:291–300.
- Peccoud J, Ollivier A, Plantegenest M, Simon J-C. 2009a. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc. Natl. Acad. Sci.* 106:7495–7500.
- Peccoud J, Simon J-C. 2010. The pea aphid complex as a model of ecological speciation. *Ecol. Entomol.* 35:119–130.
- Peccoud J, Simon J-C, von Dohlen C, Coeur d’acier A, Plantegenest M, Vanlerberghe-Masutti F, Jouselin E. 2010. Evolutionary history of aphid-plant associations and their role in aphid diversification. *C. R. Biol.* 333:474–487.
- Peccoud J, Simon J-C, McLaughlin HJ, Moran NA. 2009b. Post-Pleistocene radiation of the pea aphid complex revealed by rapidly evolving endosymbionts. *Proc. Natl. Acad. Sci.* 106:16315–16320.
- Pitino M, Coleman AD, Maffei ME, Ridout CJ, Hogenhout SA. 2011. Silencing of aphid genes by dsRNA feeding from plants. *PLOS ONE* 6:e25709.
- Pitino M, Hogenhout SA. 2012. Aphid protein effectors promote aphid colonization in a plant species-specific manner. *Mol. Plant. Microbe Interact.* 26:130–139.
- Pohlert T, Pohlert MT. 2016. Package ‘PMCMR.’
- Powell G, Tosh CR, Hardie J. 2006. Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. *Annu Rev Entomol* 51:309–330.
- Powell THQ, Hood GR, Murphy MO, Heilveil JS, Berlocher SH, Nosil P, Feder JL. 2013. Genetic divergence along the speciation continuum: the transition from host race to species in *Rhagoletis* (diptera: Tephritidae). *Evolution* 67:2561–2576.
- Price DRG, Duncan RP, Shigenobu S, Wilson ACC. 2011. Genome expansion and differential expression of amino acid transporters at the aphid/*Buchnera* symbiotic interface. *Mol. Biol. Evol.* 28:3113–3126.
- Qi M, Link TI, Müller M, Hirschburger D, Pudake RN, Pedley KF, Braun E, Voegelé RT, Baum TJ, Whitham SA. 2016. A small cysteine-rich protein from the asian soybean rust fungus, *Phakopsora pachyrhizi*, suppresses plant immunity. *PLoS Pathog.* 12.

- Rahbé Y, Sauvion N, Febvay G, Peumans WJ, Gatehouse AMR. 1995. Toxicity of lectins and processing of ingested proteins in the pea aphid *Acyrtosiphon pisum*. *Entomol. Exp. Appl.* 76:143–155.
- Remenant B, Coupat-Goutaland B, Guidot A, Cellier G, Wicker E, Allen C, Fegan M, Pruvost O, Elbaz M, Calteau A, et al. 2010. Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. *BMC Genomics* 11:379.
- Rispe C, Kutsukake M, Doublet V, Hudaverdian S, Legeai F, Simon J-C, Tagu D, Fukatsu T. 2007. Large gene family expansion and variable selective pressures for cathepsin B in aphids. *Mol. Biol. Evol.* 25:5–17.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47–e47.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
- Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25.
- Rodriguez PA, Escudero-Martinez C, Bos JIB. 2017. An aphid effector targets trafficking protein VPS52 in a host-specific manner to promote virulence. *Plant Physiol.* 173:1892–1903.
- Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* 43:W7–W14.
- Smadja C, Shi P, Butlin RK, Robertson HM. 2009. Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Mol. Biol. Evol.* 26:2073–2086.
- Smadja CM, Canbäck B, Vitalis R, Gautier M, Ferrari J, Zhou J-J, Butlin RK. 2012. Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution* 66:2723–2738.
- Smyth GK. 2004. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat. Appl. Genet. Mol. Biol.* 3:1–25.
- Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T, et al. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* 344:738–742.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stamatakis A, Hoover P, Rougemont J, Renner S. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* 57:758–771.



- Stergiopoulos I, Kourmpetis YAI, Slot JC, Bakker FT, De Wit PJGM, Rokas A. 2012. In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. *Mol. Biol. Evol.* 29:3371–3384.
- Team RC. 2017. R: A language and environment for statistical computing.
- The International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8:e1000313.
- Thorpe P, Cock PJ, Bos J. 2016. Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets. *BMC Genomics* 17:172.
- Van Valen L. 1977. The Red Queen. *Am. Nat.* 111:809–810.
- Vertacnik KL, Linnen CR. 2017. Evolutionary genetics of host shifts in herbivorous insects: insights from the age of genomics. *Ann. N. Y. Acad. Sci.* 1389:186–212.
- Wang W, Dai H, Zhang Y, Chandrasekar R, Luo L, Hiromasa Y, Sheng C, Peng G, Chen S, Tomich JM, et al. 2015a. Armet is an effector protein mediating aphid–plant interactions. *FASEB J.* 29:2032–2045.
- Wang W, Luo L, Lu H, Chen S, Kang L, Cui F. 2015b. Angiotensin-converting enzymes modulate aphid–plant interactions. *Sci. Rep.* 5.
- Wenger JA, Cassone BJ, Legeai F, Johnston JS, Bansal R, Yates AD, Coates BS, Pavinato VAC, Michel A. 2017. Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect Biochem. Mol. Biol.*
- Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. 2007. The genetic basis of a plant–insect coevolutionary key innovation. *Proc. Natl. Acad. Sci.* 104:20427–20431.
- Wiens JJ, Lapoint RT, Whiteman NK. 2015. Herbivory increases diversification across insect clades. *Nat. Commun.* 6:ncomms9370.
- Will T, Kornemann SR, Furch ACU, Tjallingii WF, Bel AJE van. 2009. Aphid watery saliva counteracts sieve-tube occlusion: a universal phenomenon? *J. Exp. Biol.* 212:3305–3312.
- Will T, Vilcinskis A. 2015. The structural sheath protein of aphids is required for phloem feeding. *Insect Biochem. Mol. Biol.* 57:34–40.
- Win J, Morgan W, Bos J, Krasileva KV, Cano LM, Chaparro-Garcia A, Ammar R, Staskawicz BJ, Kamoun S. 2007. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell Online* 19:2349–2369.
- Wu J, Baldwin IT. 2010. New insights into plant responses to the attack from insect herbivores. *Annu. Rev. Genet.* 44:1–24.
- Yamazaki Y, Morita T. 2004. Structure and function of snake venom cysteine-rich secretory proteins. *Toxicon* 44:227–231.

- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* 13:555–556.
- Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24:1586–1591.
- Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15:496–503.
- Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17:32–43.
- Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19:908–917.
- Yang Z, Nielsen R, Goldman N, Pedersen A-MK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.
- Yang Z, Wong WSW, Nielsen R. 2005. Bayes Empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22:1107–1118.
- Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18:292–298.
- Zhang Y, Fan J, Sun J, Chen J. 2015. Cloning and RNA interference analysis of the salivary protein C002 gene in *Schizaphis graminum*. *J. Integr. Agric.* 14:698–705.
- Zhao C, Escalante LN, Chen H, Benatti TR, Qu J, Chellapilla S, Waterhouse RM, Wheeler D, Andersson MN, Bao R, et al. 2015. A massive expansion of effector genes underlies gall-formation in the wheat pest *Mayetiola destructor*. *Curr. Biol.* 25:613–620.



### **III. Functional characterization of candidate salivary genes that are potentially involved in the pea aphid-plant interactions**

#### **1. Context**

Our analyses of evolutionary patterns of *Acyrtosiphon pisum* salivary genes provided promising effector candidates involved in the adaptation of *A. pisum* to the legume plant family (chapter II). Some fast-evolving genes and members of expanded salivary gene families may have evolved during the interactions of *A. pisum* with its host plant range and potentially driven or played a role in specialization to the Fabaceae family. The next step is therefore to better understand their roles in the adaptation to different legume species and biotype formation. For that, past (Jaquiéry et al. 2012; Nouhaud et al. 2014), ongoing (Nouhaud et al. 2017 submitted) and future genetic studies may help to identify effector genes that are divergent between aphid species and biotypes and potentially involved in plant specialization. Then, functional characterization studies targeting these specific effector candidates will be needed. Even if population genomic works may allow to refine the candidate list, the main challenge now lies in how to characterize the role of these candidate effectors in a pea aphid-host plant system. Indeed, functional characterization of effectors is essential to (i) validate positive (or negative) contribution of effectors to aphid fitness, (ii) reveal the molecular dialogue between the effectors and host plants, and (iii) understand the molecular mechanisms underlying host and non-host interactions between the *A. pisum* biotypes and the legume species.

To examine the effector functions, one may modify the parasites or their host plants. Various microbial pathogens are amenable to genetic transformation. In such a case, effector knock-out mutants can be created to examine the importance of the effectors in the parasitism. Complementation of the KO mutants by the effectors or effector mutants can verify the results and identify the effector domains that are important for their functions. Aphids are not transformable yet. However, development of EMS mutagenesis (Tagu et al. 2014) and CRISPR/Cas9 system (Jinek et al. 2012) is in progress in our team. Major issues of these techniques are a high time investment and the impossibility to mutate a gene essential for aphid survival. The gene silencing of aphids by dsRNA injection was previously established to study aphid effector functions (Mutti et al. 2006; Mutti et al. 2008; Guo et al. 2014; Naessens et al. 2015; Wang et al. 2015a; Wang et al. 2015b) and other gene functions (Jaubert-Possamai et al. 2007), but unfortunately, it does not work in our hands and we never reproduced any reported phenotype. Despite multiple tests of different concentrations of dsRNAs or siRNAs targeting

the *C002* gene (Mutti et al. 2006; Mutti et al. 2008), qPCR never detected a reduction of gene expression, and aphids never showed reduction in survival or fecundity like previous studies described (Mutti et al. 2006; Mutti et al. 2008; Naessens et al. 2015; Wang et al. 2015a; Wang et al. 2015b). Injuries caused to the aphids during the injection process, qPCR procedure (e.g. reference gene used), and the use of different aphid clones (between our lab and other labs) may explain the difficulties we encountered. Since it is still difficult to manipulate our aphids, we decided to develop a way to manipulate plant to study aphid effector genes.

Many techniques used to manipulate plants and study effector functions have been developed by using plants pathogens, such as *Agrobacterium tumefaciens*, *Pseudomonas syringae* or phytoviruses. *A. tumefaciens* DNA transmission capabilities (Schell and Montagu 1977) have been extensively used to express the gene of interest in plants. The genes are cloned in the expression cassette inserted between T-DNA left and right borders (as examples in Karimi et al. 2002, Sainsbury et al. 2009) of a binary vector and then introduced in a disarmed *A. tumefaciens* harboring a helper plasmid that contains the genes required for T-DNA transfer. Once the *A. tumefaciens* infects plants, the T-DNA reaches the nuclei and can be integrated in plant genome. Then, plant specific promoters contained in the expression cassettes ensure the expression of transgenes (mature protein or dsRNA structure for silencing) in plant cells. The *Agrobacterium* strains can be infiltrated directly in leaves of various plants. In this case, multiple T-DNA copies are transferred to the nuclei and can persist for a period of time, providing a transient expression of the transgene restricted to the infiltrated zone (Janssen and Gardner 1990). Transient expression by agroinfiltration can be achieved quickly compared to creating stable transgenic plants; therefore, it allows to screen the functions of many effector candidate genes, including those which are toxic for the plant (e.g. avirulence genes), in a short period of time. However, the limited expression period and a high variation in expression levels between infiltrated leaves may complicate the characterization of effector functions (Bashandy et al. 2015). On the contrary, stable transformation of plants with effector expression/silencing constructs allows to study the impact of the effectors in plant-parasite interactions for a long period of time. Also, we expect to have a uniformed phenotype in the progenies of a transgenic line, which facilitates the assessment of effector functions. However, not all the plants are transformable. Even if it is possible, creating transgenic plants is a time-consuming task. Furthermore, if the effector is too toxic to the plant, creation of effector expressing plant is not possible. Thus, combination of both transient expression – to identify effector candidates that induce phenotypes – and stable expression – to deeper investigate the effector functions – have been used to identify and characterize some effectors (Pitino and Hogenhout 2012; Elzinga et al. 2014).

The type three secretion system (T3SS) of plant pathogen *P. syringae* pv. tomato DC3000 (Pst) was also exploited to develop efficient system to test candidate effector genes (Sohn et al. 2007; Rentel et al. 2008). This system, called Effector Detector Vector (EDV) allows direct delivery of individual effector candidates in plant cells and was used to do high-throughput screening of effector functions in the oomycete *Hyaloperonospora arabidopsidis* (Fabro et al. 2011). Pst DC3000 injects a battery of own effectors into plants and can mask the function of the effectors delivered by EDV. Therefore, the *hrp/hrc* (hypersensitive response and pathogenicity-hr conserved) gene cluster that encodes T3SS from *Pseudomonas syringae* pv. *syringae* 61 was stably integrated in the genome of non-virulent *Pseudomonas fluorescens Pf0-1* providing another tool, called Effector-to-Host Analyzer (EtHAN), to study *in planta* effector functions (Thomas et al. 2009). These effector delivery systems present the advantage to allow large-scale effector screening in a reasonable amount of time and are probably transferable to different plants as *P. fluorescens* has a broad host range. However, variability of effector expression may affect the detection of phenotype. Moreover, contrary to *Agrobacterium* mediated transient or stable expression, T3SS delivery systems cannot be used to perform silencing experiments.

Other common functional tools are derived from phytoviruses (Qu 2001). Virus-based gene expression and virus-induced-gene-silencing (VIGS) are thus developed for various plant species from different families such as pea (Meziadi et al. 2016; Meziadi et al. 2017), soybean (Whitham et al. 2016), potato (Lacomme and Chapman 2005), tobacco (Zhang and Thomma 2014), and grapevine (Kurth et al. 2012; Park et al. 2016). The cloning of effector sequences can be achieved in a reasonable time and plants can be infected by direct inoculation of viruses (Meziadi et al. 2016) or inoculation mediated by agroinfiltration. However, as other techniques, some variations in expression may affect function characterization. In a perspective of aphid effector studies, virus-based approaches would represent a useful tool as some viruses are known to target specifically the phloem (Boissinot et al. 2017).

The generalist green peach aphid *Myzus persicae* is able to feed on many plant species from numerous plant families including the model plants, *N. benthamiana* and *A. thaliana* (Blackman and Eastop 2000). The potato aphid *Macrosiphum euphorbiae* is also highly polyphagous, feeding on hundreds of species in more than twenty plant families with a preference for the Solanaceae, allowing experiments with model plant *N. benthamiana* (Moeller 1973). Thus, in plant-aphid interactions, availabilities in expression tools and model plant genetic resources have facilitated the characterization of *M. persicae* and *M. euphorbiae* effector gene functions. Various studies used transgenic *A. thaliana* lines to express *M. persicae* salivary genes or their dsRNAs constitutively and analyzed fitness of aphids feeding on these plants (Pitino et al. 2011; Pitino and Hogenhout 2012; Mathers et al. 2017). Aphid fitness was also studied with agroinfiltrated *N. benthamiana* transiently expressing dsRNA

specific to *M. persicae* salivary genes (Bos et al. 2010; Pitino and Hogenhout 2012; Rodriguez et al. 2017) and *M. euphorbiae* salivary proteins (Atamian et al. 2012). In addition to *Agrobacterium*-mediated expression methods, both *P. syringae* pv. tomato and *P. fluorescens* were used to deliver *M. euphorbiae* salivary proteins in tomato *Solanum lycopersicum* by T3SS and then by following aphid performances (Atamian et al. 2012; Kettles and Kaloshian 2016). These tools have been developed in selected plants and allowed the functional characterization of the salivary effectors of the aphids that feed on those plants.

Unlike *M. persicae* and *M. euphorbiae*, *A. pisum* is a specialist of legumes, and it is not able to feed on *Arabidopsis* or *Nicotiana*. Its host plants are restricted to the Fabaceae family in which plants remain difficult to manipulate even if some genetic resources are now available for *Medicago truncatula* (Cannon et al. 2006) and *Pisum sativum* (Franssen et al. 2011). Stable transformation of *M. truncatula* R108 is now possible (Cosson et al. 2006), but this plant is not selective for *A. pisum* biotypes as it represents a common host like *Vicia faba*. Thus, most of the *A. pisum* salivary effectors that were shown to be required for maximum fitness of the aphid were characterized by gene silencing through injection of dsRNAs or siRNAs directly in aphid body (Mutti et al. 2006; Mutti et al. 2008; Guo et al. 2014; Naessens et al. 2015; Pan et al. 2015; Wang et al. 2015a; Wang et al. 2015b). These studies revealed that some salivary genes are essential for *A. pisum* feeding in host plant but not on artificial diet as survival rate of the silenced aphids is not affected on artificial diet (Naessens et al. 2015; Wang et al. 2015a; Wang et al. 2015b). Except for the work by Naessens et al. (2015), which showed significant decrease in aphid fecundity up to 13 days post dsRNA injection (dpi) when *Mifl* gene is silenced (even if optimal silencing period was between 4 and 6 dpi), other works did not measure fecundity likely because the silencing lasted too short time. Indeed, in Pan et al. (2015), the silencing of the salivary gene ACYPI006346 (*A. pisum* homologue of *M. persicae* *Mp1*) was no longer effective 3 days after dsRNA injection, and they did not observe reduction in survival rate within 7 days after injection. Even when the silencing is effective at 3 dpi, effector candidate phenotypes may not be detected (Guo et al. 2014). Two studies showed decreases in *A. pisum* survival rates when *Armet* and *ACE* genes were silenced by dsRNA injection (Wang et al. 2015a; Wang et al. 2015b). However, the inhibition of gene expression can be variable as different members of the *ACE* family showed expression reduction from 22% to 85% (Wang et al. 2015b), whereas *Armet* gene expression was reduced by around 40% (Wang et al. 2015a). Thus, depending of the target gene, the silencing efficiency and duration seem to vary when dsRNA are injected. In *Arabidopsis/Nicotiana-M. persicae* systems, RNA silencing by plant produced dsRNA also leads to variable silencing efficiency, with a maximum of gene silencing around 60% (Pitino et al. 2011; Pitino and Hogenhout 2012).

From our global analysis of salivary genes, we identified 740 SG-up-regulated genes from *A. pisum* that are predicted to encode secreted effectors with 35 of them truly secreted by aphids in artificial diet (chapter II). Most of them have no homology with already characterized genes and their functions are unknown. Hence, it is necessary to study the functions of those effector candidates to understand molecular plant-*A. pisum* interactions. *A. pisum* seems to have a large repertoire of effectors suggesting possible redundant functions between them. Therefore, the silencing of one effector may be masked by redundant effector functions and may not affect aphid performance. Comparative genomics and evolutionary analyses also suggested that the expansion of salivary gene families that occurred in *A. pisum* genome may play a key role in host plant adaptation. In the case of multigenic families, it can be difficult to design siRNAs that specifically target one family member. So, the *in planta* expression of effectors in legumes may overcome these limits of silencing methods and then may provide a tool to study the pea aphid effector functions. Because stable transformation of *A. pisum* host plants remains labor-intensive and time-consuming, transient expression may allow to screen the different genes from the SG-up-regulated effector candidate set in a reasonable time.

Therefore, we first aimed to develop *in planta* transient expression of salivary effectors mediated by *Agrobacterium tumefaciens* in several legumes including *M. sativa*, *Trifolium repens* and *T. pratense* and *P. sativum*, which are the host plants of the well-studied *A. pisum* biotypes, and we succeeded in optimizing the protocol in *P. sativum* (Guy et al. 2016). Once the tool was validated by the analysis of one salivary candidate gene, we examined aphid phenotypes (survival and fecundity) that could be induced by expression of some other promising effector candidates in *P. sativum*.

## **2. Paper 2: Optimization of agroinfiltration in *Pisum sativum* provides a new tool for studying the salivary protein functions in the pea aphid complex**





# Optimization of Agroinfiltration in *Pisum sativum* Provides a New Tool for Studying the Salivary Protein Functions in the Pea Aphid Complex

Endrick Guy<sup>1†</sup>, Hélène Boulain<sup>1†</sup>, Yoann Aigu<sup>1</sup>, Charlotte Le Pennec<sup>1</sup>, Khaoula Chawki<sup>1</sup>, Stéphanie Morlière<sup>1</sup>, Kristina Schädel<sup>2</sup>, Grit Kunert<sup>2</sup>, Jean-Christophe Simon<sup>1</sup> and Akiko Sugio<sup>1\*</sup>

<sup>1</sup> INRA, UMR1349, Institute of Genetics, Environment and Plant Protection, Le Rheu, France, <sup>2</sup> Department of Biochemistry, Max Planck Institute for Chemical Ecology, Jena, Germany

## OPEN ACCESS

### Edited by:

Jyoti Shah,  
University of North Texas, USA

### Reviewed by:

Saskia C. M. Van Wees,  
Utrecht University, Netherlands  
Karam B. Singh,  
Commonwealth Scientific  
and Industrial Research Organisation,  
Australia

### \*Correspondence:

Akiko Sugio  
akiko.sugio@inra.fr

<sup>†</sup>These authors have contributed  
equally to the work.

### Specialty section:

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

**Received:** 29 April 2016

**Accepted:** 20 July 2016

**Published:** 09 August 2016

### Citation:

Guy E, Boulain H, Aigu Y,  
Le Pennec C, Chawki K, Morlière S,  
Schädel K, Kunert G, Simon J-C and  
Sugio A (2016) Optimization  
of Agroinfiltration in *Pisum sativum*  
Provides a New Tool for Studying  
the Salivary Protein Functions  
in the Pea Aphid Complex.  
*Front. Plant Sci.* 7:1171.  
doi: 10.3389/fpls.2016.01171

Aphids are piercing-sucking insect pests and feed on phloem sap. During feeding, aphids inject a battery of salivary proteins into host plant. Some of these proteins function like effectors of microbial pathogens and influence the outcome of plant-aphid interactions. The pea aphid (*Acyrtosiphon pisum*) is the model aphid and encompasses multiple biotypes each specialized to one or a few legume species, providing an opportunity to investigate the underlying mechanisms of the compatibility between plants and aphid biotypes. We aim to identify the aphid factors that determine the compatibility with host plants, hence involved in the host plant specialization process, and hypothesize that salivary proteins are one of those factors. *Agrobacterium*-mediated transient gene expression is a powerful tool to perform functional analyses of effector (salivary) proteins in plants. However, the tool was not established for the legume species that *A. pisum* feeds on. Thus, we decided to optimize the method for legume plants to facilitate the functional analyses of *A. pisum* salivary proteins. We screened a range of cultivars of pea (*Pisum sativum*) and alfalfa (*Medicago sativa*). None of the *M. sativa* cultivars was suitable for agroinfiltration under the tested conditions; however, we established a protocol for efficient transient gene expression in two cultivars of *P. sativum*, ZP1109 and ZP1130, using *A. tumefaciens* AGL-1 strain and the pEAQ-HT-DEST1 vector. We confirmed that the genes are expressed from 3 to 10 days post-infiltration and that aphid lines of the pea adapted biotype fed and reproduced on these two cultivars while lines of alfalfa and clover biotypes did not. Thus, the pea biotype recognizes these two cultivars as typical pea plants. By using a combination of ZP1109 and an *A. pisum* line, we defined an agroinfiltration procedure to examine the effect of *in planta* expression of selected salivary proteins on *A. pisum* fitness and demonstrated that transient expression of one candidate salivary gene increased the fecundity of the aphids. This result confirms that the agroinfiltration can be used to perform functional analyses of salivary proteins in *P. sativum* and consequently to study the molecular mechanisms underlying host specialization in the pea aphid complex.

**Keywords:** pea aphid, *Acyrtosiphon pisum*, Leguminosae, agroinfiltration, salivary proteins, biotypes, host specialization, effector

## INTRODUCTION

Herbivorous insects present a high level of species diversity and a large majority of them is specialized to feed on certain host plant species. Specialization to different host plants also occurs within single insect species and leads to the existence of distinguishable “host races” or “biotypes” (Dres and Mallet, 2002). The mechanisms of host plant adaptation in herbivorous insects are poorly understood, although these could explain a large part of insect species richness (Simon et al., 2015). Therefore, insect species displaying an array of races or biotypes provide interesting opportunities to study the process of host plant specialization due to the possibility to compare genomes and feeding strategies between closely related races or biotypes.

The pea aphid, *Acyrtosiphon pisum* Harris, is the first aphid to be genome sequenced and owing to its long history of research, it is the model of aphids and sap-feeding insects (hemipterans; International Aphid Genomics Consortium, 2010). In addition, *A. pisum* encompasses a range of biotypes each specialized to one or a few closely related legume species but cannot survive or reproduce well on non-host legume plants. So far, 15 biotypes are described (Peccoud et al., 2015), of which alfalfa, clover and pea biotypes are the ones most studied in host specialization (Hawthorne and Via, 2001; Ferrari et al., 2008; Peccoud et al., 2009; Jaquiere et al., 2012; Via et al., 2012). In addition to show strong differences in performances on host and non-host plants, these biotypes are genetically distinct and can be distinguished by using microsatellite markers (Ferrari et al., 2008; Peccoud et al., 2009). Interestingly, all the *A. pisum* biotypes studied so far feed well on *Vicia faba*, which is considered as a universal host plant for pea aphids (Ferrari et al., 2008; Peccoud et al., 2009). Many of these *A. pisum* biotypes can be crossed with other biotypes (Peccoud et al., 2014), and QTL analyses have been used to identify aphid factors that determine the compatibility with the host plants (Hawthorne and Via, 2001; Via et al., 2012; Kanvil et al., 2015).

Aphids feed on plant phloem sap using a specialized mouthpart called stylet. During feeding, aphids may transmit plant pathogenic viruses, inject toxic saliva and remove nutrients from host plants. Hence, aphids are considered among the most serious crop pests. Recent studies gradually revealed that there are intricate molecular interactions between the proteins secreted with aphid saliva and host plant proteins (Elzinga and Jander, 2013; Rodriguez and Bos, 2013; Kaloshian and Walling, 2016). In some cases, salivary proteins trigger plant defense responses (De Vos and Jander, 2009; Chaudhary et al., 2014; Elzinga et al., 2014), in others, they suppress plant defense reactions and promote aphid proliferation (Will et al., 2007; Bos et al., 2010; Atamian et al., 2013; Elzinga et al., 2014; Naessens et al., 2015). Hence, aphid salivary proteins are considered to be analogous to effectors of plant pathogens, and their functions have been examined using similar techniques, such as silencing of salivary genes or *in planta* expression of salivary proteins (Elzinga and Jander, 2013; Rodriguez and Bos, 2013). The first characterized aphid salivary gene was an *A. pisum* gene named C002, which is strongly expressed in salivary glands and was detected in plants

infested by the aphids. Silencing of *A. pisum* C002 (*ApC002*) was achieved by injection of siRNA in aphids. It prevented aphids from feeding on *V. faba*, while aphid feeding on artificial diet was unaffected (Mutti et al., 2006, 2008). In line with these studies, transient or stable expression of *Myzus persicae* orthologue of *ApC002*, *MpC002*, in *Nicotiana benthamiana* and *Arabidopsis thaliana*, respectively, increased the fecundity of *M. persicae* feeding on these plants, indicating the conserved role of C002 as an effector required for aphid feeding on host plants (Bos et al., 2010; Pitino and Hogenhout, 2013; Elzinga et al., 2014).

Since then, several *A. pisum* salivary proteins required for aphid full performance have been identified and characterized mostly by using gene silencing induced by siRNA injection to aphids (Guo et al., 2014; Pan et al., 2015; Wang et al., 2015a,b) while several salivary proteins from other aphids, such as *M. persicae* have been identified using transient or stable *in planta* expression of salivary genes (Bos et al., 2010; Pitino and Hogenhout, 2013; Elzinga et al., 2014). However, since the *A. pisum* genome is extensively duplicated and more than 2000 gene families show massive expansion compared to published insect genomes (Rispe et al., 2008; International Aphid Genomics Consortium, 2010; Jaubert-Possamai et al., 2010), it is often difficult to select a siRNA or dsRNA fragment that specifically targets the gene of interest for silencing. In some cases, co-silencing of multiple gene family members need to be examined to determine whether the phenotype observed is due to the silencing of single gene or multiple genes. Furthermore, there is a possibility that gene silencing does not show a strong phenotypic effect on plant–aphid interactions if genes with redundant functions exist or if gene silencing is too transient.

On the other hand, *in planta* expression of saliva gene allows simple characterization of single gene in plant-aphid interactions. While the construction and multiplication of transgenic plants require several months to years of preparation before testing, *Agrobacterium* mediated transient gene expression (agroinfiltration) can be achieved in a few days; therefore, it is a commonly used technique to identify and characterize effector functions. However, the efficiency of agroinfiltration is highly variable and often depends on the compatibility between the *Agrobacterium tumefaciens* strain and the plant species or cultivar used (Wroblewski et al., 2005). The technique has been developed in *N. benthamiana* using a disarmed strain where the virulence factors encoded by the Ti plasmid were deleted (Goodin et al., 2008). Then, the technique was optimized for different plants such as potato (Bhaskar et al., 2009), lettuce (Chen et al., 2016), grapevine (Santos-Rosa et al., 2008), *Medicago truncatula* (Picard et al., 2013) and recently in soybeans (King et al., 2015). However, the technique is not established in the legume plants, which are hosts for *A. pisum*.

As mentioned earlier, *A. pisum* encompasses multiple biotypes which cannot survive on the plants they are not specialized to. We study the commonest and most studied pea aphid biotypes to identify the factors that determine the compatibility between the aphid and legume species as such factors are likely be involved in the host plant specialization process of the aphids. Based on

our recent genome analysis of three aphid biotypes respectively specialized on clover, alfalfa and pea, we hypothesized that salivary proteins are one of the factors that are involved in the host plant specialization process in *A. pisum* (Jaquierey et al., 2012). Hence, we envisaged to identify salivary proteins with biotype specific polymorphisms and to characterize their effects on specific plant–aphid interactions. Some salivary proteins from non-adapted biotypes may induce resistance responses in non-host plants while some salivary proteins from adapted biotypes may suppress specific plant defense reactions and allow non-adapted aphids to feed on non-host plants.

Here, as the first step to reach the objectives and to facilitate identification and functional characterization of *A. pisum* salivary proteins, we undertook optimization of agroinfiltration in *Medicago sativa* (alfalfa) and *Pisum sativum* (pea). We focused on these two plants because (1) significant amount of studies have been done on the aphid biotypes that feed on these plants (Hawthorne and Via, 2001; Jaquierey et al., 2012; Via et al., 2012), (2) these two biotypes show clear-cut performance difference on these two plants (Peccoud et al., 2009), and (3) seeds of various cultivars are easily available in our research center.

## MATERIALS AND METHODS

### Aphids, Bacteria Strains, Plasmids and Growth Conditions

Aphid lineages, and bacterial strains and plasmids used in this study are listed in Supplementary Tables S1 and S2, respectively. All aphid lineages were reared in a growth chamber at 18°C with a 16 h day/8 h night photoperiod on the broad bean, *Vicia faba* (Castel), at low density to avoid the production of winged individuals. *Escherichia coli* and *A. tumefaciens* strains were grown on Luria-Bertani medium at 37°C and 30°C, respectively. For solid media, agar was added at a final concentration of 1.5% (w/v). Antibiotics were used at the following concentrations: for all bacteria, 50 µg/ml kanamycin; for *A. tumefaciens*, 50 µg/ml rifampicin; for *E. coli*, 10 µg/ml gentamycin.

### Plants and Growth Conditions

*Pisum sativum* (Supplementary Table S3) and *Medicago sativa* plants were grown in a growth chamber at 18°C with a 16 h day/8 h night photoperiod for 2 and 3 weeks, respectively.

### Measurements of Aphid Performances on Pea Cultivars

Life traits of five aphid lineages from pea (Ar\_Po\_28, Ar\_Po\_58), alfalfa (L9Ms14) and clover (YR2, T8005) biotypes (Supplementary Table S1) were measured on *P. sativum* cultivars ZP1130 and ZP1109 (Supplementary Table S3). Adult aphids were installed on both pea cultivars and removed 24 h later, giving them enough time to produce 10 larvae that were left on the plants (day 1). Survival rate of the 10 larvae was measured at day 9 (when they reach adulthood), three surviving adult aphids were then reinstalled on the plants and biomass (the cumulated weight of the three adults and their offspring) of the aphid

population was weighted at day 17. The biomass is a good proxy of the number of nymphs produced by adult aphids and reflects well their overall fitness (Peccoud et al., 2009). Five replicates for each aphid lineage on the two tested plants were performed.

### Construction of Plasmids

All primers used in this study are listed in Supplementary Table S4. The genes encoding eGFP and the β-glucuronidase with a plant derived intron (GUSi; Vancanneyt et al., 1990) were amplified using GFP-Fw/GFP-Rv primers and GUS-Fw/GUS-Rv primers, respectively, and were added complete *attB1* and *attB2* sequences by the second PCR with *attB1* and *attB2* primers. In order to clone aphid salivary genes, cDNAs produced from aphid head total RNA were used to enrich transcripts encoding salivary genes. Adult aphids feeding on *V. faba* were flash frozen in liquid nitrogen, and decapitated with a scalpel between the first and second pairs of legs. Head RNA was extracted from 10 to 20 individuals using the RNeasy plant mini kit (Qiagen). cDNA synthesis was performed with poly-T primers using the AMV reverse transcriptase system (Promega) according to the manufacturers' instructions. ACYPI009919 (Ap25) and ACYPI008617 (ApC002) open reading frames encoding mature proteins were amplified from the cDNA of the Ar\_Po\_58 line (pea biotype) with Phusion DNA polymerase (ThermoFisher Scientific) using AP25-Fw/AP25-Rv and APC002-Fw/APC002-Rv primers (Supplementary Table S4), respectively. *attB1* and *attB2* sites were added with a second PCR using *attB1* and *attB2* primers. All amplicons, eGFP, GUSi and two salivary genes, were recombined by BP reaction into pDONR207 (Invitrogen) using BP clonase II (Invitrogen) and produced entry vectors (Table S2). Entry vectors were recombined by LR reaction using LR clonase II (Invitrogen) into pEAQ-HT-DEST1 expression vector (Supplementary Table S2; Sainsbury et al., 2009). Expression vectors were transformed in electro-competent *A. tumefaciens* cells (Supplementary Table S2).

### Infiltration of *Agrobacterium*

*Agrobacterium tumefaciens*-mediated transient expression was performed as described (Rivas et al., 2002). Freshly cultured cells were resuspended in induction buffer [10 mM MgCl<sub>2</sub>, 10 mM Mes (2-(*N*-morpholino) ethanesulfonic acid), pH 5.6, and 150 µM acetosyringone] to an O.D.<sub>600</sub> (optical density at 600 nm) of 0.5. Cells were syringe infiltrated into leaves of 2 week-old *P. sativum* (Supplementary Table S1) and 3 week-old *M. sativa* plants.

### GUS Staining

Plant leaves infiltrated with *Agrobacterium* were detached 3 days post infiltration (dpi) and GUS activity was visualized as described (Jefferson, 1987). Briefly, leaves were vacuum infiltrated with GUS staining solution (61 mM Na<sub>2</sub>HPO<sub>4</sub>, 39 mM NaHPO<sub>4</sub>, 0.1% triton X-100, 10 mM EDTA, 0.3% H<sub>2</sub>O<sub>2</sub> and 1.5 mM 5-bromo, 4-chloro, 3-indolyl glucuronide (X-Glc, Biosynth), pH 7.0) and incubated overnight at 37°C. Then chlorophyll discoloration was performed with successive washes with ethanol at 37°C.

## Protein Extraction and Western-Blot Analyses

Three leaf disks per leaf were sampled using a cork borer (area = 0.79 cm<sup>2</sup>) at 0, 7, and 10 days post-infiltration for GFP protein detection. Leaf disks were flash frozen in liquid nitrogen and stored at -80°C. Proteins were extracted in 120 µl extraction buffer (50 mM Tris pH 7.5, 1 µM Dithiothreitol, glycerol 10%, 1 mM PMSF (Phenylmethylsulfonyl fluoride), 0.05% Triton X-100). Extracts from pea plants were prepared as described (Canonne et al., 2011) and supernatants were resuspended in 5X loading buffer (0.5 M Tris pH 6.8, SDS 10%, glycerol 50 and 0.001% bromophenol blue). Fifteen microliters of samples were separated by SDS-PAGE (12% polyacrylamide) and transferred on PVDF (Polyvinylidene fluoride) membranes, (Merck Millipore) as described (Witte et al., 2004) with following modifications: PVDF membranes were soaked in methanol before and after transfer, and then washed in water. Methanol in transfer buffer was replaced by ethanol. The rabbit anti-GFP antibody (Biorad) and secondary antibodies (polyclonal goat anti-rabbit antibody peroxidase conjugated; Sigma-Aldrich) were both used at 1:10000. Detection was performed by chemiluminescence using Clarity Western ECL Substrate (Biorad) and CL-XPosure™ Film (Lifetechnologies) according to manufacturer's instructions. Coomassie stains were performed with 0.2% Coomassie Brilliant Blue R250 (Sigma) in 50:40:10 water, methanol, acetic acid.

## Aphid Performance Test on Agroinfiltrated Leaves

One young leaf of the *P. sativum* ZP1109 cultivar was syringe-infiltrated with *A. tumefaciens* AGL-1 strain harboring expression vectors. Three days later (at 3 dpi), 6 new-born aphids (1 day-old) born on *V. faba* were installed on *P. sativum* agroinfiltrated leaves in custom-built clip cages (area = 2.54 cm<sup>2</sup>). When aphids were 8 days-old (10 dpi), clip cages were opened and the number of surviving aphids was recorded to estimate the survival rate. From the survivors, one average sized aphid was selected and transferred to a new *P. sativum* leaf that was infiltrated with the same construct of *Agrobacterium* 3 days before the transfer. Clip cages were opened when aphids were 12 and 15 days-old to assess the fecundity by counting the number of nymphs produced by each aphid. The nymphs were removed after each counting to avoid overcrowding of the cages. In one experiment, 10 replicates per gene were performed and the same experiment was repeated twice, producing 20 replicates. All the experiments were conducted at 20°C, 16 h day/8 h night photoperiod.

## Statistical Analyses

All statistical analyses were conducted in R version 3.1.2 (R Core Team, 2014). Data were checked for approximate normal distribution by graphical visualizing of residuals. The effects of the different factors (pea cultivar, aphid lineage, expressed gene) were tested and the simplest model explaining the data was used. Analyses of survival rates (Figures 2A and 3A) and fecundity counts (Figure 3B) were performed by classical linear regressions using generalized linear models (GLM) with binomial

and Poisson distributions, respectively. Both tests were followed by multiple comparisons of means by the Tukey contrast method implemented in the package "multcomp" (Hothorn et al., 2008). The influence of pea cultivars and aphid lineage on aphid biomass (Figure 2B) was analyzed by a two-way ANOVA. Tukey's *post hoc* multiple comparisons of means from the R package "agricolae" (De Mendiburu, 2014) were used to reveal differences between groups.

## RESULTS

### Screening of *P. sativum* and *M. sativa* Cultivars for Agroinfiltration

Combinations of *A. tumefaciens* and various *M. sativa* and *P. sativum* cultivars were tested using the β-glucuronidase containing a plant derived intron (GUSi) as a reporter gene (Vancanneyt et al., 1990). Green fluorescence protein (GFP) could not be used as a reporter due to strong auto fluorescence induced in the leaf surface by the infiltration. Initially, we tested two plant expression vectors pGWB402Ω (Nakagawa et al., 2007) and pEAQ-HT-DEST1 (Sainsbury et al., 2009) in some pea cultivars, but the difference in expression levels between the two vectors was not very clear or slightly better when pEAQ-HT-DEST1 was used. Therefore, we used pEAQ-HT-DEST1 for the rest of screening. Also, our initial test showed that a bacterial suspension with an O.D.<sub>600</sub> less than 0.3 resulted in a weak transgene expression and more than 0.7 triggered leaf chlorosis a few days after infiltration. Therefore, for the rest of the screening, agroinfiltrations were performed using syringe infiltration method and a bacterial suspension with an O.D.<sub>600</sub> = 0.5. Seventeen *P. sativum* (Supplementary Table S3) and five *M. sativa* cultivars were selected based on geographic origin and phylogenetic groups in order to screen a large genetic diversity. Each cultivar was infiltrated with three *Agrobacterium* strains [C58C1, GV3101 and AGL-1 (Supplementary Table S2)] each harboring pEAQ-HT-DEST1-GUSi to identify the combination of plant and bacterium genotypes that produce high amount of GUS proteins. Leaves were analyzed histochemically for GUS activity at 3 dpi. At least three independent experiments were performed for each combination and results are summarized in Table 1. None of the *M. sativa* cultivars was suitable for *Agrobacterium*-mediated transient expression in leaves as no GUS staining could be observed in these plants under the tested conditions. High differences between pea cultivars were observed. Most of the pea cultivars had no or weak intensities of GUS staining. Of the three *Agrobacterium* strains used in this study, AGL-1 induced the highest expression of GUS, and C58C1 was the lowest inducer. Two pea genotypes, ZP1130 and ZP1109, inoculated with AGL-1 showed most intense coloration during GUS staining (Figure 1A). GUS staining could be observed at 3 dpi for both cultivars, ZP1130 and ZP1109. To confirm protein expression in these two cultivars, transient expression of eGFP and detection by western-blot was performed (we could not visualize GFP fluorescence due to autofluorescence induced by wounding). eGFP protein was detected at 7 and 10 dpi for both ZP1109 and

**TABLE 1 | Results of screening of *P. sativum* and *M. sativa* cultivars for agroinfiltration.**

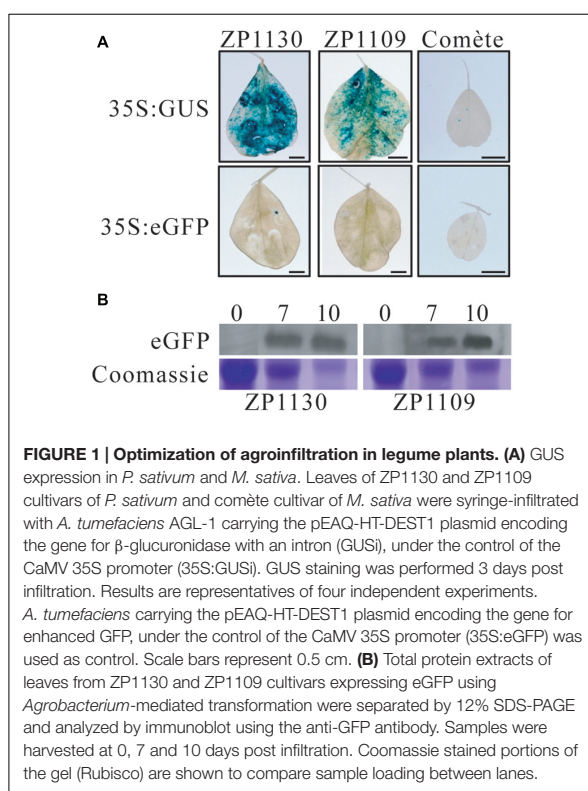
	C58C1 <sup>b</sup>	GV301	AGL-1
<b><i>Pisum sativum</i><sup>a</sup></b>			
AP3783	N	I	W
AP3830	N	N	N
WP1018	W	W	W
ZP690	N	N	N
ZP748	N	N	N
ZP750	N	W	W
ZP793	N	W	W
ZP747	N	N	W
ZP1109	N	I	I
ZP1124	N	N	W
ZP1130	N	W	S
ZP3495	N	W	N
ZP3508	N	N	N
ZP3514	N	N	N
ZP3535	N	N	N
ZP3570	N	N	N
ZP3664	N	W	W
<b><i>Medicago sativum</i></b>			
Comète	nd	nd	N
Harpe	nd	nd	N
Lux Timbale	nd	nd	N
Lux Galaxie	nd	nd	N
Cannelle	nd	nd	N

<sup>a</sup>pea or alfalfa cultivars used in this study. <sup>b</sup>*A. tumefaciens* strains used in this study. N, no coloration; W, weak; I, intermediate coloration; S, strong coloration (Figure 1A), nd, not determined.

ZP1130 (Figure 1B). During this study, yellowing of the leaves starting at 9–10 dpi for ZP1130 and at 12–13 dpi for ZP1109 was observed. This leaf yellowing was probably due to AGL-1 infection as the yellowing was observed in the leaves infiltrated with *Agrobacterium* with empty vector control, and no yellowing was observed in buffer infiltrated leaves (data not shown). Taken together, we identified two pea cultivars, ZP1130 and ZP1109, and the *A. tumefaciens* strain AGL-1 as the combinations that are suitable for transient gene expression, and we presumed that 3–8 dpi for ZP1130 and 3–10 dpi for ZP1109 are the timing to examine the effect of transgene expression in the plant or plant–aphid interactions.

### Pea Cultivars ZP1130 and ZP1109 Are Hosts Only for the *A. pisum* Pea Biotype

Survival rate and biomass of the five *A. pisum* lineages belonging to three biotypes (pea, alfalfa and clover; Supplementary Table S1) were assessed on the ZP1130 and ZP1109 pea cultivars we identified as suitable for agroinfiltration (Figure 2). Analysis revealed that the two plant cultivars did not influence the survival rate and produced aphid biomass [ $\chi^2 = 0.14$ ,  $P = 0.243$ ;  $F_{(5,44)} = 129.7$ ,  $P = 0.261$ ; for survival and biomass, respectively], but pea aphid lineages differed significantly in their survival rates ( $\chi^2 = 19.04$ ,  $P < 0.001$ ) and biomass production [ $F_{(4,45)} = 128.9$ ,

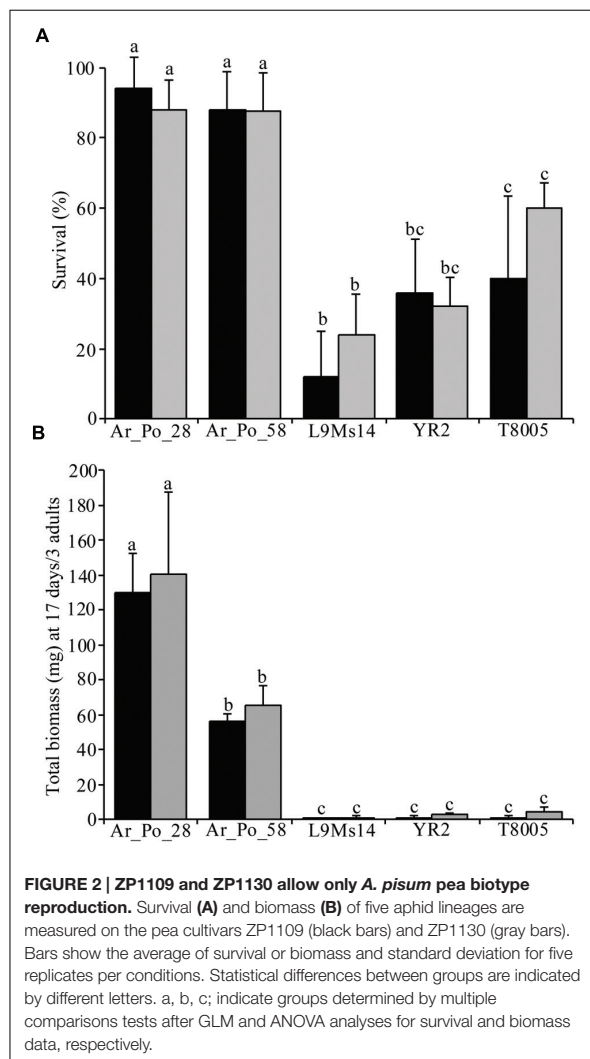


**FIGURE 1 | Optimization of agroinfiltration in legume plants. (A)** GUS expression in *P. sativum* and *M. sativa*. Leaves of ZP1130 and ZP1109 cultivars of *P. sativum* and comète cultivar of *M. sativa* were syringe-infiltrated with *A. tumefaciens* AGL-1 carrying the pEAQ-HT-DEST1 plasmid encoding the gene for  $\beta$ -glucuronidase with an intron (GUSi), under the control of the CaMV 35S promoter (35S:GUSi). GUS staining was performed 3 days post infiltration. Results are representatives of four independent experiments. *A. tumefaciens* carrying the pEAQ-HT-DEST1 plasmid encoding the gene for enhanced GFP, under the control of the CaMV 35S promoter (35S:eGFP) was used as control. Scale bars represent 0.5 cm. **(B)** Total protein extracts of leaves from ZP1130 and ZP1109 cultivars expressing eGFP using *Agrobacterium*-mediated transformation were separated by 12% SDS-PAGE and analyzed by immunoblot using the anti-GFP antibody. Samples were harvested at 0, 7 and 10 days post infiltration. Coomassie stained portions of the gel (Rubisco) are shown to compare sample loading between lanes.

$P < 0.001$ ]. The pea adapted lineages Ar\_Po\_28 and Ar\_Po\_58 showed a higher survival rate on the pea cultivars at day 9 compared to L9Ms14 (alfalfa biotype), YR2 and T8005 (clover biotype). The difference in survival was very pronounced between pea and alfalfa specialized lineages, and intermediate for lineages of the clover biotype (Figure 2A). On both ZP1130 and ZP1109 cultivars, only the lineages of the pea biotype (Ar\_Po\_28 and Ar\_Po\_58) produced a substantial biomass. Although Ar\_Po\_28 had a significantly higher biomass than Ar\_Po\_58, both lineages performed well on the tested cultivars that they seem to recognize as favorable hosts. By contrast, alfalfa and clover adapted lineages hardly reproduced on the pea cultivars that seem to be non-host plants in these interactions (Figure 2B). Thus, the ZP1130 and ZP1109 cultivars are selective hosts for *A. pisum* biotypes, allowing to assess host and non-host interactions using agroinfiltration experiments.

### Transient Expression of AP25 in ZP1109 Increased *A. pisum* Fecundity

Next, we expressed two salivary genes in ZP1109 by agroinfiltration using strain AGL-1 and examined their effects on *A. pisum* feeding on the infiltration site. We chose Ar\_Po\_58 as a test aphid line as it belongs to the pea biotype and harbors no secondary symbiont, which may interfere with plant–aphid interactions. Mature proteins encoding ACYPI008617 (ApC002) and ACYPI009919, which we named Ap25, were



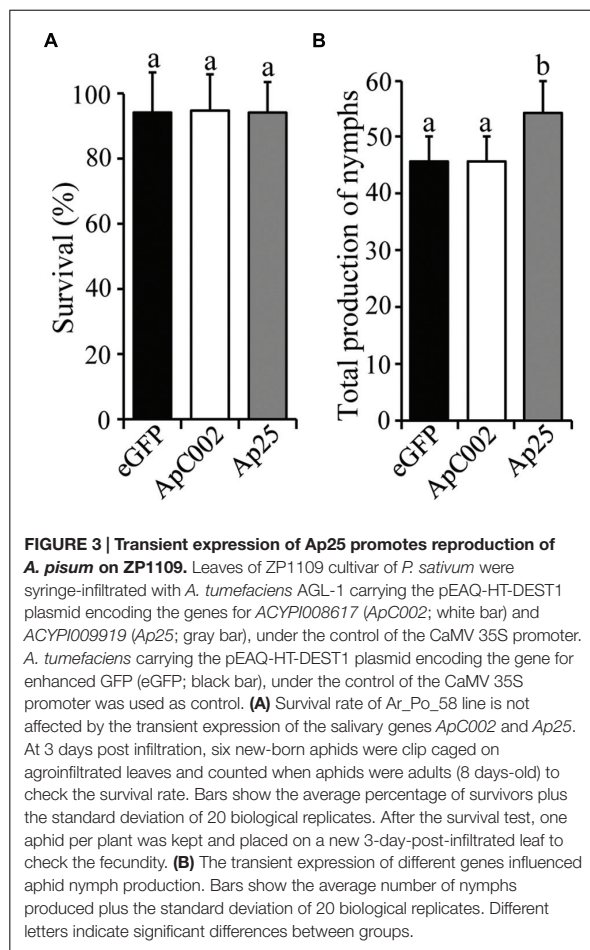
transiently expressed using pEAQ-HT-DEST1 vector. The genes were expressed by CaMV 35S promoter, which is known to be ubiquitously and constitutively activated in various plant tissues including epidermal, mesophyll and phloem tissues (Stockhaus et al., 1989). In the process of establishing phloem feeding, *A. pisum* punctures various tissues and salivates (Schwarzkopf et al., 2013). When the aphid attempts to feed on non-host legume plant, it punctures epidermal and mesophyll cells but cannot establish phloem feeding; therefore, the factors that determine the compatibility between *A. pisum* and host plants are present in those tissues (Schwarzkopf et al., 2013). Based on these informations, we thought it is important to express salivary proteins ubiquitously to fully assess their functions in plants and used 35S promoter for transient expression. 35S promoter has been successfully used in other studies on aphid salivary proteins (Bos et al., 2010; Naessens et al., 2015). *ApC002* was

chosen because it is one of the most studied salivary proteins and is shown to be essential for *A. pisum* to feed on the universal host plant (*V. faba*; Ferrari et al., 2008; Peccoud et al., 2009). *Ap25* was selected because the gene presents the same features as that of *ApC002*: the gene was identified in salivary glands by transcriptomic analyses (Carolan et al., 2011), is specifically expressed in salivary glands (Akiko Sugio et al., unpublished data), and encodes a signal peptide and a small (13.9 kDa) mature protein with no predicted function. Although many genes are duplicated in *A. pisum* genome, *Ap25*, like *ApC002*, is single copy in *A. pisum* and its orthologues exist only in the Aphididae family (Hélène Boulain et al., unpublished data).

In this study, transient protein expression was observed from 3 (detected by GUS activity) to 10 days (detected by western blot) at 20°C after infiltration of *Agrobacterium. A. pisum* starts to reproduce around 9th day after birth, reaches its peak of reproduction around 5 days later, and slows down but continues to reproduce until its death at an age of approximately 30 days (Tsuchida et al., 2004). By supplying newly infiltrated leaves, we extended the duration of the experiment to characterize the effect of transgene expression on aphid fecundity. Leaves of ZP1109 were infiltrated with AGL-1 harboring expression plasmids of eGFP, *ApC002* or *Ap25*. Three days after the infiltration, six new-born aphids of the pea adapted clone Ar\_Po\_58 were clip caged on the infiltrated leaves. When the aphids were 8 days-old (at 10 dpi) the cages were opened to count the number of survivors. One aphid was transferred to a new 3-day-post-infiltrated leaf. Production of nymphs of the caged adult was measured when the aphid was 12 and 15-day old corresponding to the peak of reproduction of adults. Survival rate and total number of nymphs of the aphids are shown in Figure 3. There was no difference in the survival rate of the aphids that were fed on the leaves expressing the three tested genes ( $\chi^2 = 0.01$ ,  $P = 0.96$ ). Production of nymphs of Ar\_Po\_58 feeding on *ApC002* expressing leaves was same as that of aphids feeding on eGFP expressing leaves, while the aphids produced approximately 12% more offspring on *Ap25* expressing leaves than on eGFP expressing leaves (20 biological replicates,  $\chi^2 = 18.75$ ,  $P < 0.001$ ). The results indicate that *Ap25* plays a role in promoting *A. pisum* feeding on *P. sativum*.

## DISCUSSION

Here, we screened cultivars of *P. sativum* and *M. sativa* using GUS activity as a reporter and identified two *P. sativum* cultivars, ZP1130 and 1109, that are amenable to *Agrobacterium* mediated transient gene expression. We noted that *A. tumefaciens* strain AGL-1 was the most efficient strain among the three strains tested. This can be explained by the presence of extra virulent factors in this strain (Jin et al., 1987). We also noted that a few days upon infiltration with high concentration of *A. tumefaciens* (O.D.<sub>600</sub> > 0.7), chlorosis appeared and was restricted to the agroinfiltrated area. Pruss et al. (2008) also observed that fully virulent and disarmed *A. tumefaciens* strains also triggered chlorosis restricted to the infiltrated area in tobacco plants.



Although the mechanisms underlying this chlorosis have not been well understood, it could be due to a defense response to the *A. tumefaciens* involving the chloroplasts (Pruss et al., 2008).

Two tested *A. pisum* lines belonging to the pea biotype reproduced well on these two cultivars, while members of the alfalfa and clover biotypes could not survive and reproduce well on them. This indicates that these two cultivars serve as host plants of the pea biotype only and can be used to characterize candidate aphid salivary genes that may determine the compatibility of *A. pisum* biotypes with *P. sativum*. Interestingly, we found differences in aphid performances, as measured by survival and biomass, between the two *P. sativum* adapted lines on both pea cultivars. In particular, biomass production by *Ar\_Po\_28* was about twice more than that of *Ar\_Po\_58*. Since the two lines differ in both genotype and symbiont composition (*Ar\_Po\_28* harbors *Rickettsia* and *Serratia* secondary symbionts while *Ar\_Po\_58* is free of any secondary symbiont, Supplementary Table S1), it is difficult to tell which factor (aphid genome or symbiont status), alone or in interaction, accounts for these differences in performances.

Although we optimized agroinfiltration in *P. sativum* to study the host specialization mechanisms in *A. pisum*, the system can be used to study the functions of *P. sativum* genes or effectors of other pea parasites. *P. sativum* is an important legume crop used in arable rotations for the production of nutritious food for both humans and animals. Various projects to identify genes involved in *P. sativum* biotic and abiotic stress resistances are ongoing (Lejeune-Henaut et al., 2008; Hamon et al., 2013; Desgroux et al., 2016) and whole-genome sequencing of *P. sativum* is underway (Alves-Carvalho et al., 2015). Therefore, the *P. sativum* research community is in need of various tools to analyze the genes of agronomical interest that will be identified in near future. Though *P. sativum* is reported to be stably transformed (Svabova et al., 2005), it remains to be a time consuming and difficult task. Recent application of virus vectors in *P. sativum* provides a new tool to express transgene in pea plant relatively quickly, but it is still time consuming (in a few weeks; Meziadi et al., 2016) and the agroinfiltration method described here provides another way to express transgenes in a few days. By using various Gateway™ compatible vectors available for agroinfiltration (Karimi et al., 2002; Nakagawa et al., 2007), fusion proteins or dsRNA will be easily produced in *P. sativum* leaves. Furthermore, coexpression of a few proteins may be realized by infiltration of *A. tumefaciens* with different expression constructs.

We transiently expressed *ApC002* and *Ap25* in *P. sativum* leaves and examined the survival and fecundity of an *A. pisum* line of the pea adapted biotype. The aphids grew well in the clip cages fixed on the agroinfiltrated leaves and produced offspring. Since *ApC002* is required for *A. pisum* feeding on *V. faba* plant, which is a universal plant of all *A. pisum* biotypes (Ferrari et al., 2008; Peccoud et al., 2009), and *in planta* (*Arabidopsis* and *N. benthamiana*) expression of *MpC002* increases the fecundity of *M. persicae* feeding on the plants, we expected that *ApC002* expression in *P. sativum* leaves would also increase the fecundity of the aphids. However, the survival and fecundity of the aphids fed on *ApC002* expressing plants were at the same level as that of the aphids feeding on *eGFP* expressing plants. As *C002* is one of the abundantly expressed salivary genes in *A. pisum* (Mutti et al., 2006), the aphids may produce enough of this protein and may not benefit significantly from extra production of *ApC002* in *P. sativum* leaves. On the other hand, expression of *Ap25* in *P. sativum* leaves increased the fecundity of the aphids. *Ap25* is an Aphididae specific gene which encodes a small protein with a signal peptide. As the protein does not show homology with known proteins, the function of *Ap25* is unknown. It is possible that the protein interferes with plant defense reactions triggered by aphid feeding and facilitates nutrient acquisition from the pea plant. Carolan et al. (2011) identified more than 300 salivary genes in *A. pisum* and more than half of the identified genes encode proteins with unknown function (Carolan et al., 2011). The agroinfiltration method described here provides a mean to examine the functions of those salivary proteins in relatively short time and also allows us to investigate whether those genes are determinants of compatibility between *P. sativum* and *A. pisum* biotypes. As the second step of this study, we envisage to express salivary proteins with biotype specific sequences in the pea leaves and examine how they affect the performance

of different pea biotypes installed on the leaves. Further, the agroinfiltration technique can be combined with aphid gene silencing to investigate whether a gene expressed in leaves can complement the silenced gene function (Naessens et al., 2015). Studies on plant–insect interactions at a molecular level are less advanced compared to plant–microbe interaction studies partly because it is not yet possible to transform insect herbivores. The tools to manipulate host plants, like the method described here, can provide alternative ways to examine plant–insect interactions at a molecular level and will be able to contribute to advance the field.

## AUTHOR CONTRIBUTIONS

EG, HB, YA, CLP, KC, SM, J-CS, and AS designed the experiments. EG, HB, YA, CLP, KC, SM, KS, J-CS and AS conducted the experiments. EG, HB, GK, J-CS and AS edited the manuscript. GK, J-CS and AS provided funding for the project.

## REFERENCES

- Alves-Carvalho, S., Aubert, G., Carrere, S., Cruaud, C., Brochet, A. L., Jacquin, F., et al. (2015). Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant J.* 84, 1–19. doi: 10.1111/tpj.12967
- Atamian, H. S., Chaudhary, R., Dal Cin, V., Bao, E., Girke, T., and Kaloshian, I. (2013). In planta expression or delivery of potato aphid *Macrosiphum euphorbiae* effectors Me10 and Me23 enhances aphid fecundity. *Mol. Plant Microbe Interact.* 26, 67–74. doi: 10.1094/mpmi-06-12-0144-fi
- Bhaskar, P. B., Venkateshwaran, M., Wu, L., Ane, J. M., and Jiang, J. M. (2009). *Agrobacterium*-mediated transient gene expression and silencing: a rapid tool for functional gene assay in potato. *PLoS ONE* 4:e5812. doi: 10.1371/journal.pone.0005812
- Bos, J. I., Prince, D., Pitino, M., Maffei, M. E., Win, J., and Hogenhout, S. A. (2010). A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (green peach aphid). *PLoS Genet.* 6:e1001216. doi: 10.1371/journal.pgen.1001216
- Canonne, J., Marino, D., Jauneau, A., Pouzet, C., Briere, C., Roby, D., et al. (2011). The Xanthomonas type III effector XopD targets the *Arabidopsis* transcription factor MYB30 to suppress plant defense. *Plant Cell* 23, 3498–3511. doi: 10.1105/tpc.111.088815
- Carolan, J. C., Caragea, D., Reardon, K. T., Mutti, N. S., Dittmer, N., Pappan, K., et al. (2011). Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrtosiphon pisum*): a dual transcriptomic/proteomic approach. *J. Proteome Res.* 10, 1505–1518. doi: 10.1021/pr100881q
- Chaudhary, R., Atamian, H. S., Shen, Z. X., Brigg, S. P., and Kaloshian, I. (2014). GroEL from the endosymbiont *Buchnera aphidicola* betrays the aphid by triggering plant defense. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8919–8924. doi: 10.1073/pnas.1407687111
- Chen, Q., Dent, M., Hurtado, J., Stahnke, J., McNulty, A., Leuzinger, K., et al. (2016). Transient protein expression by agroinfiltration in lettuce. *Methods Mol. Biol.* 1385, 55–67. doi: 10.1007/978-1-4939-3289-4\_4
- De Mendiburu, F. (2014). *Agricolae: Statistical Procedures for Agricultural Research*. R Package Version 1.2-1. Available at: <http://CRAN.R-project.org/package=agricolae>
- De Vos, M., and Jander, G. (2009). *Myzus persicae* (green peach aphid) salivary components induce defence responses in *Arabidopsis thaliana*. *Plant Cell Environ.* 32, 1548–1560. doi: 10.1111/j.1365-3040.2009.02019.x
- Desgroux, A., L'Anthoene, V., Roux-Duparque, M., Riviere, J. P., Aubert, G., Tayeh, N., et al. (2016). Genome-wide association mapping of partial resistance

## ACKNOWLEDGMENTS

Authors acknowledge to Gaëtan Denis and Jean-François Le Gallic for experimental support, Drs Marie-Laure Pilet-Nayel and Alain Baranger and their team for providing pea seeds, Anne Moussart for providing alfalfa seeds, Dr. George P. Lomonosoff for providing pEAQ vectors, Prof. Tsuyoshi Nakagawa for providing pGWB vectors and Mathew Smoker for *Agrobacterium* strains. This project was supported by the French Agence Nationale de la Recherche grant (13-JSV7-0012-01-Bugsplit) to AS and by Procope German-French collaboration program funded by the German Academic Exchange Service (DAAD) and Campus France to GK (57049694) and J-CS (30753SE). Ph.D. grant to HB is funded by INRA SPE and Région Bretagne.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01171>

to *Aphanomyces euteiches* in pea. *BMC Genomics* 17:124. doi: 10.1186/s12864-016-2429-4

- Dres, M., and Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. B Biol. Sci.* 357, 471–492. doi: 10.1098/rstb.2002.1059
- Elzinga, D. A., De Vos, M., and Jander, G. (2014). Suppression of plant defenses by a *Myzus persicae* (green peach aphid) salivary effector protein. *Mol. Plant Microbe Interact.* 27, 747–756. doi: 10.1094/mpmi-01-14-0018-r
- Elzinga, D. A., and Jander, G. (2013). The role of protein effectors in plant-aphid interactions. *Curr. Opin. Plant Biol.* 16, 451–456. doi: 10.1016/j.pbi.2013.06.018
- Ferrari, J., Via, S., and Godfray, H. C. (2008). Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. *Evolution* 62, 2508–2524. doi: 10.1111/j.1558-5646.2008.00468.x
- Goodin, M. M., Zaitlin, D., Naidu, R. A., and Lommel, S. A. (2008). *Nicotiana benthamiana*: its history and future as a model for plant-pathogen interactions. *Mol. Plant Microbe Interact.* 21, 1015–1026. doi: 10.1094/mpmi-21-8-1015
- Guo, K., Wang, W., Luo, L., Chen, J., Guo, Y., and Cui, F. (2014). Characterization of an aphid-specific, cysteine-rich protein enriched in salivary glands. *Biophys. Chem.* 189, 25–32. doi: 10.1016/j.bpc.2014.03.006
- Hamon, C., Coyne, C. J., McGee, R. J., Lesne, A., Esnault, R., Mangin, P., et al. (2013). QTL meta-analysis provides a comprehensive view of loci controlling partial resistance to *Aphanomyces euteiches* in four sources of resistance in pea. *BMC Plant Biol.* 13:45. doi: 10.1186/1471-2229-13-45
- Hawthorne, D. J., and Via, S. (2001). Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* 412, 904–907. doi: 10.1038/35091062
- Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* 50, 346–363. doi: 10.1002/bimj.200810425
- International Aphid Genomics Consortium (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8:e1000313. doi: 10.1371/journal.pbio.1000313.g001
- Jaquiere, J., Stoeckel, S., Nouhaud, P., Mieuze, L., Maheo, F., Legeai, F., et al. (2012). Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex. *Mol. Ecol.* 21, 5251–5264. doi: 10.1111/mec.12048
- Jaubert-Possamai, S., Risse, C., Tanguy, S., Gordon, K., Walsh, T., Edwards, O., et al. (2010). Expansion of the miRNA pathway in the Hemipteran Insect *Acyrtosiphon pisum*. *Mol. Biol. Evol.* 27, 979–987. doi: 10.1093/molbev/msp256



- Jefferson, R. A. (1987). Assaying chimeric genes in plants: the GUS Gene fusion system. *Plant Mol. Biol. Rep.* 5, 387–405. doi: 10.1007/BF02667740
- Jin, S., Komari, T., Gordon, M. P., and Nester, E. W. (1987). Genes responsible for the supervirulence phenotype of *Agrobacterium tumefaciens* A281. *J. Bacteriol.* 169, 4417–4425.
- Kaloshian, I., and Walling, L. L. (2016). Hemipteran and dipteran pests: effectors and plant host immune regulators. *J. Integr. Plant Biol.* 58, 350–361. doi: 10.1111/jipb.12438
- Kanvil, S., Collins, C. M., Powell, G., and Turnbull, C. G. N. (2015). Cryptic virulence and avirulence alleles revealed by controlled sexual recombination in pea aphids. *Genetics* 199, 581–593. doi: 10.1534/genetics.114.173088
- Karimi, M., Inze, D., and Depicker, A. (2002). GATEWAYTM vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/s1360-1385(02)02251-3
- King, J. L., Finer, J. J., and McHale, L. K. (2015). Development and optimization of agroinfiltration for soybean. *Plant Cell Rep.* 34, 133–140. doi: 10.1007/s00299-014-1694-4
- Lejeune-Henaut, I., Hanocq, E., Bethencourt, L., Fontaine, V., Delbreil, B., Morin, J., et al. (2008). The flowering locus Hr colocalizes with a major QTL affecting winter frost tolerance in *Pisum sativum* L. *Theor. Appl. Genet.* 116, 1105–1116. doi: 10.1007/s00122-008-0739-x
- Meziadi, C., Blanchet, S., Richard, M. M. S., Pilet-Nayel, M. L., Geffroy, V., and Pflieger, S. (2016). Bean pod mottle virus: a new powerful tool for functional genomics studies in *Pisum sativum*. *Plant Biotechnol. J.* 14, 1777–1787. doi: 10.1111/pbi.12537
- Mutti, N. S., Louis, J., Pappan, L. K., Pappan, K., Begum, K., Chen, M. S., et al. (2008). A protein from the salivary glands of the pea aphid, *Acyrtosiphon pisum*, is essential in feeding on a host plant. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9965–9969. doi: 10.1073/pnas.0708958105
- Mutti, N. S., Park, Y., Reese, J. C., and Reece, G. R. (2006). RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrtosiphon pisum*. *J. Insect Sci.* 6, 1–7. doi: 10.1673/031.006.3801
- Naessens, E., Dubreuil, G., Giordanengo, P., Baron, O. L., Minet-Kebdani, N., Keller, H., et al. (2015). A secreted MIF cytokine enables aphid feeding and represses plant immune responses. *Curr. Biol.* 25, 1898–1903. doi: 10.1016/j.cub.2015.05.047
- Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., et al. (2007). Improved gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci. Biotechnol. Biochem.* 71, 2095–2100. doi: 10.1271/bbb.70216
- Pan, Y. M., Zhu, J. J., Luo, L., Kang, L., and Cui, F. (2015). High expression of a unique aphid protein in the salivary glands of *Acyrtosiphon pisum*. *Physiol. Mol. Plant Pathol.* 92, 175–180. doi: 10.1016/j.pmp.2015.04.006
- Peccoud, J., de la Huerta, M., Bonhomme, J., Laurence, C., Outreman, Y., Smadja, C. M., et al. (2014). Widespread host-dependent hybrid unfitness in the pea aphid species complex. *Evolution* 68, 2983–2995. doi: 10.1111/evo.12478
- Peccoud, J., Maheo, F., De La Huerta, M., Laurence, C., and Simon, J. C. (2015). Genetic characterisation of new host-specialised biotypes and novel associations with bacterial symbionts in the pea aphid complex. *Insect Conserv. Divers.* 8, 484–492. doi: 10.1111/icad.12131
- Peccoud, J., Ollivier, A., Plantegenest, M., and Simon, J. C. (2009). A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7495–7500. doi: 10.1073/pnas.0811117106
- Picard, K., Lee, R., Hellens, R., and Macknight, R. (2013). Transient gene expression in *Medicago truncatula* leaves via Agroinfiltration. *Methods Mol. Biol.* 1069, 215–226. doi: 10.1007/978-1-62703-613-9\_15
- Pitino, M., and Hogenhout, S. A. (2013). Aphid protein effectors promote aphid colonization in a plant species-specific manner. *Mol. Plant Microbe Interact.* 26, 130–139. doi: 10.1094/MPMI-07-12-0172-FI
- Pruss, G. J., Nester, E. W., and Vance, V. (2008). Infiltration with *Agrobacterium tumefaciens* induces host defense and development-dependent responses in the infiltrated zone. *Mol. Plant Microbe Interact.* 21, 1528–1538. doi: 10.1094/MPMI-21-12-1528
- R Core Team (2014). *A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rispe, C., Kutsukake, M., Doublet, V., Hudaverdian, S., Legeai, F., Simon, J. C., et al. (2008). Large gene family expansion and variable selective pressures for cathepsin B in aphids. *Mol. Biol. Evol.* 25, 5–17. doi: 10.1093/molbev/msm222
- Rivas, S., Romeis, T., and Jones, J. D. G. (2002). The Cf-9 disease resistance protein is present in an approximately 420-kilodalton heteromultimeric membrane-associated complex at one molecule per complex. *Plant Cell* 14, 689–702. doi: 10.1105/tpc.010357
- Rodriguez, P. A., and Bos, J. I. B. (2013). Toward understanding the role of aphid effectors in plant infestation. *Mol. Plant Microbe Interact.* 26, 25–30. doi: 10.1094/mpmi-05-12-0119-fi
- Sainsbury, F., Thuenemann, E. C., and Lomonosoff, G. P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* 7, 682–693. doi: 10.1111/j.1467-7652.2009.00434.x
- Santos-Rosa, M., Poutaraud, A., Merdinoglu, D., and Mestre, P. (2008). Development of a transient expression system in grapevine via agro-infiltration. *Plant Cell Rep.* 27, 1053–1063. doi: 10.1007/s00299-008-0531-z
- Schwarzkopf, A., Rosenberg, D., Niebergall, M., Gershenzon, J., and Kunert, G. (2013). To feed or not to feed: plant factors located in the epidermis, mesophyll, and sieve elements influence pea aphid's ability to feed on legume species. *PLoS ONE* 8:e75298. doi: 10.1371/journal.pone.0075298
- Simon, J. C., d'Alençon, E., Guy, E., Jacquin-Joly, E., Jaquiere, J., Nouhaud, P., et al. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Brief. Funct. Genomics.* 14, 413–423. doi: 10.1093/bfpg/evl015
- Stockhaus, J., Schell, J., and Willmitzer, L. (1989). Correlation of the expression of the nuclear photosynthetic gene ST-LS1 with the presence of chloroplasts. *EMBO J.* 8, 2445–2451.
- Svabova, L., Smykal, P., Griga, M., and Ondrej, V. (2005). *Agrobacterium*-mediated transformation of *Pisum sativum* in vitro and in vivo. *Biol. Plant.* 49, 361–370. doi: 10.1007/s10535-005-0009-6
- Tsuchida, T., Koga, R., and Fukatsu, T. (2004). Host plant specialization governed by facultative symbiont. *Science* 303:1989. doi: 10.1126/science.1094611
- Vancannet, G., Schmidt, R., O'Connor-Sanchez, A., Willmitzer, L., and Rocha-Sosa, M. (1990). Construction of an intron-containing marker gene: splicing of the intron in transgenic plants and its use in monitoring early events in *Agrobacterium*-mediated plant transformation. *Mol. Gen. Genet.* 220, 245–250. doi: 10.1007/BF00260489
- Via, S., Conte, G., Mason-Foley, C., and Mills, K. (2012). Localizing F(ST) outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow. *Mol. Ecol.* 21, 5546–5560. doi: 10.1111/mec.12021
- Wang, W., Dai, H. E., Zhang, Y., Chandrasekar, R., Luo, L., Hiromasa, Y., et al. (2015a). Armet is an effector protein mediating aphid-plant interactions. *Faseb J.* 29, 2032–2045. doi: 10.1096/fj.14-266023
- Wang, W., Luo, L., Lu, H., Chen, S. L., Kang, L., and Cui, F. (2015b). Angiotensin-converting enzymes modulate aphid-plant interactions. *Sci. Rep.* 5:8885. doi: 10.1038/srep08885
- Will, T., Tjallingii, W. F., Thonnessen, A., and van Bel, A. J. (2007). Molecular sabotage of plant defense by aphid saliva. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10536–10541. doi: 10.1073/pnas.0703535104
- Witte, C. P., Noël, L. D., Gielbert, J., Parker, J. E., and Romeis, T. (2004). Rapid one-step protein purification from plant material using the eight-amino acid StrepII epitope. *Plant Mol. Biol.* 55, 135–147. doi: 10.1007/s11103-004-0501-y
- Wroblewski, T., Tomczak, A., and Michelmore, R. (2005). Optimization of *Agrobacterium*-mediated transient assays of gene expression in lettuce, tomato and *Arabidopsis*. *Plant Biotechnol. J.* 3, 259–273. doi: 10.1111/j.1467-7652.2005.00123.x

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

Copyright © 2016 Guy, Boulain, Aigu, Le Penne, Chawki, Morlière, Schädel, Kunert, Simon and Sugio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

### 3. Functional characterization of salivary effector candidates in the *A. pisum*-pea plant system

#### INTRODUCTION

The optimization of *Agrobacterium* mediated transient expression established by Guy et al. (2016) offers now the possibility to investigate the role of *A. pisum* salivary candidate genes that may be required for the interaction between *A. pisum* and *P. sativum* host. Thus, we could explore the functions of SG-up-regulated candidate effectors for which we identified evolutionary patterns suggesting a role in pea aphid adaptation to legumes (chapter II).

Among the SG-up-regulated candidate effectors identified in the pea aphid genome, 17% are found only in Aphididae species and have probably evolved upon salivary glands specialization occurred during aphid phloem-feeding style establishment. Some of these Aphididae specific genes evolved faster than the others, corresponding to those that are most likely involved in the arms race with host plant defenses. By contrast, slow evolving ones may be linked to general feeding functions in aphids. Even if we have no idea about their role in biotype specialization yet, some of the Aphididae fast-evolving genes may contribute to successful feeding of pea biotype aphids on pea plant, as it is the case for *Ap25* (Guy et al. 2016, chapter II).

The SG-up-regulated candidate effector gene set also contains 25% of genes that are found exclusively in the *A. pisum* genome. They may result from recent duplications in *A. pisum* and/or be too divergent to be clustered with orthologues of other aphid/insect species. These *A. pisum* specific salivary effectors also constitute strong candidates for controlling host plant adaptation.

Thus, expressing the Aphididae specific fast-evolving or *A. pisum* specific effector candidates in pea plant may be beneficial for the aphids from the pea biotype. If they function as effectors that overcome plant defenses, we expect better performances of pea biotype aphids (e.g. higher survival rate and fecundity) on *P. sativum* leaves expressing these genes.

Here, we aim to use the *Agrobacterium*-mediated transient expression system to examine the function of pea biotype effector candidates on pea plant (Table 1). We expect that *in planta* expression of candidates will positively or negatively impact aphid fecundity.

## MATERIAL AND METHODS

**Aphid rearing:** Aphid lineages used in this study are listed in Supplementary Table S1. All aphids were reared in a growth chamber at 18°C with a 16h day/8h night photoperiod on the universal host, *Vicia faba* (Castel), at low density to avoid the production of winged individuals.

**Construction of vectors:** The open reading frame encoding mature proteins of selected salivary genes were amplified from cDNA of the ArPo58 line (pea biotype) with the Phusion DNA polymerase (ThermoFisher Scientific) using specific primers (Supplementary Table S5). Then Gateway recombination was performed to clone the genes in the plant expression vector pEAQ-HT-DEST1 (Sainsbury et al. 2009) as described in Guy et al. 2016.

**Agroinfiltration:** Transformation of expression and silencing vectors in *A. tumefaciens* AGL-1 strain and agroinfiltration steps followed the protocol established by Guy et al. 2016.

**Aphid assays:** Aphid life traits (survival and fecundity) were followed on agroinfiltrated leaves of *P. sativum* ZP1109 cultivar. Three days after infiltration (3 dpi), six one-day-old aphids born on *V. faba* were installed on *P. sativum* leaves in clipcages. When larvae reached 8 days old, the number of survivors within each clipcage was recorded to assess survival rate. After that, one survivor by clipcage was kept and installed on a new infiltrated leaf (3 dpi). When aphids were 12 and 15 days old, the clipcages were open to count the number of larvae. After each counting, the larvae were removed to avoid over-crowding in the clipcages. For each condition, 16 replicates were performed and experiments were conducted at 22°C with 16h day/8h night photoperiod.

**Statistical analysis:** Difference on survival rates and fecundity between tested and control aphids were assessed by generalized linear model (GLM) implemented in R (R CoreTeam 2017).

## RESULTS

### *Selected candidate genes*

Because identification and evolutionary analysis of the SG-up-regulated effectors set and functional analysis were conducted in parallel, the selection of candidate genes was mainly based on their SG expression (454 sequencing and RNA-seq pilot experiment) and on a preliminary analyses of all different published and unpublished data on aphid salivary effectors (Harmel et al. 2008; Carolan et al. 2009; Bos et al. 2010; Carolan et al. 2011; Rao et al. 2013, and unpublished saliva proteomics). However, once the analysis of salivary effector sets (chapter II) was finished, it confirmed that the selected genes were up-regulated in salivary glands. So, five salivary genes out of the 100 SG top-

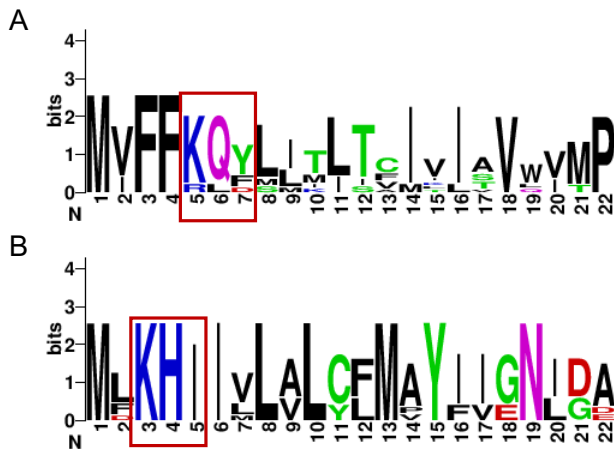
expressed genes within the SG-up-regulated effector candidate gene set were analyzed in this study (Table 1). Three genes (*Ap490*, *Ap1606*, *Ap56*) were Aphididae specific genes that showed accelerated evolutionary (dN/dS) rates among the aphid species (chapter II). These genes may play a fundamental role, and the amino acid changes may have been selected during the adaptation to a particular host. Based on high SG expression and high dN/dS ratio they represent strong effector candidates although only *Ap1606* protein was detected in saliva by artificial diet experiment. The low sensitivity of saliva proteomics and/or possible requirement of plant cues (Powell et al. 2006) to trigger aphid secretion of particular effectors may explain the non-detection of these proteins in artificial diet experiments. As most of the aphid effectors, the encoded proteins do not harbor known functional domains.

**Table 1.** Characteristic of tested salivary effector candidate genes

Gene name	NCBI ID Aphidbase ID	Protein size (aa)	Proteomics		Effector set	Expression rank	Orthology level	dN/dS	dN/dS	dN/dS
			Saliva	SG				<i>A. pisum</i> vs <i>A. glycines</i>	<i>A. pisum</i> vs <i>D. noxia</i>	<i>A. pisum</i> vs <i>M. persicae</i>
<i>Ap490</i>	LOC100159087 ACYPI000490	258	No	Yes	SG-up-regulated	8	Aphididae	0.33	0.72	0.40
<i>Ap1606</i>	LOC100160301 ACYPI001606	228	Yes	Yes	SG-up-regulated	39	Aphididae	0.73	0.79	0.50
<i>Ap56</i>	LOC100574728 ACYPI56566	128	No	Yes	SG-up-regulated	83	Aphididae	0.50	1.02	0.45
<i>Ap223</i>	LOC100158789 ACYPI000223	204	No	Yes	SG-up-regulated	18	<i>A. pisum</i> specific	NA	NA	NA
<i>Ap7553</i>	LOC100166702 ACYPI007553	157	Yes	Yes	SG-up-regulated	13	<i>A. pisum</i> specific	NA	NA	NA

Two other genes, *Ap223* and *Ap7553* were selected among *A. pisum* specific SG-up-regulated candidate effectors. They belong to two groups of candidate effectors (Supplementary Table S6) that encode proteins with similar signal peptides and divergent C-terminal regions. The two groups of candidate effectors were named KQY and KHI in reference to conserved signal peptide motifs (Figure 1; Supplementary Figures S1 and S2). All the members of the groups were classified in the SG-up-regulated candidate effectors. Some of them present a particular gene structure with many small exons (3-15 bp) distributed on large genomic region, especially for *Ap223* which contains 37 exons in 45,000 bp and encodes only 204 amino acids. From these two groups, *Ap223* and *Ap7553* genes were selected based on their high SG expression level and successful detection in saliva proteomics. *Ap223* was not detected in the saliva proteomics presented in chapter II but was previously detected in another saliva proteomics dataset (Frank White personal communication). The characteristics of KQY or KHI salivary gene groups resemble those of effectors of other plant pathogens, such as RxLR or CRN effectors of oomycetes (Hass et al. 2009) that have conserved N-terminal sequences and variable C-

terminal parts which accumulated mutations that may be positively selected allowing the emergence of new functions (Win et al. 2007).

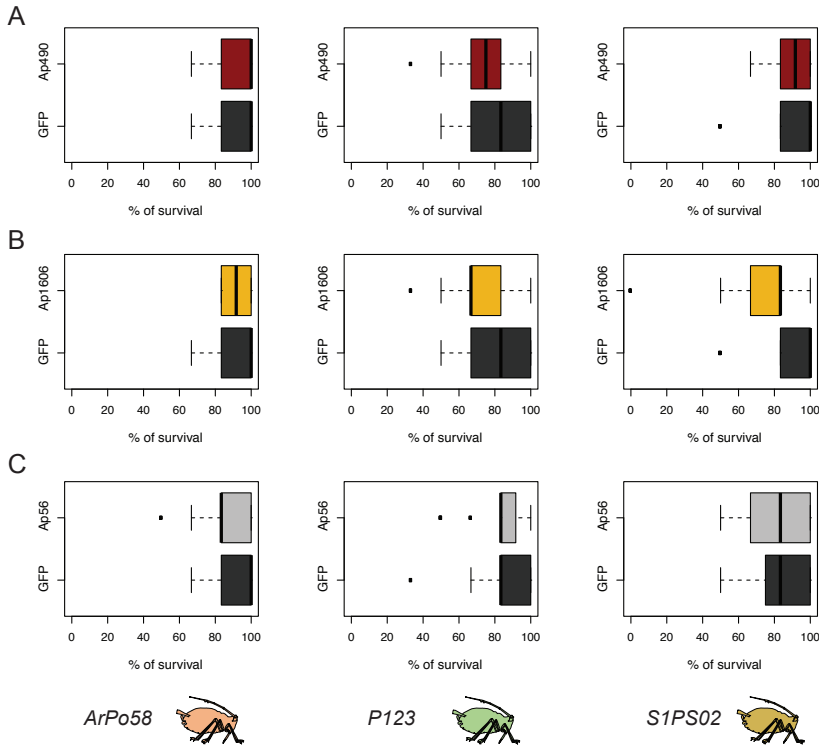


**Figure 1.** Frequencies of amino acids in signal peptides from the KQY (A) and the KHI (B) groups of proteins. Red boxes indicate conserved motifs used to name the groups. WebLogos (Crooks et al. 2004) were constructed after alignment of 11 KQY and 15 KHI proteins with ClustalW (Thompson et al. 1994).

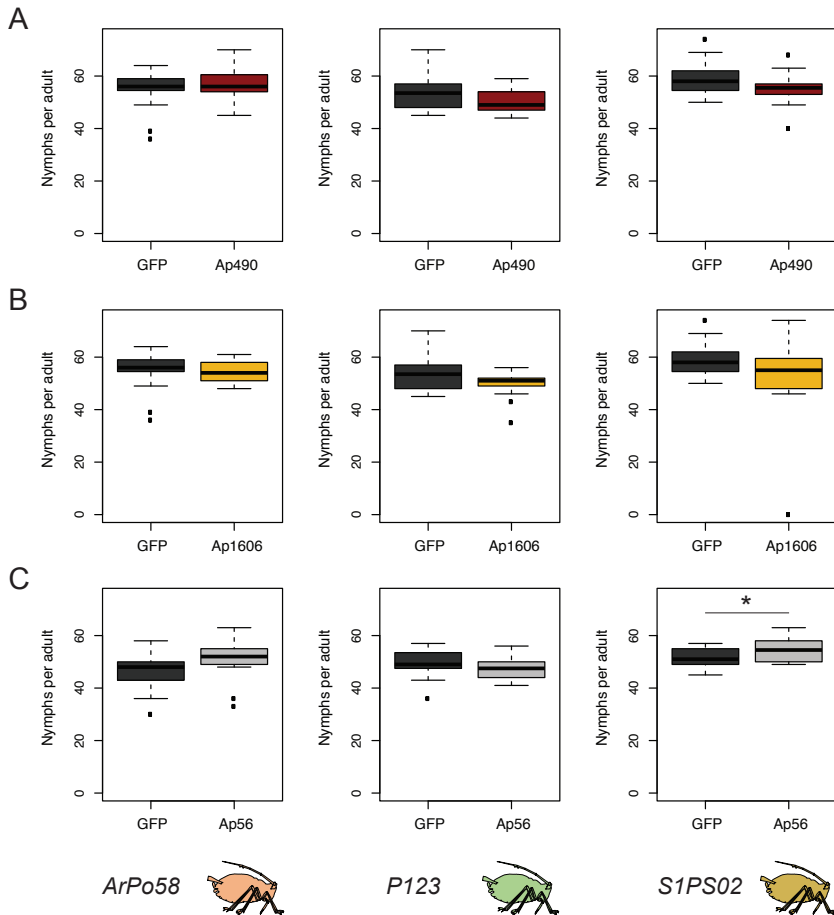
### ***In planta* expression of salivary candidate genes**

To examine the functions of the five candidate genes as effectors, we expressed them one by one in pea plant using agroinfiltration technique and observed the impact of gene expression on aphid survival and fecundity. As a negative control, we expressed the Green Fluorescent Protein (GFP) which is not supposed to affect aphid survival and fecundity. If the tested genes function as effectors (virulence factors) and suppress plant defense reactions, we expect to see an increase of aphid survival rate and/or fecundity. To generalize the results, we examined the effect of salivary effector expression on 3 clones of aphids that belong to the pea biotype and are free of any secondary symbionts.

The three Aphididae specific (*Ap490*, *Ap1606*, and *Ap56*) candidate effectors did not influence aphid survival when they were expressed in pea plant (Figure 2). Aphids from the 3 pea biotype clones presented the same survival rates on the leaves where a candidate effector was expressed as on leaves where GFP gene was expressed. No effect on aphid fecundity was detected except when S1PS02 aphids fed on the pea leaves expressing the *Ap56* gene (Figure 3). Indeed, aphid from clone S1PS02 had a higher fecundity (total at day 15) on *Ap56* infiltrated leaves than on control leaves ( $p=0.0291$ ).

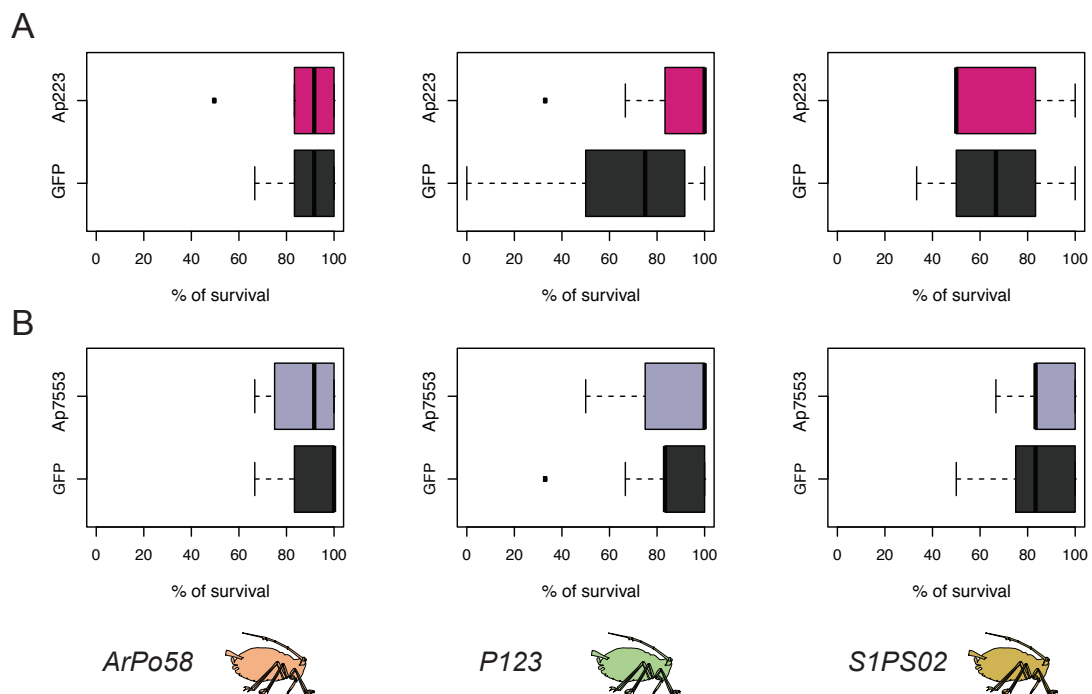


**Figure 2.** Effect of the *in planta* expression of three Aphididae conserved salivary effector candidates *Ap490* (A), *Ap1606* (B) and *Ap56* (C) on aphid survival. Six one-day-old larvae were installed in clipages fixed on agroinfiltrated leaves expressing an effector gene or GFP as a negative control. The number of surviving aphids was counted 8 days after installation. The test was conducting using three aphid lineages (ArPo58, P123, and S1PS02) belonging to the pea biotype. No significant difference was observed by GLM test between tested and control aphids (N=16).

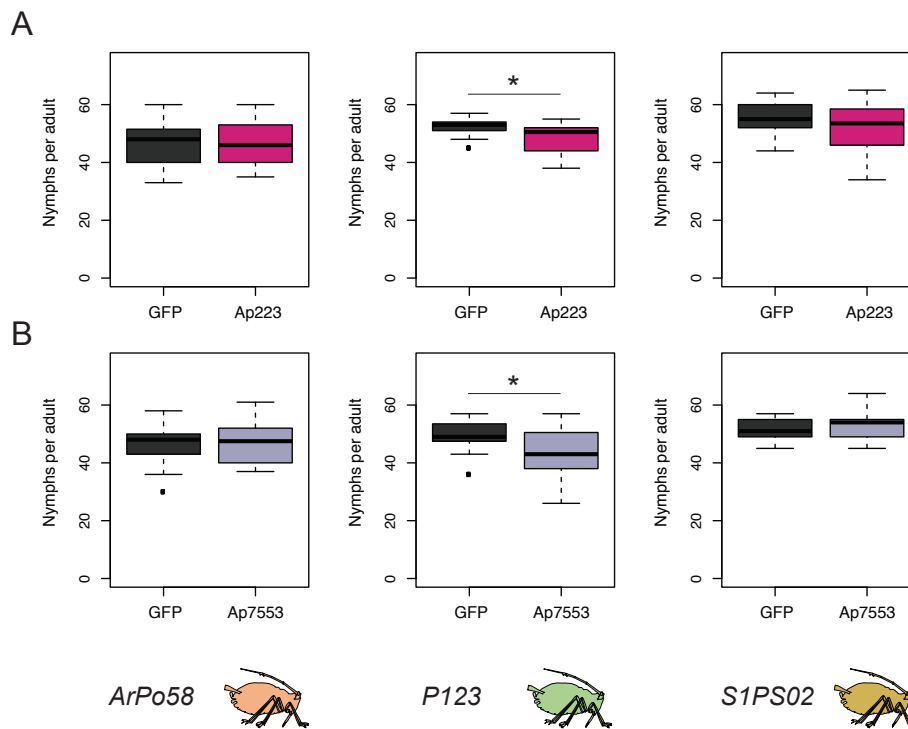


**Figure 3.** Effect of the *in planta* expression of three Aphididae conserved salivary effector candidates *Ap490* (A), *Ap1606* (B) and *Ap56* (C) on aphid fecundity. One aphid from the 6 one-day-old larvae installed for measuring survival, was kept in clipages and put on new agroinfiltrated leaves expressing an effector gene or GFP as a negative control. The test was conducting using three aphid clones (ArPo58, P123, and S1PS02) belonging to the pea biotype. Fecundity of aphid was counted at day 12 and 15 and larvae were removed to avoid over-crowding in the clipages. The global fecundity (cumulative at day 15) is presented here, and the fecundity between tested and control aphids were compared statistically. Statistical differences were assessed by generalized linear models (N=16). \* indicates  $p < 0.05$ .

The expression of the two *A. pisum* specific genes, *Ap223* and *Ap7553*, did not affect survival rate of the three clones on pea plants (Figure 4). The fecundity of ArPo58 and S1PS02 clones was not affected by these KQY and KHI gene expression. However, unlike what was expected, both genes slightly reduced fecundity of the P123 clone when they were expressed in pea leaves (Figure 5). P123 aphids produced on average 52.6 offspring in 15 days on control leaves and 48.3 on *Ap223* infiltrated ones ( $p=0.0198$ ). The *Ap7553* expression led to a slight reduction in fecundity with the production of 43.6 larvae in average compared to the control where 49.5 larvae were produced ( $p=0.0323$ ).



**Figure 4.** Effect of the *in planta* expression of two *A. pisum* specific salivary effector candidates *Ap223* (A), and *Ap7553* (B) on aphid survival. Six one-day-old larvae were installed in clipcages fixed on agroinfiltrated leaves expressing an effector gene or GFP as a negative control. The number of surviving aphids was counted 8 days after installation. The test was conducting using three aphid clones (ArPo58, P123, and S1PS02) belonging to the pea biotype. No significant difference was observed by GLM test between tested and control aphids (N=16).



**Figure 5.** Effect of the *in planta* expression of two *A. pisum* specific salivary effector candidates *Ap223* (A) and *Ap7553* (B) on aphid fecundity. One aphid from the 6 one-day-old larvae installed for measuring survival, was kept in clipcages and put on new agroinfiltrated leaves expressing an effector gene or GFP as a negative control. The test was conducted using three aphid clones (ArPo58, P123, and S1PS02) belonging to the pea biotype. Fecundity by aphid was counted at day 12 and 15 to avoid over-crowding in the clipcages. The global fecundity (cumulative at day 15) is presented here, and the fecundity between tested and control aphids were compared statistically. Statistical differences were assessed by generalized linear models (N=16). \* indicates  $p < 0.05$ .

## DISCUSSION

*Agrobacterium*-mediated transient gene expression allowed us to express salivary effector candidates in pea leaves and examine some induced phenotypes on pea biotype aphids. Contrarily to our working hypothesis, that assumes that effector expression will benefit to aphids, we could not detect positive effect of the tested effector candidate genes on aphid fecundity except for *Ap56* and only for one clone. Indeed, *Ap56* expression in pea induced higher fecundity of S1PS02 aphids. Moreover, we detected a slight negative effect of both *Ap223* and *Ap7553* *A. pisum* specific genes on fecundity of P123 clone. Nonetheless, some points remain to be elucidated before concluding on



candidate gene functions. In Guy et al. (2016) we were able to detect GFP proteins up to 10 days after agroinfiltration in pea using the expression vector pEAQ-HT-DEST1 (Sainsbury et al. 2009). Based on this result, we assumed that aphid effectors were expressed in *P. sativum* leaves during the period of aphid infestation but we did not check for that. Further measures of gene expression (RT-qPCR) or protein detections (Western-Blot) will be necessary to confirm the expression of the effectors. The absence of phenotype may be due to a low expression of *A. pisum* effector in pea leaves and gene expression variations between plants and between gene constructions may complicate the detection of clear phenotype. Absence of phenotype may be also due to sub-optimal localization of the effectors in plant tissues. We do not know if each effector protein reached the phloem sieve elements and whether they must do so to perform their function. Hence, we used the Cauliflower Mosaic Virus (CaMV) 35S promoter, which constitutively activates the expression of down-stream gene in many different plant tissues (epidermal, mesophyll and phloem), to express effector candidates in the leaves of *P. sativum* by hoping that the expressed effectors can reach the cells that exert their functions. However, we cannot discard the possibility that tissue (e.g. phloem) specific expression of effectors gives clearer phenotype than the ubiquitous expression of the effectors by 35S promoter. Nevertheless, both phloem-specific (Bos et al. 2010; Pitino et al. 2011; Pitino and Hogenhout 2012) and ubiquitous (Atamian et al. 2012) promoters have been used for the detection of aphid effector phenotypes.

Assuming that effector genes were well expressed in the tissue they were required, lack of phenotype in these transient expression experiments does not necessarily mean that the genes are not contributing to aphid performances. Indeed, the aphids are well-adapted to pea and may not benefit from the extra proteins that we expressed in these plants. If so, less well-adapted aphids from other biotypes or gene silenced aphids may show an increase of fecundity. In addition to the absence of phenotypes, both *Ap223* (KQY group) and *Ap7553* (KHI group) *A. pisum* specific genes slightly reduced the fecundity of the P123 clone when they are expressed in pea. Further experiments such as measure of expression levels *in planta* and gene silencing could reveal whether a too high expression may cause increased plant defenses reactions.

These experiments also raised an important issue in functional analysis of *A. pisum* effectors: even within a single biotype, the aphid clone factor may have an influence as we noticed an effect of *Ap56*, *Ap223* and *Ap7553* in a clone-dependent manner. The expression level of effector candidate genes in each clone may be the key of these differences (Eyres et al. 2016; our unpublished data). For example, expression level of *Ap56* in the aphid may be significantly lower in S1PS02 compared to other two clones; therefore, S1PS02 might have benefitted from the *in planta* expression of *Ap56*. Pitino and Hogenhout (2012) observed different phenotypes between the expression of *Mp1* and *Mp2* *M. persicae* salivary genes. *Mp1*, which is highly expressed in salivary glands, did not promote aphid

fecundity when it is expressed in *Arabidopsis* whereas *Mp2*, which is 3-fold lowly expressed than *Mp1*, did. Nonetheless, both genes function as effectors because silencing of these genes reduces *M. persicae* performances. So, quantification of expression levels of the five tested genes in the three pea biotype clones is required to understand the observed phenotypes. Effector expression levels may be taken into account to decide the strategies that will be used for further examination of effector functions.

## 4. Discussion and perspectives on functional characterization of salivary effectors

### *Screening of effector phenotypes*

We have established the *Agrobacterium* mediated transient gene expression as a tool to explore the functions of candidate salivary effectors in *A. pisum*-*P. sativum* interactions. Before allocating much time in the development of transgenic pea lines (Svabova et al. 2005), this method offers the possibility to screen functions of *A. pisum* effectors in a reasonable time period, though within and between leaves variations of agroinfiltration efficiency may disturb phenotype detection (Bashandy et al. 2015). *In planta* expression of effectors is supposed to increase aphid performances; however, it was observed in the *M. persicae*-*A. thaliana* system that aphids may not benefit from *in planta* extra expression of effectors which are already highly expressed in salivary glands. In contrast, *in planta* silencing of these highly expressed genes reduces aphid fecundity (Pitino and Hogenhout 2012).

Since gene silencing by dsRNA injection into *A. pisum* does not last long (Guo et al. 2014; Naessens et al. 2015; Pan et al. 2015; Wang et al. 2015a; Wang et al. 2015b) and was not effective in our hands, it is essential to develop Host-Induced-Gene-Silencing (HIGS) in order to better characterize effector candidates. *Agrobacterium* mediated transient expression system could be used with vectors that would express dsRNA in the plants (Karimi et al. 2002). Binary silencing vectors were used in several studies and led to silencing of *M. persicae* salivary genes (Pitino et al. 2011; Pitino and Hogenhout 2012). Gene expression was reduced by up to 60% in the case of *C002* silencing by stable *Arabidopsis* transformants expressing dsRNA (Pitino et al. 2011). Gene silencing would offer the possibility to characterize the functions of genes that do not show phenotype when they are expressed. Moreover, compared to silencing by dsRNA injection, effect of HIGS would be maintained longer and would let us follow aphid performances from larval stage to adulthood, and even transgenerational persistence of silencing as demonstrated in the system *Arabidopsis*-*M. persicae* (Coleman et al. 2015).

We started to perform HIGS of *Ap25* in pea plants using the binary silencing vector pB7GWIWG2(II) (Karimi et al. 2002) but no phenotype was observed. The silencing was not confirmed because qPCR of targeted gene seemed to be disturbed by the possible presence of dsRNA in the aphid body. A new construction, with restricted fragment of *Ap25* gene, would allow to measure the expression level based on a part of the gene which is absent from the dsRNA structure. Moreover, relative expression assessed by qPCR seems to be highly dependent of reference genes that are used. So, further tests to find stable references are needed.

Other methods of producing dsRNA could be envisaged such as Virus-Induced-Gene-Silencing (VIGS) developed for *P. sativum* with the Bean Pod Mottle Virus (Meziadi et al. 2016), the Pea Early Browning Virus (Constantin et al. 2004) and the White clover mosaic virus (Ido et al. 2012). After virus inoculation, the inserted fragment of gene of interest is replicated and spread in the plant as part of the recombinant virus and dsRNA are produced (Lange et al. 2013).

Finally, another way to screen effector phenotype within the pea aphid complex would be to create stable transgenic plants expressing effectors or their dsRNA, allowing to study aphid performances as well as plant phenotypes in a long-lasting interaction. Although some existing possibilities for legume plants (Pereira and Erickson 1995; Trieu et al. 2000; Svabova et al. 2005; Cosson et al. 2006), this method may be time-consuming.

### ***Further functional characterization of aphid effectors***

These different functional tools would be useful to screen phenotypes of both highly conserved candidate genes and Aphididae specific candidate genes that evolved faster (chapter II). Then, to examine the role of fast-evolving effector genes in plant adaptation, it would be necessary to characterize their role in different aphid species. Some fast-evolving Aphididae effectors were shown to act in a species-specific manner. Indeed, *in planta* expression of *C002*, *Mp1* and *Mp2* from *M. persicae* increases *M. persicae* performances whereas *A. pisum* homologues of these genes do not (Pitino and Hogenhout 2012), suggesting their specialized functions adapted to the host plant species. Because *Ap25* is a fast-evolving gene (chapter II) and appears to be important for *A. pisum* feeding (Guy et al. 2016), we planned to examine its role in plant adaptation. So, we started to investigate the role of *M. persicae* homologue of *Ap25* (named *Mp25*). Comparing the functions of this gene between *A. pisum* and the generalist aphid *M. persicae* could be one way to examine whether it acts in a species-specific manner and has evolved during arms race with host plant defenses. In this perspective, several experiments were initiated.

First, we wanted to test whether *Mp25* functions as an effector in *M. persicae*-host plant interactions. Thus, we stably transformed *A. thaliana* plants with *Mp25* expression and silencing

constructs. These lines will be soon used in performance assays with *M. persicae*. If *Mp25* is indeed an effector, we expect a better performance of *M. persicae* when *Mp25* is expressed and a lower performance when it is silenced.

Second, we aimed to test whether *Ap25* and *Mp25* act as effectors in a species-specific manner. To do so, the effect of transient *Mp25* expression on *A. pisum* performances on pea is currently examined. Moreover, *Arabidopsis* transgenic lines expressing *Ap25* were created, and the effect of *Ap25* expression on *M. persicae* performances on *Arabidopsis* will be soon studied. Like the previous study showed (Pitino and Hogenhout 2012), we expect that the expression of a gene of one species will not impact performances of the other species on its host. This would suggest that both genes have evolved to targets their specific host plants.

*Agrobacterium* mediated transient gene expression in pea represents an outstanding advance for the characterization of *A. pisum* salivary effectors that we consider as candidates for plant adaptation. However, previous population genetic and experimental works (Jaquiéry et al. 2012; Nouhaud et al. 2014; Sanchez-Arcos et al. 2016) also suggest that biotype-specific salivary effectors may have evolved to allow the pea aphid biotypes to feed on their specific plants. Those effectors, might have biotype-specific polymorphisms or may be highly expressed in one biotype compared to the others. Thus, to understand their role in plant specialization of the pea aphid biotypes, we will have to study them in several biotype-plant associations.

Therefore, to validate effector function across different biotypes, we will need tools to express transgene in other host plants of the complex in addition to pea. Among these plants, the stable transformations of *Medicago truncatula* (Trieu et al. 2000) and *Medicago sativa* (Pereira and Erickson 1995) are possible. Expression of effector genes using viruses will be another direction to explore as it is effective in *M. truncatula* and *Lathyrus odorata* using the Pea Early Browning Virus (Constantin et al. 2004). These possibilities to use different legume plants for transient expression mediated by *Agrobacterium*, viruses or stable transformation may offer new tools to study *A. pisum*-plant interactions. The development of these techniques will be the main challenge if we aim to characterize genes involved in the specialization of biotypes to each host legume plant. In addition, VIGS and HIGS could be useful tools to characterize the functions of expanded salivary gene families that we previously identified in *A. pisum* genome (chapter II). By using a conserved sequence, it will allow to silence several gene copies in the same time. However, to understand the function of each family member, we need to examine each member independently and this will probably be easier to achieve by expression of each protein separately rather than silence them.

### ***In-depth characterization of the identified effectors***

After determining the salivary effectors that affect the outcome of plant-aphid interactions, next step will be to understand their function and how they interact with host plant defenses. Protein-protein interaction analyses as Yeast-Two-Hybrid screening, Co-Immunoprecipitation (Lee 2007) or Bimolecular Fluorescence Complementation (BifC) (Kerppola 2006) are commonly used to identify plant targets of pathogen effectors (Shan et al. 2008; Sugio et al. 2011) as well as aphid effectors (Rodriguez et al. 2017). Identification of effector targets may help to reveal which plant defense pathways are targeted and suppressed by aphids. In addition, exploring the specificity of aphid effector-plant target interactions may provide insights on the involvement of effectors in host plant adaptation of aphids.

Next, do go further in the characterization of effectors functions, specific techniques may be required to examine if aphid effectors, as other pathogen effectors, degrade, cleave, sequester, phosphorylate/ dephosphorylate, or again inhibit their targets (Büttner 2016). As the *A. pisum* effector set contains many metalloproteases (such as ACE and apN, chapter II), we may analyze peptidase substrates in host plants and track the extent of cleavage by gel-based methods and N-terminal identifications (Agard and Wells 2009).

Because I had an opportunity to visit Saskia Hogenhout lab for 5 months, we envisaged to exploit *Arabidopsis-M. persicae* system to identify the targets of *A. pisum* effector Ap25. We hypothesized that Mp25 and Ap25 target orthologous proteins in *Arabidopsis* and legumes, respectively. So, we rationalized that identification of Mp25 target(s) in *Arabidopsis* will facilitate the identification of Ap25 target(s) in legumes, such as *P. sativum*. *Arabidopsis* Y2H library was available in the Hogenhout lab and knockout mutants are easily available from The Arabidopsis Information Resource (TAIR). Hence, we thought that this strategy would accelerate the identification of Ap25 target(s). Unfortunately, screening of the *Arabidopsis* library revealed only 4 potential Mp25 interactors that proved to be false positive by further Yeast-Two-Hybrid and BifC experiments. So, a second *Arabidopsis* library screening will be started, after confirming the effector function of Mp25, with particular efforts to increase efficiency using the Matchmaker Gold Yeast Two-Hybrid System (Clontech) which allows high performances thanks to a combination of a new yeast strain (Y2H Gold), stringent reporters, easy-to-use libraries, and high-level expression vectors. This approach would allow to identify plant targets of effectors that are found in both aphid species and could be extended to other top candidate effectors from *A. pisum*.

## 5. Supplementary materials

**Table S1.** Aphid clones used in this study.

Clone	Color	Plant origin	Location	Collection date	Secondary symbiont
Ar_Po_58	Green	Pea	Ardillières, France	November 2011	none
Ar_Po_28	Green	Pea	Ardillières, France	November 2011	<i>Rickettsia+Serratia symbiotica</i>
P123	Green	Pea	St Prex, Switzerland	May 2012	none
S1PS02	Green	Pea	Mauze-le-Mignon, France	April 1999	none
T8005	Pink	Clover	Saint-Augustin, Canada	August 2010	<i>Regiella insecticola</i>
YR2	Pink	Clover	York, Great Britain	December 2002	<i>Regiella insecticola</i>
L9Ms14	Pink	Alfalfa	Bugey, France	August 2011	<i>Hamiltonella defensa</i>

**Table S2.** Bacteria and plasmids used in this study.

Bacteria and plasmids	Features	Reference or source
<b>Bacteria</b>		
<i>Escherichia coli</i> Top 10	F- <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\Phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74 recA1 araD139 <math>\Delta</math>(<i>araleu</i>)7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG</i></i>	ThermoFisher Scientific
<i>Agrobacterium tumefaciens</i> C58C1	Rif <sup>r</sup> , harbors pGV2260 (pTiB6S3 $\Delta$ T-DNA)	(Deblaere <i>et al.</i> , 1985) (Koncz and Schell 1986, Holsters <i>et al.</i> , 1980)
<i>Agrobacterium tumefaciens</i> GV3101	Rif <sup>r</sup> , harbors pMP90 (pTiC58 $\Delta$ T-DNA)	(Lazo <i>et al.</i> , 1991)
<i>Agrobacterium tumefaciens</i> AGL-1	Rif <sup>r</sup> , harbors plasmid pTiBo542 $\Delta$ T-DNA	
<b>Plasmids</b>		
pDONR207	Gateway vector for entry clone construction, Gent <sup>r</sup> , Cm <sup>r</sup>	Invitrogen
pENTRY-eGFP	pDONR207 derivative, eGFP ORF cloned by BP reaction, Gent <sup>r</sup>	This study
pENTRY-GUSi	pDONR207 derivative, GUS ORF cloned by BP reaction, Gent <sup>r</sup>	This study
pENTRY-Ap25	pDONR207 derivative, ACYPI009919 <sub>67-435</sub> ORF from pea aphid line Ar_Po_58 cloned by BP reaction, Gent <sup>r</sup>	This study
pENTRY-ApC002	pDONR207 derivative, ACYPI008617 <sub>70-660</sub> ORF from pea aphid line Ar_Po_58 cloned by BP reaction, Gent <sup>r</sup>	This study
pENTRY-ApMIF	pDONR207 derivative, ACYPI002465 ORF from pea aphid line LL01 cloned by BP reaction, Gent <sup>r</sup>	This study
pEAQ-HT-DEST1	Empty vector, Kan <sup>r</sup> , Cm <sup>r</sup>	(Sainsbury, <i>et al.</i> , 2009)
pEAQ-HT-DEST1-eGFP	p35S::eGFP Kan <sup>r</sup> , Cm <sup>r</sup>	This study
pEAQ-HT-DEST1-GUSi	p35S::GUS Kan <sup>r</sup> , Cm <sup>r</sup>	This study
pEAQ-HT-DEST1-Ap25	p35S::ACYPI009919 <sub>67-435</sub> , Kan <sup>r</sup> , Cm <sup>r</sup>	This study
pEAQ-HT-DEST1-ApC002	p35S:: ACYPI008617 <sub>70-660</sub> , Kan <sup>r</sup> , Cm <sup>r</sup>	This study

**References**

- Deblaere, R., Bytebier, B., De Greve, H., Deboeck, F., Schell, J., Van Montagu, M., *et al.*, (1985). Efficient octopine Ti plasmid-derived vectors for *Agrobacterium*-mediated gene transfer to plants. *Nucleic Acids Res* 13, 4777-4788. doi: 10.1093/nar/13.13.4777
- Holsters, M., Silva, B., Van Vliet, F., Genetello, C., De Block, M., Dhaese, P., *et al.* (1980). The Functional Organization of the Nopaline *A. tumefaciens* Plasmid pTiC58. *Plasmid* 3, 212-230. doi: 10.1016/0147-619X(80)90110-9
- Koncz, C., and Schell, J. (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* 204, 383–396. doi: 10.1007/BF00331014
- Lazo, G.R., Stein, P.A., Ludwig R.A. (1991). A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Nat Biotechnol* 10, 963-967. doi: 10.1038/nbt1091-963
- Sainsbury, F., Thuenemann, E.C., and Lomonosoff, G.P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J* 7, 682-693. doi: 10.1111/j.1467-7652.2009.00434.x

**Table S3.** Pea cultivars used in this study.

Cultivars and Groups	Other names	Phylogenetic group <sup>a</sup>	CRB code <sup>b</sup>	Country of origin <sup>c</sup>	Registration year	Reference
ZP3508	Champagne	II	CRB385	FRA	-	(Baranger <i>et al.</i> , 2004)
ZP3514	DP	II	CRB390	-	-	(Baranger <i>et al.</i> , 2004)
ZP750	552	-	CRB130	USA	-	(Wicker <i>et al.</i> , 2003)
ZP747	90-2079	-	CRB119	USA	1992	(Kraft JM, 1992)
ZP748	90-2131	-	CRB135	USA	1992	(Wicker <i>et al.</i> , 2003), (Kraft JM, 1992)
ZP1109	AeD99OSW-50-2-5	-	-	FRA	-	(Roux-Duparque <i>et al.</i> , 2004 ; Moussart <i>et al.</i> , 2007)
ZP793	DSP (Dark Skin Perfection)	-	CRB134	USA	1995	-
AP3783	Enduro	-	-	FRA	2007	-
ZP3535	ISARD	-	-	FRA	2005	-
ZP1130	PI180693	-	CRB128	DEU	-	(Wicker <i>et al.</i> , 2003)
ZP690	PUGET	VI sgIII	CRB412	GBR	-	(Baranger <i>et al.</i> , 2004)
ZP1124	J1296	VI sgIII	CRB448	FRA	-	(Baranger <i>et al.</i> , 2004)
AP3830	FP (CE101)	VII	CRB383	FRA	-	(Baranger <i>et al.</i> , 2004)
ZP3495	BACCARA	VIII sgI	CRB423	FRA	1991	(Wicker <i>et al.</i> , 2003)
WP1018	EIFFEL	VIII sgI	-	FRA	1993	(Baranger <i>et al.</i> , 2004)
ZP3664	SOLARA	VIII sgII	-	HOL	1986	(Baranger <i>et al.</i> , 2004)
ZP3570	TERESE	VIII sgII	CRB413	DEN	1988	(Baranger <i>et al.</i> , 2004)

a: as described in Baranger *et al.* 2004, sg (sub-group); b: Line codes in the Biological resource center (CRB) database of INRA Dijon, France; c: Country codes ISO3166-1, FRA (France), USA (United states of America), DEU (Germany), GBR (Great Britain), Hol (Holland), DEN (Denmark).



## References

- Baranger, A., Aubert, G., Arnau, G., Laine, A.L., Deniot, G., Potier, J., *et al.*, (2004). Genetic diversity within *Pisum sativum* using protein- and PCR-based markers. TAG. Theoretical and applied genetics. *Theoretische und angewandte Genetik* 108, 1309-1321. doi: 10.1007/s00122-003-1540-5
- Kraft JM. Registration of 90-2079, 90-2131, and 90-2322 pea germplasms (1992). *Crop Sci.* 32, 1076. doi: 10.2135/cropsci1992.0011183X003200040063x
- Moussart A., Devaux C., Muel F., Pilet-Nayel M., Baranger A., Tivoli B. *et al.*, (2007). Improving partial resistance to *Aphanomyces* root rot in GSP breeding program. 3<sup>rd</sup> International *Aphanomyces* Workshop on Legumes, France.
- Roux-Duparque M., Boitel C., (2004). Breeding peas for resistance to *Aphanomyces* root rot: current main outputs of three breeding programmes. Eur Conf Grain Legumes, France.
- Wicker, E., Moussart, A., Duparque, M., and Rouxel, F. (2003). Further contributions to the development of a differential set of pea cultivars (*Pisum sativum*) to investigate the virulence of isolates of *Aphanomyces euteiches*. *Eur J Plant Pathol* 109, 47-60. doi: 10.1023/A:1022020312157

**Table S4.** Primers used in this study.

Name	Sequence (5'→3') <sup>a</sup>
ATTB1	ACAAGTTTGTACAAAAAAGCAGGC
ATTB2	ACCACTTTGTACAAGAAAGCTGGG
GFP-Fw	<u>AAAAAGCAGGCTCCACCATGGT</u> GAGCAAGGGCGAGGAGCTG
GFP-Rv	<u>AGAAAGCTGGGTGTTACTT</u> GTACAGCTCGTCCATGCCGAG
GUS-Fw	<u>AAAAAGCAGGCTCCACCATGTTACGTCCTGTAGAAACC</u>
GUS-Rv	<u>AGAAAGCTGGGTGTCATTGTTTGCCTCCCTGCTG</u>
AP25-Fw	<u>AAAAAGCAGGCTCCACCATGACAAAAACTGAAAAATCTGAC</u>
AP25-Rv	<u>AGAAAGCTGGGTTTATTTTTTGTTCATTTTT</u>
APC002-Fw	<u>AAAAAGCAGGCTCCACCATGGATTGGTCTGCCGCTGAACC</u>
APC002-Rv	<u>AGAAAGCTGGGTTTAAAAACGTCTGAAG</u>

a: underlined and bold sequences indicate *att* sites for cloning purposes and Kozak consensus, respectively.

**Table S5.** Primers used to clone salivary genes in expression vectors

<b>Primer name</b>	<b>Sequence (5'→3')<sup>a</sup></b>
<i>Ap490-F</i>	<u>AAAAAGCAGGCTTGGCC</u> <b>ATG</b> AATAAAAAGAACCCGCACTGTG
<i>Ap490-R</i>	<u>AGAAAGCTGGGTCTACTTGTGAGCTTGGACG</u>
<i>Ap1606-F</i>	<u>AAAAAGCAGGCTTGGCC</u> <b>ATG</b> TATTCTATTGAAACTACAAATTTTCAC
<i>Ap1606-R</i>	<u>AGAAAGCTGGGTGATGATCAGCAGAAAAACATTTCG</u>
<i>Ap56-F</i>	<u>AAAAAGCAGGCTTGGCC</u> <b>ATG</b> AGTTTGTTCACCGATCCC
<i>Ap56-R</i>	<u>AGAAAGCTGGGTTTATCCGTTGTTGGT</u> CGAAG
<i>Ap223-F</i>	<u>AAAAAGCAGGCTCCACC</u> <b>ATG</b> ATTTTTTTAAGCAATA
<i>Ap223-R</i>	<u>AGAAAGCTGGGTTTATGTATCAATTTTAAATAC</u>
<i>Ap7553-F</i>	<u>AAAAAGCAGGCTCCACC</u> <b>ATG</b> TTGAAACATATTATTGTG
<i>Ap7553-R</i>	<u>AGAAAGCTGGGTTCAA</u> AAGAAACTGTAAGCATTAC
<i>AttB1</i>	ACAAGTTTGTACAAAAAAGCAGGC
<i>AttB2</i>	ACCACTTTGTACAAGAAAGCTGGG

<sup>a</sup>underlined sequences correspond to *att* adaptors and Kozak consensus for cloning, bold sequences indicate gene specific regions.

**Table S6.** Characteristics of KQY and KHI salivary genes

Group	Gene name	Protein ID	Protein length	Exon number	Scaffold	Saliva proteomics	SG proteomics	Effector candidate set	Expression rank	Orthology level	Gene copy
KHI	LOC103307615	XP_008187472.1	189	3	GL350625	N	Y	SG up-regulated	55	Aphididae	Multiple
KHI	ACYP146154	NP_001280397.1	168	16	GL349740	N	Y	SG up-regulated	60	Aphididae	Multiple
KHI	LOC103308549	XP_008180313.1	163	8	GL349740	N	N	SG up-regulated	61	Aphididae	Multiple
KHI	LOC100165341	XP_008179328.1	151	11	GL349695	N	Y	SG up-regulated	66	A. pisum specific	Single
KHI	Ap7553	XP_008189863.1	170	13	GL349623	Y	Y	SG up-regulated	100	A. pisum specific	Single
KHI	LOC100572837	XP_016660699.1	187	13	GL350003	N	Y	SG up-regulated	111	A. pisum specific	Single
KHI	LOC100571631	XP_008181302.1	166	12	GL349629	N	Y	SG up-regulated	150	A. pisum specific	Single
KHI	LOC100302383	NP_001156448.1	205	3	GL354397	N	Y	SG up-regulated	202	A. pisum specific	Single
KHI	LOC100159750	XP_008182706.1	175	12	GL349884	Y	Y	SG up-regulated	280	A. pisum specific	Single
KHI	LOC100570234	XP_016657694.1	131	6	GL349740	N	N	SG up-regulated	447	A. pisum specific	Single
KHI	LOC100534636	NP_001191953.1	189	14	GL349658	N	N	SG up-regulated	625	Aphididae	Multiple
KQY	LOC100302480	NP_001156816.1	147	1	GL365281	N	Y	SG up-regulated	13	A. pisum specific	Single
KQY	Ap223	NP_001155914.1	204	37	GL349767	N	Y	SG up-regulated	18	A. pisum specific	Single
KQY	LOC100302370	NP_001156343.1	245	32	GL349709	N	Y	SG up-regulated	25	A. pisum specific	Single
KQY	LOC103307616	XP_008181764.1	224	6	GL349825	N	Y	SG up-regulated	52	A. pisum specific	Single
KQY	LOC100302403	NP_001156609.1	120	24	GL349643	N	Y	SG up-regulated	59	A. pisum specific	Single
KQY	NV22	NP_001156835.1	154	1	GL349784	N	N	SG up-regulated	87	A. pisum specific	Single
KQY	LOC100573495	XP_016662973.1	161	17	GL350443	N	Y	SG up-regulated	88	A. pisum specific	Single
KQY	LOC100302376	NP_001156435.1	192	24	GL349725	N	Y	SG up-regulated	122	A. pisum specific	Single
KQY	LOC100301916	NP_001153885.1	139	15	GL349642	N	Y	SG up-regulated	151	A. pisum specific	Single
KQY	NV12	NP_001156817.1	229	19	GL349717	N	Y	SG up-regulated	158	A. pisum specific	Single
KQY	LOC107884599	XP_016662560.1	136	8	GL350321	N	N	SG up-regulated	205	A. pisum specific	Single
KQY	LOC100302439	XP_016659404.1	158	9	GL349872	N	Y	SG up-regulated	233	A. pisum specific	Single
KQY	LOC100161848	NP_001156130.1	224	2	GL351735	N	N	SG up-regulated	301	A. pisum specific	Single
KQY	LOC107882999	XP_016657779.1	100	5	GL349745	N	N	SG up-regulated	344	A. pisum specific	Single
KQY	LOC107882998	XP_016657778.1	112	7	GL349745	N	N	SG up-regulated	440	A. pisum specific	Single

Effector candidate set was determined in chapter II, and salivary gland expression rank refers to SG- up-regulated effector gene set. Grey color indicates the two genes that were functionally analyzed.

Figure S1. KQY protein alignment

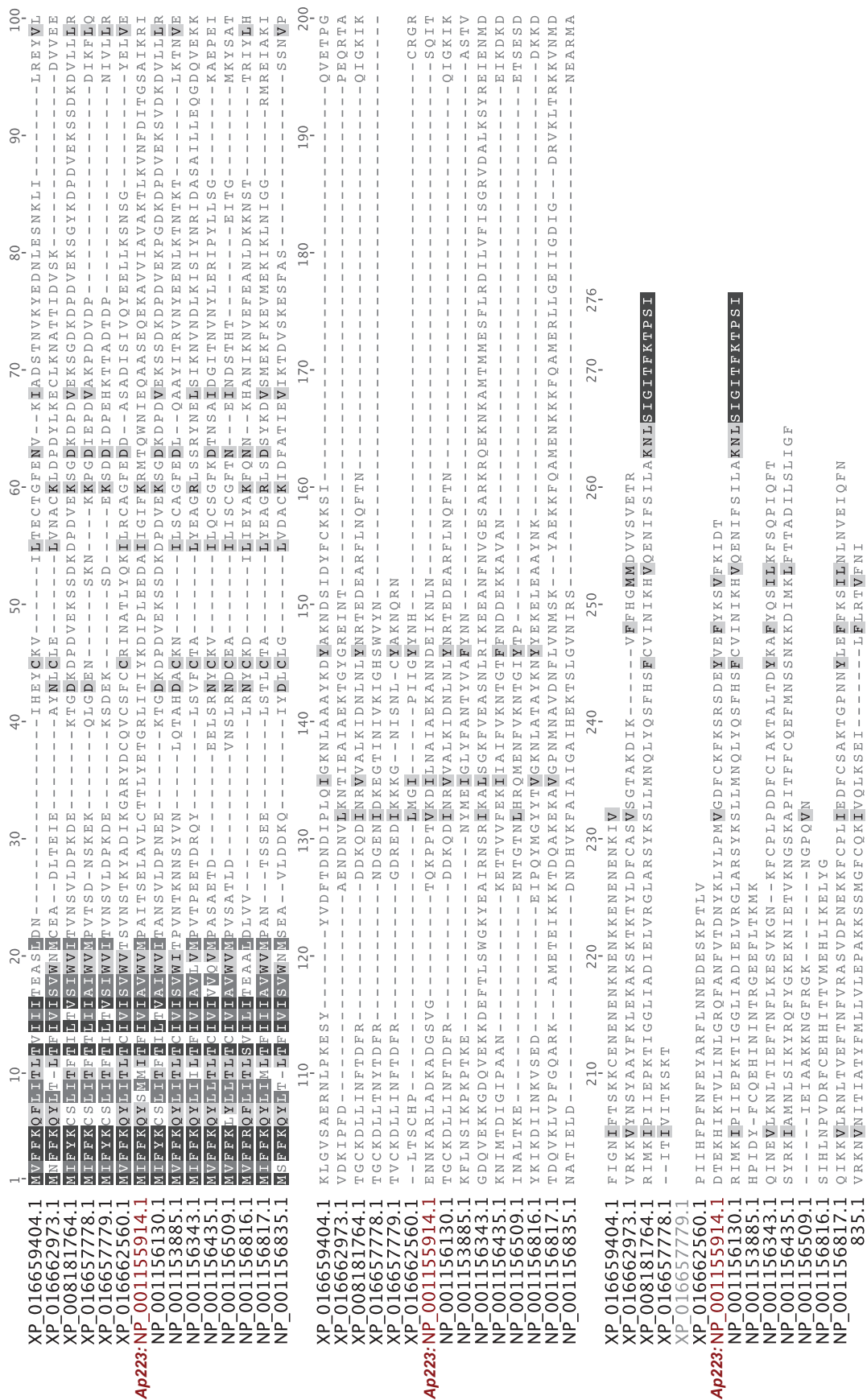
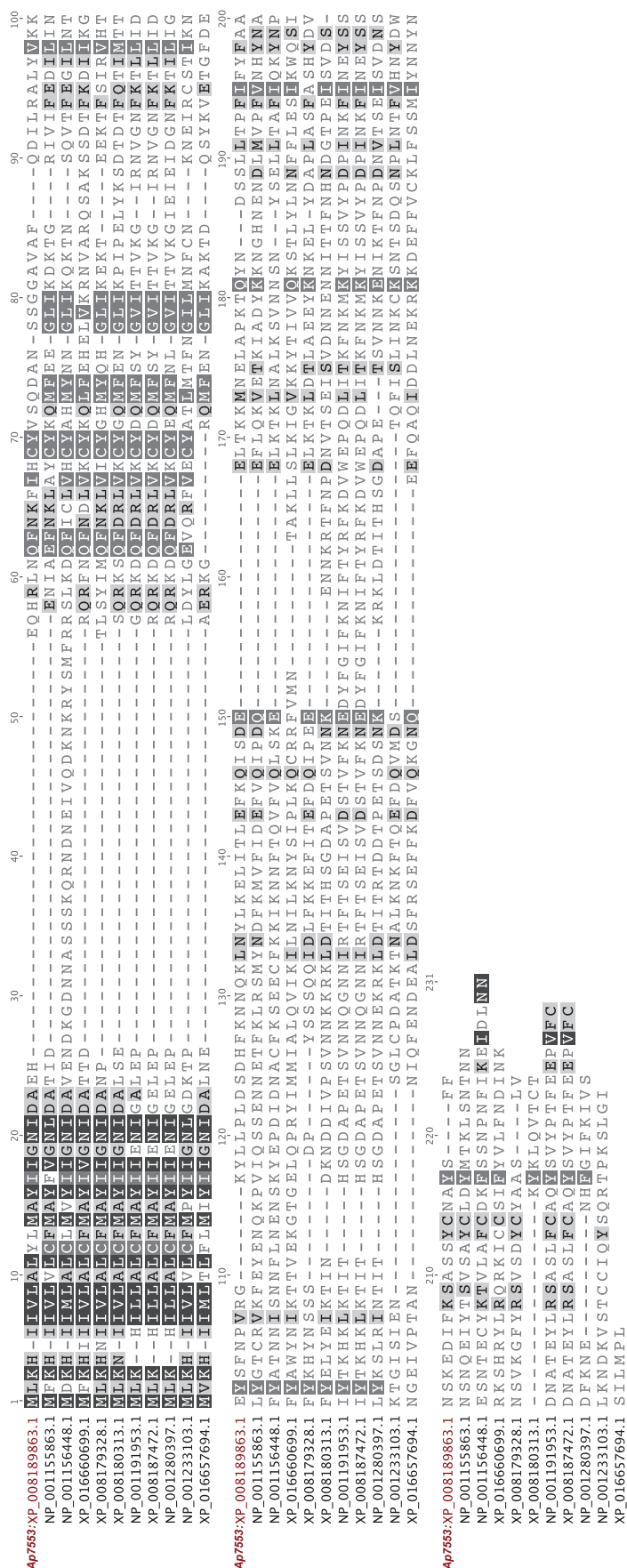


Figure S2. KHI protein alignment



## IV. Discussion générale et perspectives

Cette discussion générale est organisée autour des trois axes des travaux de thèse : l'identification des effecteurs salivaires du puceron du pois, les patrons d'évolution de ces effecteurs et leur caractérisation fonctionnelle dans le système puceron du pois-plante hôte. Elle exposera les principales avancées apportées, les limites des différentes analyses et les perspectives qu'ouvrent mes travaux de thèse.

### 1. Effecteurs salivaires du puceron du pois

#### 1.1 Identification des effecteurs candidats

Les travaux de ces dernières années montrent que les insectes herbivores, tout comme les pathogènes de plante, secrètent des effecteurs dont certains manipulent les défenses de l'hôte, leur permettant ainsi d'améliorer leur survie et reproduction (Hogenhout et al. 2009; Hogenhout and Bos 2011; Rodriguez and Bos 2012; Stuart 2015; Kaloshian and Walling 2016). Chez le puceron du pois, un premier catalogue d'effecteurs salivaires a été établi grâce à une approche combinant le séquençage des transcrits (*Expressed Sequence Tags*) et l'analyse protéomique des glandes salivaires (Carolan et al. 2011). Parmi les gènes identifiés, 324 encodent des protéines prédites comme sécrétées. Une nouvelle analyse du transcriptome des glandes salivaires par pyroséquençage 454 réalisée par des collaborateurs a prédit un plus grand nombre d'effecteurs potentiels sur la base des gènes exprimés identifiés (données non-publiées), soulignant ainsi l'incomplétude du premier catalogue. De plus, lors de différentes analyses fonctionnelles des effecteurs salivaires, il a été montré que certains étaient exprimés de manière spécifique dans les glandes salivaires (Pan et al. 2015; Wang et al. 2015a; Wang et al. 2015b).

C'est pourquoi un premier objectif de ma thèse a été de définir un set plus complet d'effecteurs salivaires du puceron du pois. Ainsi, l'expression des gènes dans les glandes salivaires et l'intestin d'*A. pisum* a été mesuré par RNA-seq, permettant une plus grande profondeur de séquençage et une quantification de l'expression de l'ensemble des transcrits. Cela a permis de déterminer dans un premier temps les gènes exprimés dans les glandes salivaires et dans un deuxième temps, ceux qui sont surexprimés dans celles-ci par rapport à l'intestin (chapitre II). La prédiction de motifs de sécrétion

tels que les peptides signaux ou des signaux de sécrétion non-canoniques a permis d'identifier 3603 gènes exprimés dans les glandes salivaires encodant des protéines potentiellement sécrétées par le puceron. Parmi ces gènes, 740 se sont révélés être surexprimés dans les glandes salivaires par rapport à l'intestin. Ces deux sets de gènes formant respectivement les effecteurs candidats exprimés dans les glandes salivaires (3603 gènes) et les effecteurs candidats surexprimés dans les glandes salivaires (740 gènes).

Le nombre d'effecteurs salivaires candidats du puceron du pois a donc été considérablement augmenté par rapport à la précédente étude (Carolan et al. 2011). L'utilisation de la technique RNA-seq, plus puissante et plus sensible que le séquençage d'EST, est à l'origine de la différence entre le premier catalogue de 324 effecteurs et le nombre plus important d'effecteurs candidats identifiés par ces travaux. En plus de cela, la part des gènes de ménage – sécrétés pour le fonctionnement des glandes salivaires – présents parmi les potentiels effecteurs a pu être réduite grâce à la comparaison de l'expression avec l'intestin. Ceci peut alors expliquer le faible chevauchement entre les effecteurs surexprimés identifiés dans notre étude et ceux prédits par Carolan et al. (2011).

Malgré l'utilisation de la méthode RNA-seq, l'expression spécifique des effecteurs dans les glandes salivaires devra être validée par la comparaison avec d'autres organes du puceron. Ainsi, le séquençage des organes reproducteurs, antennes, ou encore pattes, dont certains sont en cours de réalisation dans notre équipe pourra permettre d'affiner le set d'effecteurs candidats surexprimés dans les glandes salivaires. Toutefois, certains effecteurs peuvent être non-spécifiques et donc exprimés dans d'autres organes, comme c'est le cas pour le gène *Mif1* qui influe sur l'alimentation des pucerons et réprime les défenses de la plante (Naessens et al. 2015). En effet, le gène *Mif1* est exprimés dans les glandes salivaires (Naessens et al. 2015, chapitre II), mais son expression est aussi détectée dans le reste du corps (Dubreuil et al. 2014; Naessens et al. 2015), l'intestin (chapitre II) et les hémocytes des pucerons (Dubreuil et al. 2014). De plus, dans notre étude, seule l'expression des gènes salivaires a été prise en compte pour définir le set des effecteurs « spécifiques » des glandes salivaires (ici surexprimés par rapport à l'intestin), mais des gènes exprimés dans différents organes et produisant des transcrits alternatifs, dont certains seraient spécifiques des glandes salivaires, pourraient également participer au set d'effecteurs spécifiques de ce tissu. L'analyse des transcrits alternatifs dans les glandes salivaires ainsi que dans d'autres tissus pour lesquels nous avons (ou allons avoir) des données de séquençage permettrait alors d'identifier ce type d'effecteurs.

Néanmoins, le fait que presque tous les effecteurs salivaires déjà caractérisés pour leur rôle dans la modulation des interactions plantes-pucerons soient surexprimés dans les glandes salivaires (chapitre II) vient étayer notre hypothèse de spécialisation des effecteurs dans les glandes salivaires. Dans ce cas, l'étude des transcrits alternatifs permettrait d'identifier, parmi les gènes effecteurs

surexprimés dans les glandes salivaires, ceux qui sont sujet à un épissage alternatif intense suggérant une diversification des fonctions de ces effecteurs en réponse à des pressions de sélection exercées par la plante hôte (Thorpe et al. 2014; Betz et al. 2016; Noon et al. 2016).

L'identification des effecteurs salivaires du puceron du pois repose également sur la prédiction des signaux de sécrétion. Bien que les outils de prédictions (SignalP, SecretomeP) soient aujourd'hui performants, reposant sur des bases de données toujours plus larges et incluant de plus en plus de protéines d'insectes, certaines protéines d'*A. pisum* ne sont pas prédites comme sécrétées alors qu'elles le sont. Il est possible que d'autres mécanismes, encore inconnus, de sécrétion des protéines existent chez les pucerons, expliquant ainsi pourquoi certaines protéines détectées dans la salive récoltée sur milieu artificiel ne sont pas prédites comme sécrétées (chapitre II). Les mauvaises prédictions de signaux de sécrétion peuvent également être dues aux erreurs d'annotation du génome du puceron du pois. Ainsi, cinq des protéines identifiées par analyse protéomique de la salive présentent des annotations partielles (étant le plus souvent tronquées dans la partie N-terminale) ce qui peut expliquer la non prédiction d'un signal peptide. Le génome de puceron du pois (Acyr\_2.0) est actuellement fragmenté en 23 925 *scaffolds*, et plus de la moitié des *scaffolds* de taille supérieure à 150 Kilobases sont en réalité des chimères du chromosome X et des chromosomes autosomes (Jaquiery et al. 2017a). Ces erreurs d'assemblage peuvent avoir un impact sur l'annotation des gènes (notamment sur les petits *scaffolds* ou en limite de chimères X-autosomes) et donc sur l'identification des effecteurs salivaires. Un nouvel assemblage du génome d'*A. pisum*, en cours dans notre équipe, vise à réduire le nombre de *scaffolds* et à limiter leur chimérisme en séquençant de longs fragments d'ADN par la technique PacBio (Schadt et al. 2010), fragments qui seront combinés avec les *reads* Sanger ou Illumina déjà disponibles (The International Aphid Genomics Consortium 2010). L'amélioration du génome du puceron du pois offrira alors la possibilité de résoudre certaines difficultés de prédictions des signaux de sécrétions des protéines tronquées, de mieux décrire les éventuels transcrits alternatifs des effecteurs (e.g. *mapping* et annotation des exons), mais aussi, d'étudier la localisation et l'organisation sur le génome des gènes codant pour les effecteurs.

## 1.2 Apports de l'analyse protéomique de la salive

Venant compléter l'analyse transcriptomique et la prédiction de signaux de sécrétion, l'analyse protéomique de la salive récoltée sur milieu artificiel permet de détecter les protéines réellement sécrétées par le puceron. L'analyse protéomique de la salive d'*A. pisum* a confirmé la sécrétion de 51 protéines, ce qui représente une avancée considérable par rapport aux neuf protéines précédemment



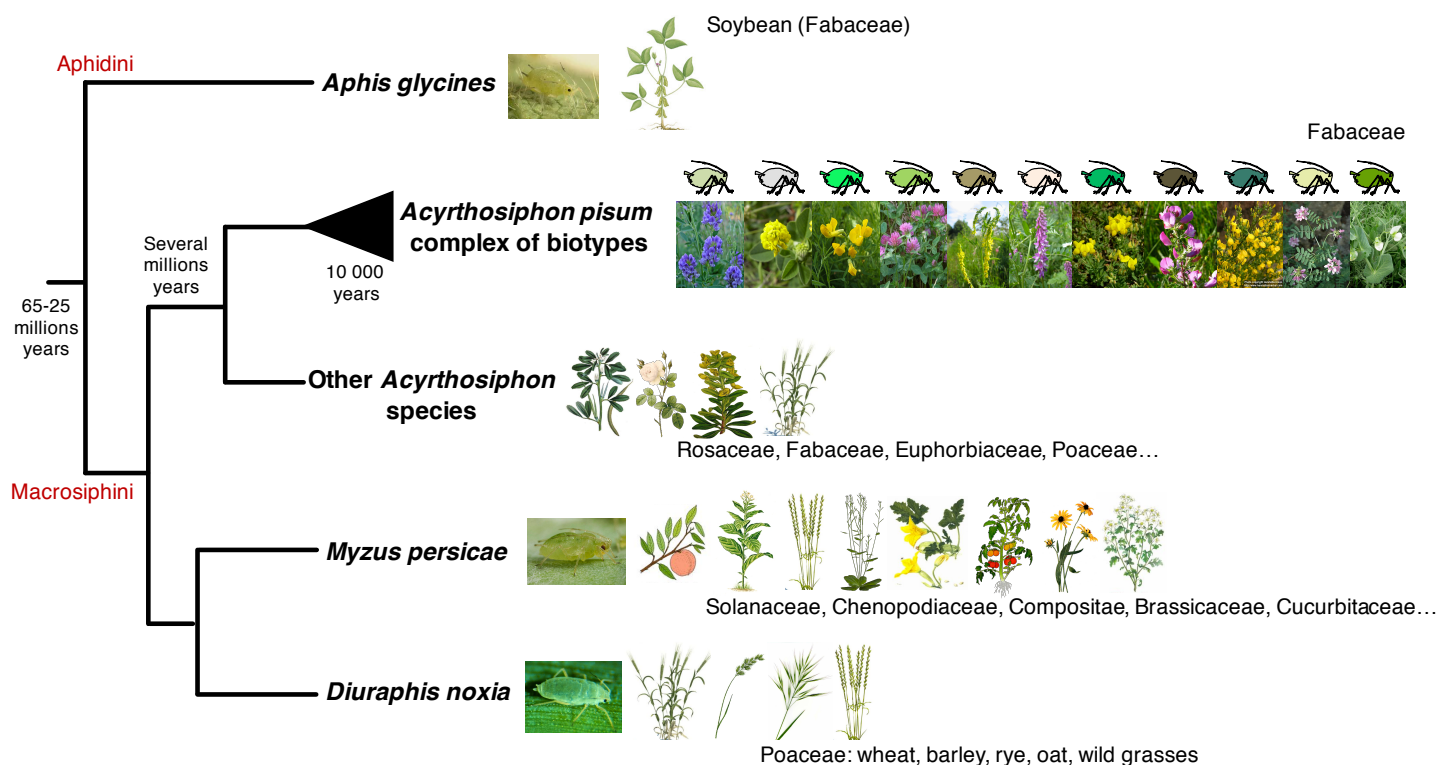
détectées dans la salive du puceron du pois (Carolan et al. 2009). Parmi les protéines identifiées, 90% sont encodées par des gènes surexprimés dans les glandes salivaires, soulignant une nouvelle fois l'importance d'inclure une analyse transcriptomique différentielle pour prédire les effecteurs salivaires du puceron du pois. Cette technique d'identification des protéines sécrétées a néanmoins ses limites.

Malgré de récents progrès techniques, l'analyse protéomique connaît encore des limites de sensibilité. Les pucerons ne sécrétant qu'une infime quantité de salive, les protéines faiblement représentées sont difficilement détectables en spectrométrie de masse, malgré des étapes de concentrations de la salive récoltée. De plus, certains effecteurs pourtant exprimés dans les glandes salivaires d'*A. pisum* et détectés dans la plante grâce à des anticorps spécifiques (*Western-Blot*), comme les protéines C002 (Mutti et al. 2008) ou Armet (Wang et al. 2015a), ne sont pas détectés dans le milieu artificiel (chapitre II, Carolan et al. 2009). La sécrétion de certains effecteurs semble alors dépendante de signaux chimiques de la plante hôte (Powell et al. 2006), ainsi une stimulation de la plante hôte peut déclencher une plus forte expression des effecteurs dans les glandes salivaires des pucerons par rapport au milieu artificiel (Guo et al. 2014; Wang et al. 2015a; Wang et al. 2015b). L'absence de corrélation positive entre le niveau d'expression des gènes salivaires et l'intensité de détection protéique dans la salive – alors qu'une telle corrélation positive existe entre expression et intensité dans les glandes salivaires – pourrait être due à un manque de puissance lié au faible nombre de protéines identifiées dans la salive, mais également à des mécanismes de régulation de la sécrétion dépendant du dialogue moléculaire avec la plante (chapitre II). Ce biais se traduit aussi par le fait que certains effecteurs n'influencent pas la *fitness* des pucerons sur milieu artificiel alors qu'ils y participent lors de l'interaction avec la plante hôte (Naessens et al. 2015; Wang et al. 2015a; Wang et al. 2015b). En l'absence d'un système « haut-débit » qui permettrait de détecter directement les effecteurs injectés par les pucerons dans la plante, l'analyse de salive à partir de milieu artificiel reste une approche valable pour confirmer la sécrétion des effecteurs candidats identifiés par des approches transcriptomiques.

Cette première partie de mes travaux de thèse a considérablement enrichi et consolidé le « catalogue » des effecteurs du puceron du pois, même si ce dernier doit continuer à être amélioré. Ce catalogue global représente ainsi une référence pour l'ensemble des travaux sur les effecteurs aphidiens.

## 2. Evolution des effecteurs salivaires et adaptation à la plante hôte

Afin d'identifier des candidats impliqués dans l'adaptation à la plante hôte parmi les effecteurs salivaires du puceron du pois, nous avons, dans un second objectif de ma thèse, exploré leurs patrons d'évolution et recherché des signatures de sélection positive. Ces traces de sélection peuvent être détectées à une échelle de temps large (macro-évolutive), au sein des différentes espèces de pucerons ayant divergé il y a plusieurs millions d'années (Kim et al. 2011; Misof et al. 2014), ou à une échelle de temps récente (micro-évolutive), au sein du complexe de biotypes d'*A. pisum*, dont le dernier ancêtre maternel commun est daté entre 16 000 et 8000 ans (Peccoud et al. 2009b). Compte tenu que le genre *Acyrtosiphon* (environ 60 espèces validées) comprend des espèces s'alimentant sur des familles botaniques aussi variées que les Rosaceae, Fabaceae (légumineuses), Euphorbiaceae ou Poaceae (Eastop 1971), le scénario le plus plausible est qu'*A. pisum* se soit tout d'abord adapté aux légumineuses puis que cette espèce ait évolué beaucoup plus récemment en un complexe de biotypes, chacun spécialisé sur une ou quelques espèces de légumineuses (Peccoud et al. 2009a) (Figure 5).



**Figure 5.** Phylogénie et gamme d'hôtes des différentes espèces de pucerons considérées dans cette thèse. La phylogénie schématique des pucerons est inspirée de celle déterminée dans le chapitre II.

## 2.1 Les effecteurs aphidiens vus à une échelle macro-évolutive

### *Les effecteurs du puceron du pois à travers la phylogénie*

Le séquençage récent de plusieurs génomes de pucerons (Nicholson et al. 2015; Mathers et al. 2017; Wenger et al. 2017) nous a permis de retracer l'histoire évolutive des effecteurs salivaires au sein des Aphididae, grâce à des études de génomique comparative. Dans un premier temps, nous avons cherché à caractériser les niveaux d'orthologie des différents effecteurs du puceron du pois identifiés par l'approche transcriptomique. Le protéome d'*A. pisum* a été comparé à 4 protéomes de pucerons, 11 protéomes d'insectes (parmi les ordres majeurs) et le protéome d'un arachnide, *Tetranychus urticae*, afin d'attribuer un niveau d'orthologie à chaque gène d'*A. pisum* en fonction des orthologues retrouvés chez les autres espèces et de leur relations phylogénétiques (chapitre II). Cette analyse a révélé des différences entre les deux sets d'effecteurs candidats (surexprimés ou exprimés dans les glandes salivaires) et le génome du puceron du pois. Les effecteurs candidats surexprimés dans les glandes salivaires sont enrichis en gènes qui sont retrouvés uniquement chez les Aphididae. Lorsque l'on regroupe les gènes spécifiques soit d'*A. pisum*, des Aphididae et des Aphidomorpha, ils composent plus de la moitié du set des effecteurs candidats surexprimés. A l'inverse, les effecteurs candidats exprimés (comprenant tous les gènes exprimés dans les glandes salivaires potentielles sécrétés) sont enrichis en gènes conservés chez les insectes et arthropodes, comptant pour plus de 50% de ce set. Ces résultats suggèrent que les effecteurs candidats surexprimés sont davantage liés au développement du mode particulier d'alimentation des pucerons et à l'évolution des glandes salivaires au cours de la diversification des Aphidomorpha, pour répondre au défi que représente l'ingestion de phloème. Chez la mouche de Hesse, *Mayetiola destructor*, dont les larves se nourrissent de tissus végétaux en induisant la formation de galles (Harris et al. 2006), la majeure partie des effecteurs salivaires candidats (*secreted salivary gland proteins* SSGPs) est spécifique de cette espèce et ne présente pas d'homologie avec les gènes des autres espèces d'insectes dont le génome est séquencé (Zhao et al. 2015). De plus, parmi la famille SSGP-71, qui contient des protéines caractérisées pour leur rôle d'effecteurs lors de l'interaction avec la plante hôte, le blé (Stuart et al. 2012; Aggarwal et al. 2014), la plupart ne sont retrouvés que chez *M. destructor* (Zhao et al. 2015). Comme plus de 50% des effecteurs surexprimés dans les glandes salivaires du puceron du pois sont spécifiques de cette espèce ou des pucerons, nous supposons que les meilleurs candidats impliqués dans l'adaptation aux légumineuses chez cet insecte se trouvent parmi ces effecteurs.

### ***Evolution des effecteurs au sein des Aphididae***

L'analyse des taux d'évolution (i.e. dN/dS) des gènes présents en copie simple au sein des espèces d'Aphididae a montré que les effecteurs candidats surexprimés évoluent en moyenne plus rapidement que ceux exprimés dans les glandes salivaires de manière non différentielle avec l'intestin et que les autres gènes du puceron du pois (chapitre II). Ceci démontre une nouvelle fois les propriétés particulières du set d'effecteurs surexprimés dans les glandes salivaires par rapport au set des effecteurs exprimés. Ainsi, leur évolution rapide suggère un rôle de effecteurs surexprimés dans l'adaptation à la plante hôte chez les Aphididae. Une évolution rapide des effecteurs est décrite chez de nombreux pathogènes de plantes aussi bien procaryotes qu'eucaryotes (Hogenhout et al. 2009). Les effecteurs bactériens appartenant au système de sécrétion de type III (*type three secretion system* T3SS) ou encore les effecteurs RXLR des oomycètes présentent des domaines polymorphes dont les taux d'évolution sont différents en fonction de leur rôle dans la sécrétion ou l'activité de l'effecteur. Ainsi, les domaines C-terminaux qui assurent l'activité de ces effecteurs évoluent plus rapidement que les autres domaines et sont bien souvent sélectionnés positivement au cours de la coévolution avec la plante hôte (Ma et al. 2006; Win et al. 2007).

Néanmoins, l'analyse des taux d'évolution des effecteurs salivaires précédemment caractérisés chez différentes espèces de pucerons, révèle qu'à la fois des gènes évoluant rapidement et d'autres évoluant sous sélection purifiante peuvent être impliqués dans l'interaction avec la plante hôte. Les effecteurs qui sont spécifiques des Aphididae ou Aphidomorpha évoluent plus rapidement que ceux qui sont conservés chez les insectes ou arthropodes (chapitre II). Nous suggérons ici que l'évolution plus rapide des effecteurs spécifiques des pucerons résulte d'un processus de course aux armements lors de la coévolution avec les plantes hôtes. Certains d'entre eux, comme les gènes *C002*, *Mp1* ou *Mp2* sont connus pour agir de manière plante-hôte-spécifique (Pitino and Hogenhout 2012). Ainsi, *Myzus persicae* bénéficie de l'expression *in planta* de ses propres versions de ces trois gènes, alors que les versions d'*A. pisum* n'influencent pas sa *fitness*. Ceci suggère une interaction spécifique des effecteurs avec des protéines de la plante hôte sur laquelle chaque espèce de puceron est adaptée. Une telle interaction spécifique a été montrée dans le cas de l'effecteur Mp1. Rodriguez et al. (2017) ont identifié la cible de l'effecteur Mp1 comme étant la *Vacuolar Protein Sorting Associated Protein52* (VPS52) chez deux hôtes de *M. persicae*, *Arabidopsis thaliana* et la pomme de terre *Solanum tuberosum*. Alors que l'effecteur Mp1 de *M. persicae* est capable d'interagir avec les VPS52 de ses hôtes, aucune interaction n'est détectée avec les VPS52 provenant de l'orge *Hordeum vulgare* ni de *Medicago truncatula*, qui sont des hôtes respectifs des pucerons *Rhopalosiphum padi* et *A. pisum*. De plus, cette association semble spécifique de l'interaction entre *M. persicae* et ses hôtes car les protéines Mp1 de *R. padi* et *A. pisum* ne ciblent pas les VPS52 de leurs hôtes respectifs (Rodriguez et al. 2017).

L'histoire évolutive des VPS52 des plantes n'a pas encore été étudiée, mais leur évolution rapide suggérerait une coévolution antagoniste avec les gènes *Mp1* des pucerons, qui eux évoluent rapidement sous sélection positive (chapitre II, Pitino and Hogenhout 2012)

Malgré des taux d'évolution rapide chez ces effecteurs et des études suggérant leur rôle dans l'adaptation à la plante hôte chez les pucerons, seulement deux d'entre eux semblent évoluer sous sélection positive, *Mp1* et *Me10* (Pitino and Hogenhout 2012, Chapitre II). L'absence de signature de sélection positive chez les autres effecteurs à "évolution rapide" pourrait être due à certaines limites dans l'analyse du dN/dS. En effet, les régions non-alignées (gaps) ne sont pas prises en compte dans l'estimation du dN/dS alors qu'elles participent, au même titre que les SNPs (*single nucleotide polymorphisms*), à la variabilité des effecteurs au sein des différentes espèces de pucerons. Ces régions, comme certaines insertions présentes dans les séquences de *C002* et *Mp2* chez *M. persicae* (Pitino and Hogenhout 2012), pourraient donc être impliquées dans l'interaction spécifique plante hôte-puceron. Ainsi une insertion correspondant à cinq répétitions du motif NDNQGEE présente dans la séquence du *C002* de *M. persicae*, et non-retrouvée chez les autres espèces de pucerons, semble jouer un rôle dans l'interaction de *M. persicae* avec son hôte *Arabidopsis*. Effectivement, l'expression de la version mutée du gène *C002* (délétion des cinq répétitions) n'augmente pas la fitness de *M. persicae* sur son hôte alors que c'est le cas pour la version non mutée (Pitino and Hogenhout 2012). Malgré les limites inhérentes à la détection de signatures de sélection avec les méthodes actuelles, ces effecteurs évoluant rapidement n'en demeurent pas moins de bons candidats dans l'adaptation des pucerons à des plantes hôtes différentes.

Contrastant avec les gènes effecteurs évoluant rapidement, que nous supposons impliqués dans l'adaptation à la plante hôte, les effecteurs encodés par des gènes évoluant plus lentement (i.e. sous sélection purifiante plus forte) sont probablement impliqués dans des fonctions plus générales et conservées de l'alimentation des pucerons. Ainsi, ces effecteurs pourraient participer à la suppression des défenses en ciblant des voies très conservées au sein des plantes. Pour explorer cette hypothèse, il faudrait par exemple analyser les interactions plante-hôte-spécifique (comme pour les gènes *C002*, *Mp1* ou *Mp2*; Pitino and Hogenhout 2012) dans le cas des gènes très conservés déjà caractérisés comme *Armet*, *Mp10*, *Mif1* (Naessens et al. 2015; Wang al. 2015a; Wang et al. 2015b) et tester si l'expression du gène d'une espèce de puceron peut bénéficier à la fitness d'une autre espèce sur sa plante hôte. Cependant, supposer des implications différentes des effecteurs dans l'adaptation ou non à la plante hôte reste difficile. Prenons comme exemple la protection du stylet lors de sa navigation entre les tissus végétaux, qui constitue un caractère important et conservé chez les pucerons, et est assurée par la gaine sétale composée majoritairement par la protéine SHP (*sheath protein*). Ainsi, la réduction de l'expression du gène *shp* perturbe l'alimentation et diminue la fitness de plusieurs espèces

de pucerons (Abdellatef et al. 2015; Will and Vilcinskas 2015). Comme ce caractère semble impliqué dans une fonction très générale de l'alimentation des pucerons, on peut s'attendre à ce que le gène *shp* soit très conservé et évolue lentement (forte sélection purifiante pour éliminer les mutations), mais notre analyse le classe plutôt parmi les gènes évoluant rapidement, démontrant ainsi la complexité de l'évolution des interactions plantes-pucerons (chapitre II).

### ***Expansion des familles d'effecteurs chez le puceron du pois***

L'analyse protéomique de la salive a révélé qu'environ 50% des protéines détectées étaient encodées par trois familles multigéniques bien représentées dans le set d'effecteurs candidats surexprimés dans les glandes salivaires d'*A. pisum* (chapitre II). L'analyse des patrons d'évolution de ces trois familles de gènes, *cystein-rich-proteins* CRPs, *angiotensin-converting-enzyme* ACEs et *aminopeptidases-N* apN a montré que les CRPs sont spécifiques des Aphididae alors que les ACEs et apN existent chez d'autres espèces d'insectes. De plus, ces trois familles présentent une nette expansion chez le puceron du pois par rapport aux autres espèces d'Aphididae étudiées. Des expansions de gènes ont déjà été observées chez le puceron du pois dans les familles codant pour les protéines cuticulaires, les transporteurs d'acides aminés, la machinerie ARNi, les cathepsines B, les récepteurs olfactifs et les récepteurs gustatifs (Rispe et al. 2007; Smadja et al. 2009; Jaubert-Possamai et al. 2010; The International Aphid Genomics Consortium 2010; Price et al. 2011). Tout comme pour les familles des récepteurs olfactifs et gustatifs (Smadja et al. 2009) ainsi que des cathepsines B (Rispe et al. 2007), la sélection positive semble avoir participé à l'évolution de ces trois familles d'effecteurs salivaires chez le puceron du pois. Ainsi, plusieurs épisodes de sélection ont été détectés après les duplications géniques survenues chez *A. pisum* dans ces trois familles (chapitre II). De fortes pressions de sélection exercées par les plantes hôtes d'*A. pisum* pourraient être à l'origine de la sélection positive agissant spécifiquement sur des sites parmi les branches ancestrales ou gènes de cette espèce (sous sélection neutre ou purifiante dans les autres branches de la phylogénie).

Les duplications géniques sont considérées comme une source de nouveautés, permettant dans certains cas la création de nouvelles fonctions biologiques (Zhang 2003). Ainsi, chez les insectes herbivores les duplications peuvent être à l'origine d'innovations clés dirigeant l'adaptation à la plante hôte, comme c'est le cas des *nitrile-specifier proteins* (NSPs) impliqués dans la détoxification des glucosinolates des Brassicales par les Piéridés (Wheat et al. 2007; Fischer et al. 2008; Fischer et al. 2008; Edger et al. 2015). En plus de leur implication dans les mécanismes de coévolution entre plantes et insectes (Edger et al. 2015), ces duplications géniques peuvent favoriser les sauts d'hôte et ainsi contribuer à l'adaptation et spéciation des insectes herbivores. Chez le puceron *M. persicae*, une expansion massive du nombre de copies d'un gène codant pour un cytochrome P450 a semble-t-il

permis à cette espèce de coloniser un nouvel hôte, le tabac (Puinean et al. 2010; Bass et al. 2013; Bass et al. 2014). Un seul exemple d'expansion des gènes codant pour les effecteurs salivaires avait été décrit jusqu'à maintenant, chez *M. destructor*. Les SSGPs ont en effet subi une expansion massive chez la mouche de Hesse et comptent maintenant pour 7% de son génome. La famille des SSGP-71, dont deux membres sont caractérisés pour leur fonction d'effecteurs dans l'interaction avec le blé, représente même la plus large famille multigénique jamais décrite chez les arthropodes (Zhao et al. 2015). Dans les génomes de pathogènes de plantes, les effecteurs sont bien souvent organisés en clusters issus de duplications en tandem dans une même région du génome (Win et al. 2007; Haas et al. 2009; Brefort et al. 2014). Chez le puceron du pois, il est difficile d'analyser l'architecture génomique des trois familles codant pour des effecteurs candidats à cause de la fragmentation du génome. Bien que quelques membres de chaque famille semblent être localisés respectivement sur des *scaffolds* identiques (chapitre II), un nouvel assemblage sera nécessaire pour retracer l'histoire de ces expansions dans le génome du puceron du pois et peut-être identifier des clusters d'effecteurs résultant de duplication en tandem.

Ces effecteurs CRP, ACE et apN constituent néanmoins, de par leur forte représentation dans la salive, leur expansion spécifique et des signatures de sélection positive sur certaines branches, de bons effecteurs candidats dans l'adaptation du puceron du pois à ses hôtes légumineuses.

## 2.2 Les effecteurs aphidiens vus à une échelle micro-évolutive

### *Régions génomiques et gènes liés à l'adaptation du complexe de biotypes*

Au sein du complexe biotypes du puceron du pois, une étude génétique des populations sans *a priori* par balayage génomique (*genome scan*) a révélé plusieurs régions potentiellement liées à l'adaptation à la plante hôte (Jaquiéry et al. 2012). Parmi les *outliers* présentant une différenciation génétique plus forte que celle attendue sous évolution neutre entre les trois biotypes analysés (pois, luzerne, trèfle), deux sont proches de gènes codant pour des récepteurs olfactifs et trois proches de gènes salivaires (Jaquiéry et al. 2012). Ces trois gènes salivaires (ACYPI089376, ACYPI002172, ACYPI54712) sont prédits dans notre analyse (chapitre II) comme effecteurs salivaires surexprimés dans les glandes salivaires et sont tous des gènes Aphididae spécifiques sans fonctions connues. ACYPI089376 et ACYPI002172 appartiennent au même cluster d'orthologues (3 copies chez *A. pisum* appartenant au set d'effecteurs surexprimés) et ACYPI54712 est un gène en simple copie évoluant rapidement au sein des Aphididae (chapitre II). Ainsi, les gènes candidats à l'adaptation de l'espèce aux légumineuses peuvent être les mêmes que ceux potentiellement impliqués dans la spécialisation

au sein du complexe de biotypes. En plus de cette étude par Jaquiéry et al. (2012), l'analyse de données pool-seq de 3 des biotypes du complexe a été réalisée au sein de notre équipe (Nouhaud et al. 2017, en révision). Parmi les gènes se trouvant dans les *hotspots* de différenciation identifiés entre ces biotypes (régions génomiques avec une plus forte différenciation que l'attendu sous évolution neutre), il y a un net enrichissement des effecteurs salivaires, confirmant ainsi un rôle prépondérant de ceux-ci dans l'adaptation des biotypes à des plantes hôtes différentes. Néanmoins, ces deux études de génétique des populations ont été réalisées avant l'établissement du set d'effecteurs candidats plus complet (chapitre II) et sont donc basées sur le premier catalogue établi par Carolan et al. (2011) et sur les travaux réalisés chez d'autres espèces de pucerons (Harmel et al. 2008; Bos et al. 2010). Par conséquent, nous sommes actuellement en train de rechercher de nouveaux gènes effecteurs présents dans les *hotspots* ou *outliers* de différenciation entre les biotypes en utilisant le set d'effecteurs candidats défini par ces travaux de thèse. Ce qui permettra d'avoir une vue d'ensemble de tous les effecteurs candidats potentiellement liés à l'adaptation des biotypes.

Les approches de génétique quantitative représentent également des outils puissants pour identifier les régions génomiques et gènes impliqués dans les processus adaptatifs. Ainsi, des QTL (*quantitative trait loci*) liées à la préférence et à la performance des pucerons du pois ont été identifiées et révèlent une base polygénique de l'adaptation à la plante hôte au sein du complexe de biotypes (Hawthorne and Via 2001; Via and Hawthorne 2002; Via and Hawthorne 2005). Plus récemment, les clones N116 et PS01 d'*A. pisum*, appartenant respectivement aux biotypes luzerne et pois ont montré des performances contrastées sur différents génotypes *Medicago truncatula* (Kanvil et al. 2014). En effet le clone N116 est virulent sur la plupart des génotypes alors que le clone PS01 est avirulent sur les génotypes qui possèdent le gène de résistance *RAP1 – Resistance to Acyrthosiphon pisum 1* (Stewart et al. 2009; Kanvil et al. 2014). L'analyse des descendants résultant du croisement de ces deux clones a permis de révéler l'existence d'allèles de virulence et d'avirulence chez le puceron du pois (Kanvil et al. 2015). Des projets sont en cours dans l'équipe et via des collaborations pour identifier les régions génomiques sous les QTLs et tenter de révéler le(s) gène(s) contrôlant les facteurs de virulence. Les approches de génétique quantitative offrent beaucoup de possibilités dans l'identification des régions et gènes impliqués dans l'adaptation à la plante hôte des biotypes. En effet, les biotypes présentent peu d'isolement reproducteur les uns avec les autres ce qui permet de réaliser des croisements en induisant la production de formes sexuées par diminution de la photopériode (Peccoud et al. 2014). De futures analyses pourront ainsi révéler si des effecteurs salivaires candidats sont situés dans les QTL liés à l'adaptation des biotypes. Néanmoins, toutes ces approches visant à identifier les régions et l'architecture génomique de l'adaptation du complexe à ses hôtes bénéficieront grandement d'un génome réassemblé et de meilleure qualité.



Des approches de génétique des populations peuvent également être menées à une échelle plus restreinte (e.g. approche gène candidat) en ciblant les fonctions qui semblent impliquées dans l'adaptation à la plante hôte. Certains gènes de récepteurs olfactifs appartenant à des familles multigéniques spécifiques au génome d'*A. pisum* et présentant des signatures de sélection positive (Smadja et al. 2009), ont également montré des empreintes de sélection divergente par des approches de capture de gènes sur différents biotypes (Smadja et al. 2012; Eyres et al. 2017). De telles approches pourraient être conduites en ciblant les gènes effecteurs candidats à l'adaptation d'*A. pisum* aux légumineuses que nous avons identifiés lors de ces travaux de thèse (chapitre II), permettant ainsi de faire le lien entre les échelles macro et micro-évolutives.

### ***Polymorphisme et expression différentielles des gènes salivaires candidats***

En utilisant des méthodes de détection de signatures de sélection divergente (e.g. Gautier et al. 2013; Vitalis et al. 2013; Gautier 2015) les approches de génétique/génomique des populations ont déjà permis d'identifier des gènes fortement différenciés entre les biotypes adaptés à différentes plantes hôtes. Pour certains gènes les SNPs identifiés au niveau génomique induisent un changement d'acides aminés, suggérant une fonction différente de la protéine, mais pour d'autres la séquence protéique reste inchangée. La production de protéines différentes, spécifiques des biotypes, à partir d'un gène présentant des signatures de sélection divergente entre ceux-ci, suggère un rôle important de ce gène dans l'adaptation des biotypes à leur hôte spécifique. La recherche de polymorphismes non-synonymes dans les données de séquençage acquises lors des approches de génétique/génomique des populations mais également parmi les génomes reséquencés de 33 clones représentatifs de 11 biotypes du complexe (Gouin et al. 2015) disponibles dans notre équipe, permettra de faire un tri parmi les gènes salivaires candidats afin d'identifier ceux qui semblent avoir des fonctions biotypes-spécifiques, et qui sont donc potentiellement liés à la spécialisation des biotypes.

Outre ce polymorphisme spécifique des biotypes, les gènes candidats à l'adaptation peuvent présenter des expressions différentielles entre les biotypes en réponse à l'utilisation de plantes hôtes spécifiques. Une première étude a montré que des différences d'expression liées à la spécialisation des biotypes concernaient notamment des gènes salivaires et chimiosensoriels (Eyres et al. 2016) et que les gènes différentiellement exprimés en fonction de la plante hôte (universelle vs « native ») ne représentaient pas une grande proportion des différences observées entre les biotypes. Ces résultats soulignent que les gènes différentiellement exprimés entre les biotypes et potentiellement impliqués dans la spécialisation, ne sont pas nécessairement les mêmes que ceux qui montrent une plasticité lors du changement entre l'hôte « natif » et l'hôte universel.

Une nouvelle étude transcriptomique a été menée récemment dans notre équipe sur des clones appartenant aux biotypes pois et luzerne, et comparant les patrons d'expression sur l'hôte universel *Vicia faba* et sur l'hôte « natif » pour chacun des deux biotypes. Les données sont en cours de traitement mais les premiers résultats montrent très peu de gènes différentiellement exprimés en fonction de la plante comme montré par Eyres et al. (2016), et que les plus grandes différences d'expression sont observées entre les biotypes. Il ressort également des premières analyses que certains effecteurs candidats identifiés à l'échelle macro-évolutive comme les familles multigéniques CRP, ACE et apN et des gènes simples copiant évoluant rapidement (notamment *C002* et *Ap25*), sont différentiellement exprimés entre les deux biotypes. De plus, les gènes KQY/KHI (chapitre III) présentent également des différences d'expression dépendantes du biotype. Ces résultats préliminaires permettent d'émettre l'hypothèse que les différents biotypes présentent des profils d'expression des gènes effecteurs spécifiques à leur hôte, et que le fait qu'ils se développent tous sur *V. faba* refléterait plutôt une grande permissivité de cet hôte qu'une plasticité des pucerons.

### 2.3 Adaptation à la plante hôte

L'exploration des patrons d'évolution et profils d'expression des effecteurs salivaires aux différentes échelles macro et micro-évolutives révèlent différents mécanismes impliqués dans l'adaptation du puceron du pois à ses hôtes fabacées et dans la spécialisation du complexe de biotypes.

Ainsi, les familles de gènes présentant une expansion spécifique au génome d'*A. pisum* semblent jouer un rôle fondamental dans l'adaptation du puceron du pois aux légumineuses. Cette expansion particulière des familles d'effecteurs salivaires est retrouvée également chez les pucerons *Diuraphis noxia* et *Aphis glycines* mais dans une moindre mesure. Tout comme *A. pisum* ces deux espèces de pucerons peuvent être classées comme spécialiste, la première se nourrissant uniquement sur des graminées et la deuxième étant adaptée au soja (Figure 5). A l'inverse, chez le puceron généraliste *M. persicae*, les trois familles d'effecteurs salivaires sont très réduites (une seule copie pour les apN (clade 4), ACE (clade 4) et CRP). Une tendance à posséder de plus larges familles multigéniques codant pour des effecteurs salivaires chez les spécialistes que chez les généralistes paraît se dégager au sein des pucerons, néanmoins la comparaison est difficile car une seule espèce généraliste a été étudiée.

Les pucerons du pois appartenant à différents biotypes semblent aussi présenter une expression différentielle des gènes codant pour les effecteurs salivaires (Eyres et al. 2016). Mais la comparaison entre hôtes universel et « natif » montre peu de plasticité transcriptionnelle chez les biotypes de cette

espèce. En revanche, chez *M. persicae*, une grande plasticité transcriptionnelle, notamment des gènes codant pour les cathepsines B intestinales, semble impliquée dans la capacité qu'a ce généraliste à coloniser un grand nombre d'hôtes appartenant à des familles botaniques différentes (Mathers et al. 2017).

Ainsi, ces deux mécanismes (duplication génique et plasticité transcriptionnelle) semblent jouer un rôle important dans l'adaptation des pucerons à une gamme d'hôtes restreinte (spécialiste) ou plus large (généraliste), respectivement, mais cela reste une hypothèse. De plus, la néofonctionnalisation par duplications et la plasticité transcriptionnelle participent bien souvent de concours à l'adaptation des espèces. Ainsi, chez *T. urticae*, un acarien herbivore généraliste, les deux mécanismes semblent avoir favorisé l'acquisition d'une large gamme d'hôte (Grbić et al. 2011).

La consolidation du catalogue des effecteurs salivaires a permis d'identifier des effecteurs candidats potentiellement impliqués dans l'adaptation du puceron du pois aux légumineuses. Les analyses génomiques et transcriptionnelles à l'échelle des biotypes suggèrent également l'implication des effecteurs salivaires dans la spécialisation des biotypes. L'ensemble de ces analyses permettent ainsi « faire le tri » et de restreindre les études fonctionnelles aux candidats les plus probablement impliqués dans l'adaptation et la spécialisation d'*A. pisum* à ses hôtes légumineuses.

### **3. Outils et stratégies pour une meilleure compréhension des interactions moléculaires dans le système puceron du pois-légumineuses**

#### **3.1 Différents types d'effecteurs pour différentes fonctions ?**

L'étude des patrons d'évolution des effecteurs du puceron du pois réalisée lors de cette thèse révèle différents types d'effecteurs salivaires candidats, pouvant avoir des rôles variés dans l'adaptation à la plante hôte chez les pucerons. Ainsi, parmi les effecteurs candidats d'*A. pisum* on peut proposer de définir trois catégories : 1) des gènes conservés au sein des insectes/arthropodes évoluant sous sélection purifiante, 2) des gènes évoluant rapidement et retrouvés uniquement chez les pucerons et 3) d'autres enfin spécifiques d'*A. pisum* et pouvant évoluer sous sélection positive (e.g. membres de familles multigéniques, chapitre II). Plusieurs effecteurs appartenant aux deux premières catégories sont connus pour influencer positivement les performances de différentes espèces de pucerons sur leurs

plantes hôtes (Mutti et al. 2006; Mutti et al. 2008; Bos et al. 2010; Pitino et al. 2011; Atamian et al. 2012; Atamian et al. 2012; Pitino and Hogenhout 2012; Elzinga et al. 2014; Naessens et al. 2015; Wang et al. 2015a), en revanche seuls deux gènes appartenant à une famille présentant une expansion spécifique d'*A. pisum* sont pour l'instant caractérisés (Wang et al. 2015b).

Ces différents types d'effecteurs soulèvent de nombreuses questions quant à leur fonctions et leur rôle dans l'adaptation des pucerons à leur plante hôte et plus particulièrement dans l'adaptation du complexe de biotypes chez le puceron du pois aux différentes espèces de légumineuses.

Les effecteurs très conservés ont-ils des fonctions également conservées au sein des différentes interactions plantes-pucerons ? Ainsi, les effecteurs de différentes espèces de pucerons ont-ils les mêmes cibles dans les différentes plantes hôtes ? Sont-ils plutôt impliqués dans des fonctions communes aux pucerons, comme prévenir l'occlusion du phloème, plutôt que liés à l'adaptation à la plante hôte ?

Les effecteurs évoluant rapidement au sein des Aphididae fonctionnent-ils de manière pucerons-plantes hôtes spécifiques ? Leur évolution rapide a-t-elle permis de cibler des fonctions différentes chez les plantes hôtes ? Les gènes cibles chez les plantes évoluent-ils également rapidement en résultat d'une coévolution entre les espèces de pucerons et leurs plantes hôtes ? Certains de ces effecteurs sont-ils impliqués dans l'adaptation du puceron du pois à une famille de plante particulière ?

Les effecteurs spécifiques d'*A. pisum* sont-ils impliqués dans l'adaptation aux légumineuses ? Présentent-ils des polymorphismes et/ou expressions biotype-spécifiques qui influencent les interactions spécifiques des biotypes avec leurs plantes associées ?

Ainsi, avant d'approfondir ces différentes questions, il est nécessaire de caractériser le rôle effecteur des candidats précédemment identifiés, dans les interactions plantes-pucerons.

### **3.2 Outils et stratégies pour caractériser la fonction des différents effecteurs dans les interactions plantes-pucerons**

Comme présenté dans le chapitre III, les outils pour caractériser les effecteurs de pathogènes de plantes ou d'herbivores sont majoritairement disponibles chez les plantes modèles, telles que *Arabidopsis* et *Nicotiana*. Dans le cadre du complexe de biotypes du puceron du pois, les plantes hôtes légumineuses ne sont pas des plantes modèles et donc peu d'outils ont à ce jour été développés pour étudier leurs parasites. Au cours de cette thèse, nous avons cherché à mettre au point/transférer et optimiser plusieurs outils permettant l'analyse fonctionnelle des effecteurs du puceron du pois.

L'injection de dsRNA dans les pucerons s'étant avérée jusqu'à présent infructueuse entre nos mains, nous avons développé en parallèle l'expression transitoire médiée par *Agrobacterium tumefaciens* chez les légumineuses, cette technique n'ayant jamais été optimisée chez ces plantes non modèles. Cette technique a pu être établie, mais seulement chez le pois *Pisum sativum* (Guy et al. 2016), ce qui a limité le type d'interactions étudiées (chapitre III).

### **Caractérisation du rôle effecteur des différents candidats**

Les effecteurs conservés peuvent être étudiés par des approches d'ARN interférence (*silencing*) chez différentes espèces de pucerons (Naessens et al. 2015; Wang et al. 2015a). Ainsi, une fois l'impact négatif sur les performances démontré chez différentes espèces de pucerons lorsque les différents homologues sont réprimés, les réactions de la plante en réponse à ces effecteurs peuvent être étudiées dans les systèmes pour lesquels les outils sont bien développés, comme le système *Arabidopsis/Nicotiana-M. persicae* (Naessens et al. 2015). Néanmoins, le *silencing* par injection de dsRNA/siRNA ayant certaines limites, notamment en terme de mortalité induite chez les pucerons et de durée de répression de l'expression, des méthodes de répression des effecteurs salivaires *in planta* peuvent s'avérer nécessaires (chapitre III). Ces méthodes déjà bien développées dans le système *Arabidopsis/Nicotiana-M. persicae* (Pitino et al. 2011; Pitino and Hogenhout 2012) sont en cours d'optimisation sur le système puceron pois-légumineuses, comme le *silencing* médié par *Agrobacterium* et par VIGS (*Virus-Induced-Gene-Silencing*) (chapitre III).

La répression de l'expression des effecteurs aphidiens évoluant rapidement ne sera en revanche pas adaptée pour rechercher une fonction puceron-plante hôte spécifique de ces effecteurs. En effet, pour étudier l'effet des différents homologues dans les interactions hôtes et non-hôtes, leur expression *in planta* est nécessaire, comme cela a été réalisé avec les versions de *M. persicae* et *A. pisum* des gènes *Mp1*, *Mp2* et *C002* dans le système *Arabidopsis/Nicotiana-M. persicae* (Pitino and Hogenhout 2012). Le développement de l'expression transitoire médiée par *Agrobacterium* offre maintenant la possibilité d'explorer le rôle de ces effecteurs dans le système *A. pisum*-pois (Guy et al. 2016). Ainsi, l'expression des versions *A. pisum*-spécifiques des effecteurs aphidiens à évolution rapide *Ap25* et *Ap56* a montré un impact bénéfique de ces deux effecteurs sur la fécondité d'*A. pisum* lorsque celui-ci s'alimente sur le pois (Guy et al. 2016, chapitre III). La caractérisation du rôle de ces effecteurs dans l'interaction *Arabidopsis/Nicotiana-M. persicae* permettra de révéler si ces gènes fonctionnent comme effecteurs également chez cette espèce (développement en cours de lignées d'*Arabidopsis* exprimant *Mp25*, l'homologue d'*Ap25*, voir chapitre III). Enfin, dans l'hypothèse d'un rôle de ces effecteurs dans l'adaptation des différentes espèces de pucerons à leurs hôtes respectifs, leur action plante hôte-

spécifique pourra être examinée en exprimant les gènes de *M. persicae* chez le pois, et les gènes d'*A. pisum* chez *Arabidopsis* ou *Nicotiana* (chapitre III).

L'étude du rôle des effecteurs candidats dans la spécialisation des biotypes va nécessiter la mise en place d'outils fonctionnels permettant d'exprimer les différentes versions des gènes chez les légumineuses hôtes du complexe. Ainsi, comme l'expression transitoire médiée par *Agrobacterium* n'a pu être établie que chez le pois (Guy et al. 2016), l'utilisation d'autres méthodes d'expression *in planta*, grâce à des virus de plantes ou à la machinerie de sécrétion de type III de *Pseudomonas syringae* sont à l'étude dans notre équipe (plus de détails sur ces différentes techniques sont apportés dans le chapitre III). Cependant l'outil développé chez le pois peut déjà permettre de tester si l'expression des effecteurs spécifiques du biotype pois aide les autres biotypes non-adaptés à se développer sur cet hôte, suggérant ainsi le rôle de ces effecteurs dans l'adaptation à *P. sativum*. De telles expériences ont déjà été menées par notre équipe, notamment avec le gène *Ap25*, mais jusqu'à présent aucun impact positif n'a été observé sur les biotypes non adaptés, ceux-ci se développant très mal voire pas du tout sur le pois (Guy et al. 2016). L'agroinfiltration ne fonctionnant pas chez la luzerne *M. sativa* (Guy et al. 2016), le développement de la transformation stable de cette plante (Pereira and Erickson 1995) nous permettra de bénéficier de deux outils afin de comparer les fonctions des effecteurs spécifiques des biotypes pois et luzerne pour lesquels les études transcriptomiques révèlent des différences. Ainsi, des effecteurs pourront être associés à une interaction spécifique au sein du complexe de biotype suggérant leur rôle dans la spécialisation des biotypes.

La caractérisation de l'impact des effecteurs candidats sur la *fitness* des pucerons connaît néanmoins certaines limites. En effet, compte tenu du grand nombre d'effecteurs candidats révélé chez *A. pisum*, il apparaît évident que la manipulation de la plante hôte est assurée par plusieurs effecteurs. Les effecteurs peuvent agir en synergie, ce qui rend plus difficile la détection des phénotypes induits lors de leur étude un par un, leurs effets individuels pouvant être très faibles. Dans le cas des familles multigéniques, des effets phénotypiques plus forts pourraient être induits par le *silencing* de plusieurs membres à la fois, mais pour définir quelles copies sont les plus déterminantes, il faudra ensuite les exprimer *in planta* une par une et tester des co-expressions.

### ***Dialogue moléculaire et interaction des effecteurs avec les défenses des plantes***

Après avoir démontré le rôle effecteur des candidats chez les pucerons, l'exploration du rôle de ces effecteurs dans l'adaptation à la plante hôte passe par la compréhension du dialogue moléculaire plante-puceron, et donc par l'identification des cibles des effecteurs dans la plante. Différentes

méthodes permettant de réaliser cette étape sont présentées dans le chapitre III. Jusqu'à maintenant, seule une cible de plante a été identifiée dans les interactions plantes-pucerons (celle de l'effecteur Mp1 de *M. persicae* chez *Arabidopsis* et la pomme de terre), mais des études sont en cours dans différents groupes de recherche afin d'identifier notamment les cibles des effecteurs Me10 et Mif1 (Atamian et al. 2012; Naessens et al. 2015). En revanche, chez *A. pisum* aucune étude n'a encore été menée dans ce sens. Une fois identifiées, ces cibles des effecteurs peuvent permettre d'apporter des réponses quant à la fonction des différents effecteurs lors de l'interaction avec la plante. D'autres analyses sur les plantes hôtes (e.g. études de mutants KO pour la cible) seront alors requises pour comprendre quelles voies de défense sont, d'un côté mises en jeu par la plante et de l'autre ciblées par les effecteurs des pucerons. La compréhension fine du dialogue moléculaire entre le puceron et sa plante est également nécessaire pour caractériser le rôle des effecteurs dans l'adaptation à la plante hôte chez les pucerons.

Les cibles des effecteurs conservés parmi les pucerons, si elles sont conservées chez plusieurs plantes hôtes, permettront de valider l'hypothèse d'un rôle de ces effecteurs dans des fonctions communes aux pucerons, et d'examiner si elles sont liées à l'établissement du mode d'alimentation des pucerons de manière générale.

Les gènes aphidiens évoluant rapidement, s'ils agissent de manière plante hôte-spécifique et qu'une spécificité effecteurs-cibles est caractérisée, peuvent être impliqués dans l'adaptation de différentes espèces de pucerons à leur plantes hôtes, par exemple à une famille spécifique dans le cas d'*A. pisum*. L'analyse phylogénétique et évolutive des gènes cibles parmi les espèces de plantes sera nécessaire pour tester l'existence d'une éventuelle coévolution des effecteurs et des défenses des plantes induite lors d'interactions antagonistes. Dans le cas d'*A. pisum*, pour identifier un rôle de ces effecteurs dans l'adaptation à la famille des légumineuses, l'étude de ces effecteurs chez des espèces proches du genre *Acyrtosiphon* pourra s'avérer nécessaire.

Finalement dans le cas des biotypes, la caractérisation d'interactions effecteur-cible, qui seraient spécifiques d'associations biotype-hôte et notamment liées à des polymorphismes biotype-spécifique suggérerait l'implication de ces effecteurs dans la diversification du complexe de biotypes.

## 4. Implications évolutives et appliquées

La classification des effecteurs en trois catégories, comme présentée précédemment, est une base de travail simplifiée et hypothétique. Ainsi, sans exploration fonctionnelle dans différents systèmes plantes-pucerons, le rôle des effecteurs dans l'adaptation reste difficile à prédire. De plus, les études de génomique comparative entre espèces de puceron et de génétique des populations et/ou transcriptionnelle aux sein du complexe de biotypes du puceron du pois montrent que les gènes candidats à l'adaptation aux légumineuses et à la diversification des biotypes peuvent être les mêmes (Jaquiéry et al. 2012, chapitre II). Comprendre les mécanismes évolutifs dans lesquels sont impliqués les effecteurs des pucerons mais plus particulièrement ceux d'*A. pisum* peut apporter des connaissances fondamentales sur l'évolution et la diversification des insectes herbivores. En effet, l'adaptation d'*A. pisum* à la famille des légumineuses se situe à une échelle de temps large (plusieurs millions d'années, Jaquiery et al. 2017), au cours de laquelle des phénomènes de coévolution ont pu avoir lieu avec les défenses mises en place par cette famille de plante, ayant ainsi pu façonner le répertoire des effecteurs salivaires d'*A. pisum*. En revanche, à l'échelle du complexe de biotypes, la diversification est bien plus récente (quelques milliers d'années, Peccoud et al. 2009b). L'état actuel des biotypes situés le long d'un continuum de divergence, peut illustrer les prémisses d'événements de spéciation conduisant à une nouvelle radiation adaptative. Ainsi, l'étude du rôle des effecteurs salivaires dans la spécialisation des biotypes contribue à la compréhension générale de la grande diversification des insectes herbivores et des fonctions qui y sont associées (e.g. reconnaissance de l'hôte, suppression des défenses de l'hôte et digestion/détoxification de la ressource).

Dans un but agronomique, la compréhension fine des interactions plantes-pucerons et des mécanismes qui permettent aux pucerons de contourner les défenses (et quelles défenses) des plantes, est fondamentale pour développer des méthodes durables et plus respectueuses de l'environnement pour lutter contre les pucerons. Ces derniers étant de plus en plus résistants aux insecticides, favoriser la résistance des plantes semble alors une bonne alternative. Par croisements et sélection variétale, les QTL de résistance aux pucerons pourraient alors être transférés de génotypes résistants vers des génotypes sensibles à haut potentiel agronomique. D'autre part, l'identification des effecteurs des pucerons et de leurs cibles chez les plantes représente une source importante de connaissances pour la création de plantes transgéniques résistantes aux pucerons.



## 5. Conclusion

Ces travaux de thèse ont permis d'élargir le répertoire des effecteurs candidats chez le puceron du pois, de caractériser leurs patrons d'évolution et de révéler des signatures évolutives particulières suggérant leur rôle dans les interactions moléculaires plantes-pucerons. Les approches de génomique comparative et d'estimation du taux d'évolution au sein des espèces de pucerons ont permis d'identifier des effecteurs candidats dans l'adaptation du puceron du pois à une famille particulière de plantes, les légumineuses, et ouvrent la voie à de futures analyses visant à caractériser les bases de l'adaptation des biotypes à leurs hôtes spécifiques au travers notamment le rôle joué par les effecteurs salivaires. Les avancées fournies quant à la description de l'évolution des interactions légumineuses-*A. pisum* participent ainsi à la compréhension des mécanismes sous-jacents à l'adaptation et la spéciation au cours d'interactions antagonistes. Finalement, le développement d'un outil d'expression des effecteurs salivaires chez le pois et les perspectives de nouveaux outils fonctionnels à venir, offrent la possibilité d'explorer les fonctions des effecteurs candidats dans l'interaction avec la plante et de décrypter le dialogue moléculaire plante-puceron. Ainsi, l'acquisition de connaissances fines sur les interactions plantes-pucerons bénéficiera à des objectifs plus finalisés visant la création/sélection de plantes résistantes aux pucerons, qui sont encore aujourd'hui des ravageurs majeurs des cultures.

## V. Références bibliographiques

- Abdellatef E, Will T, Koch A, Imani J, Vilcinskas A, Kogel K-H. 2015. Silencing the expression of the salivary sheath protein causes transgenerational feeding suppression in the aphid *Sitobion avenae*. *Plant Biotechnol. J.* 13:849–857.
- Acevedo FE, Rivera-Vega LJ, Chung SH, Ray S, Felton GW. 2015. Cues from chewing insects — the intersection of DAMPs, HAMPs, MAMPs and effectors. *Curr. Opin. Plant Biol.* 26:80–86.
- Acuña R, Padilla BE, Flórez-Ramos CP, Rubio JD, Herrera JC, Benavides P, Lee S-J, Yeats TH, Egan AN, Doyle JJ, et al. 2012. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc. Natl. Acad. Sci. U. S. A.* 109:4197–4202.
- Agard NJ, Wells JA. 2009. Methods for the proteomic identification of protease substrates. *Curr. Opin. Chem. Biol.* 13:503–509.
- Aggarwal R, Subramanyam S, Zhao C, Chen M-S, Harris MO, Stuart JJ. 2014. Avirulence effector discovery in a plant galling and plant parasitic arthropod, the Hessian Fly (*Mayetiola destructor*). *PLOS ONE* 9:e100958.
- Asplen MK, Bano N, Brady CM, Desneux N, Hopper KR, Malouines C, Oliver KM, White JA, Heimpel GE. 2014. Specialisation of bacterial endosymbionts that protect aphids from parasitoids. *Ecol. Entomol.* 39:736–739.
- Atamian HS, Chaudhary R, Cin VD, Bao E, Girke T, Kaloshian I. 2012. In planta expression or delivery of potato aphid *Macrosiphum euphorbiae* effectors Me10 and Me23 enhances aphid fecundity. *Mol. Plant. Microbe Interact.* 26:67–74.
- Bashandy H, Jalkanen S, Teeri TH. 2015. Within leaf variation is the largest source of variation in agroinfiltration of *Nicotiana benthamiana*. *Plant Methods* 11.
- Bass C, Puinean AM, Zimmer CT, Denholm I, Field LM, Foster SP, Gutbrod O, Nauen R, Slater R, Williamson MS. 2014. The evolution of insecticide resistance in the peach potato aphid, *Myzus persicae*. *Insect Biochem. Mol. Biol.* 51:41–51.
- Bass C, Zimmer CT, Riveron JM, Wilding CS, Wondji CS, Kausmann M, Field LM, Williamson MS, Nauen R. 2013. Gene amplification and microsatellite polymorphism underlie a recent insect host shift. *Proc. Natl. Acad. Sci.* 110:19460–19465.
- Bernays E, Graham M. 1988. On the Evolution of host specificity in phytophagous arthropods. *Ecology* 69:886–892.
- Betz R, Walter S, Requena N. 2016. Alternative splicing – an elegant way to diversify the function of repeat-containing effector proteins? *New Phytol.* 212:306–309.
- Biere A, Bennett AE. 2013. Three-way interactions between plants, microbes and insects. *Funct. Ecol.* 27:567–573.

- Bindschedler LV, Panstruga R, Spanu PD. 2016. Mildew-Omics: How global analyses aid the understanding of life and evolution of powdery mildews. *Front. Plant Sci.* 7.
- Blackman RL, Eastop VF. 1994. *Aphids on the world's trees: an identification and information guide.* Cab International
- Blackman RL, Eastop VF. 2000. *Aphids on the world's crops: an identification and information guide.* John Wiley & Sons Ltd
- Boissinot S, Pichon E, Sorin C, Piccini C, Scheidecker D, Ziegler-Graff V, Brault V. 2017. Systemic propagation of a fluorescent infectious clone of a polerovirus following inoculation by *Agrobacteria* and aphids. *Viruses* 9:166.
- Bonaventure G, VanDoorn A, Baldwin IT. 2011. Herbivore-associated elicitors: FAC signaling and metabolism. *Trends Plant Sci.* 16:294–299.
- Bos JIB, Prince D, Pitino M, Maffei ME, Win J, Hogenhout SA. 2010. A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (Green Peach Aphid). *PLOS Genet.* 6:e1001216.
- Braby MF, Trueman JWH, Eastwood R. 2005. When and where did troidine butterflies (Lepidoptera : Papilionidae) evolve? Phylogenetic and biogeographic evidence suggests an origin in remnant Gondwana in the Late Cretaceous. *Invertebr. Syst.* 19:113–143.
- Braendle C, Davis GK, Brisson JA, Stern DL. 2006. Wing dimorphism in aphids. *Heredity* 97:192–199.
- Brefort T, Tanaka S, Neidig N, Doehlemann G, Vincon V, Kahmann R. 2014. Characterization of the largest effector gene cluster of *Ustilago maydis*. *PLoS Pathog.* 10.
- ten Broeke CJM, Dicke M, van Loon JJA. 2013. Feeding behaviour and performance of different populations of the black currant-lettuce aphid, *Nasonovia ribisnigri*, on resistant and susceptible lettuce. *Entomol. Exp. Appl.* 148:130–141.
- Buchner P. 1965. *Endosymbiosis of animals with plant microorganisms.* Krieger Pub Co
- Burd JD, Porter DR, Puterka GJ, Haley SD, Peairs FB. 2006. Biotypic variation among North American russian wheat aphid (Homoptera: Aphididae) populations. *J. Econ. Entomol.* 99:1862–1866.
- Bush GL. 1966. The taxonomy, cytology, and evolution of the genus *Rhagoletis* in North America (Diptera, Tephritidae).
- Büttner D. 2016. Behind the lines—actions of bacterial type III effector proteins in plant cells. *FEMS Microbiol. Rev.* 40:894–937.
- Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, et al. 2006. Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proc. Natl. Acad. Sci.* 103:14959–14964.
- Carolan JC, Caragea D, Reardon KT, Mutti NS, Dittmer N, Pappan K, Cui F, Castaneto M, Poulain J, Dossat C, et al. 2011. Predicted effector molecules in the salivary secretome of the pea aphid

- (*Acyrtosiphon pisum*): a dual transcriptomic/proteomic approach. *J. Proteome Res.* 10:1505–1518.
- Carolan JC, Fitzroy CIJ, Ashton PD, Douglas AE, Wilkinson TL. 2009. The secreted salivary proteome of the pea aphid *Acyrtosiphon pisum* characterised by mass spectrometry. *PROTEOMICS* 9:2457–2467.
- Casteel CL, Walling LL, Paine TD. 2006. Behavior and biology of the tomato psyllid, *Bactericerca cockerelli*, in response to the Mi-1.2 gene. *Entomol. Exp. Appl.* 121:67–72.
- Celorio-Mancera M de la P, Heckel DG, Vogel H. 2012. Transcriptional analysis of physiological pathways in a generalist herbivore: responses to different host plants and plant structures by the cotton bollworm, *Helicoverpa armigera*. *Entomol. Exp. Appl.* 144:123–133.
- Chaudhary R, Atamian HS, Shen Z, Briggs SP, Kaloshian I. 2014. GroEL from the endosymbiont *Buchnera aphidicola* betrays the aphid by triggering plant defense. *Proc. Natl. Acad. Sci.* 111:8919–8924.
- Chaudhary R, Atamian HS, Shen Z, Briggs SP, Kaloshian I. 2015. Potato aphid salivary proteome: enhanced salivation using resorcinol and identification of aphid phosphoproteins. *J. Proteome Res.* 14:1762–1778.
- Chung Park K, Hardie J. 2003. Electroantennogram responses of aphid nymphs to plant volatiles. *Physiol. Entomol.* 28:215–220.
- Chung SH, Rosa C, Scully ED, Peiffer M, Tooker JF, Hoover K, Luthe DS, Felton GW. 2013. Herbivore exploits orally secreted bacteria to suppress plant defenses. *Proc. Natl. Acad. Sci. U. S. A.* 110:15728–15733.
- Coleman AD, Wouters RHM, Mugford ST, Hogenhout SA. 2015. Persistence and transgenerational effect of plant-mediated RNAi in aphids. *J. Exp. Bot.* 66:541–548.
- Constantin GD, Krath BN, MacFarlane SA, Nicolaisen M, Elisabeth Johansen I, Lund OS. 2004. Virus-induced gene silencing as a tool for functional genomics in a legume species. *Plant J.* 40:622–631.
- Cosson V, Durand P, d’Erfurth I, Kondorosi A, Ratet P. 2006. *Medicago truncatula* transformation using leaf explants. *Methods Mol. Biol.* Clifton NJ 343:115–127.
- Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14:1188–1190.
- Das A, Lee S-H, Hyun TK, Kim S-W, Kim J-Y. 2013. Plant volatiles as method of communication. *Plant Biotechnol. Rep.* 7:9–26.
- D J Futuyma, Moreno and G. 1988. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* 19:207–233.
- De Vos M, Jander G. 2009. *Myzus persicae* (green peach aphid) salivary components induce defence responses in *Arabidopsis thaliana*. *Plant Cell Environ.* 32:1548–1560.

- De Vos M, Van Oosten VR, Van Poecke RMP, Van Pelt JA, Pozo MJ, Mueller MJ, Buchala AJ, Métraux J-P, Van Loon LC, Dicke M, et al. 2005. Signal signature and transcriptome changes of *Arabidopsis* during pathogen and insect attack. *Mol. Plant. Microbe Interact.* 18:923–937.
- Diezel C, von Dahl CC, Gaquerel E, Baldwin IT. 2009. Different Lepidopteran elicitors account for cross-talk in herbivory-induced phytohormone signaling. *Plant Physiol.* 150:1576–1586.
- Dogimont C, Bendahmane A, Chovelon V, Boissot N. 2010. Host plant resistance to aphids in cultivated crops: genetic and molecular bases, and interactions with aphid populations. *C. R. Biol.* 333:566–573.
- Dolling WR. 1991. Hemiptera. Oxford University Press
- Douglas A. 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annu. Rev. Entomol.* 43:17–37.
- Douglas AE. 2006. Phloem-sap feeding by animals: problems and solutions. *J. Exp. Bot.* 57:747–754.
- Du B, Zhang W, Liu B, Hu J, Wei Z, Shi Z, He R, Zhu L, Chen R, Han B, et al. 2009. Identification and characterization of Bph14, a gene conferring resistance to brown planthopper in rice. *Proc. Natl. Acad. Sci.* 106:22163–22168.
- 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* 15:762.
- Duvaux L, Geissmann Q, Gharbi K, Zhou J-J, Ferrari J, Smadja CM, Butlin RK. 2015. Dynamics of copy number variation in host races of the pea aphid. *Mol. Biol. Evol.* 32:63–80.
- Eastop VF. 1971. Keys for the identification of *Acyrtosiphon* (Hemiptera : Aphididae). *Bull. Br. Mus. Nat. Hist. B Entomol.* 26:1–115.
- Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci.* 112:8362–8366.
- Ehrlich PR, Raven PH. 1964. Butterflies and plants: a study in coevolution. *Evolution* 18:586–608.
- Eichenseer H, Mathews MC, Powell JS, Felton GW. 2010. Survey of a Salivary Effector in Caterpillars: Glucose oxidase variation and correlation with host range. *J. Chem. Ecol.* 36:885–897.
- Ellis C, Karafyllidis I, Turner JG. 2002. Constitutive activation of jasmonate signaling in an *Arabidopsis* mutant correlates with enhanced resistance to *Erysiphe cichoracearum*, *Pseudomonas syringae*, and *Myzus persicae*. *Mol. Plant-Microbe Interact.* MPMI 15:1025–1030.
- Elzinga DA, De Vos M, Jander G. 2014. Suppression of plant defenses by a *Myzus persicae* (Green Peach Aphid) salivary effector protein. *Mol. Plant. Microbe Interact.* 27:747–756.
- Engsontia P, Sangket U, Chotigeat W, Satasook C. 2014. Molecular evolution of the odorant and gustatory receptor genes in lepidopteran insects: implications for their adaptation and speciation. *J. Mol. Evol.* 79:21–39.

- Erb M, Meldau S, Howe GA. 2012. Role of phytohormones in insect-specific plant reactions. *Trends Plant Sci.* 17:250–259.
- Eyres I, Duvaux L, Gharbi K, Tucker R, Hopkins D, Simon J-C, Ferrari J, Smadja CM, Butlin RK. 2017. Targeted re-sequencing confirms the importance of chemosensory genes in aphid host race differentiation. *Mol. Ecol.* 26:43–58.
- Eyres I, Jaquiéry J, Sugio A, Duvaux L, Gharbi K, Zhou J-J, Legeai F, Nelson M, Simon J-C, Smadja CM, et al. 2016. Differential gene expression according to race and host plant in the pea aphid. *Mol. Ecol.* 25:4197–4215.
- Eyun S, Soh HY, Posavi M, Munro JB, Hughes DST, Murali SC, Qu J, Dugan S, Lee SL, Chao H, et al. 2017. Evolutionary history of chemosensory-related gene families across the Arthropoda. *Mol. Biol. Evol.* 34:1838–1862.
- Fabro G, Steinbrenner J, Coates M, Ishaque N, Baxter L, Studholme DJ, Körner E, Allen RL, Piquerez SJM, Rougon-Cardoso A, et al. 2011. Multiple candidate effectors from the oomycete pathogen *Hyaloperonospora arabidopsidis* suppress host plant immunity. *PLOS Pathog.* 7:e1002348.
- Feder JL, Berlocher SH, Roethele JB, Dambroski H, Smith JJ, Perry WL, Gavrilovic V, Filchak KE, Rull J, Aluja M. 2003. Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc. Natl. Acad. Sci.* 100:10314–10319.
- Ferrari J, Darby AC, Daniell TJ, Godfray HCJ, Douglas AE. 2004. Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance. *Ecol. Entomol.* 29:60–65.
- Ferrari J, Godfray HCJ, Faulconbridge AS, Prior K, Via S, Funk D. 2006. Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. *Evolution* 60:1574–1584.
- Ferrari J, Scarborough CL, Godfray HCJ. 2007. Genetic variation in the effect of a facultative symbiont on host-plant use by pea aphids. *Oecologia* 153:323–329.
- Fischer HM, Wheat CW, Heckel DG, Vogel H. 2008. Evolutionary origins of a novel host plant detoxification gene in butterflies. *Mol. Biol. Evol.* 25:809–820.
- Forbes AA, Devine SN, Hippee AC, Tvedte ES, Ward AK, Widmayer HA, Wilson CJ. 2017. Revisiting the particular role of host shifts in initiating insect speciation. *Evolution* 71:1126–1137.
- Forister ML, Dyer LA, Singer MS, Stireman JO, Lill JT. 2012. Revisiting the evolution of ecological specialization, with emphasis on insect-plant interactions. *Ecology* 93:981–991.
- Foster SP, Paul VL, Slater R, Warren A, Denholm I, Field LM, Williamson MS. 2014. A mutation (L1014F) in the voltage-gated sodium channel of the grain aphid, *Sitobion avenae*, is associated with resistance to pyrethroid insecticides. *Pest Manag. Sci.* 70:1249–1253.
- Fraenkel GS. 1959. The Raison d'Être of secondary plant substances. *Science* 129:1466–1470.
- Frago E, Dicke M, Godfray HCJ. 2012. Insect symbionts as hidden players in insect-plant interactions. *Trends Ecol. Evol.* 27:705–711.

- Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E, Weber AP. 2011. Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC Genomics* 12:227.
- Frantz A, Plantegenest M, Mieuze L, Simon J-C. 2006. Ecological specialization correlates with genotypic differentiation in sympatric host-populations of the pea aphid. *J. Evol. Biol.* 19:392–401.
- Freeman. 2008. An overview of plant defenses against pathogens and herbivores. *Plant Health Instr.*
- Furch ACU, Bel V, E AJ, Will T. 2015. Aphid salivary proteases are capable of degrading sieve-tube proteins. *J. Exp. Bot.* 66:533–539.
- Furch ACU, Zimmermann MR, Will T, Hafke JB, van Bel AJE. 2010. Remote-controlled stop of phloem mass flow by biphasic occlusion in *Cucurbita maxima*. *J. Exp. Bot.* 61:3697–3708.
- Fürstenberg-Hägg J, Zagrobelny M, Bak S. 2013. Plant defense against insect herbivores. *Int. J. Mol. Sci.* 14:10242–10297.
- Futuyma DJ, Agrawal AA. 2009. Macroevolution and the biological diversity of plants and herbivores. *Proc. Natl. Acad. Sci.* 106:18054–18061.
- Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. 2008. *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Mol. Ecol.* 17:1648–1657.
- Gatehouse JA. 2002. Plant resistance towards insect herbivores: a dynamic interaction. *New Phytol.* 156:145–169.
- Gautier M. 2015. Genome-Wide scan for adaptive divergence and association with population-specific covariates. *Genetics* 201:1555–1579.
- Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P, Kerdelhué C, Estoup A. 2013. Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Mol. Ecol.* 22:3766–3779.
- Gil R, Sabater-Muñoz B, Latorre A, Silva FJ, Moya A. 2002. Extreme genome reduction in *Buchnera spp.*: Toward the minimal genome needed for symbiotic life. *Proc. Natl. Acad. Sci.* 99:4454–4458.
- Giraud M, Hilliou F, Fricaux T, Audant P, Feyereisen R, Le Goff G. 2015. Cytochrome P450s from the fall armyworm (*Spodoptera frugiperda*): responses to plant allelochemicals and pesticides. *Insect Mol. Biol.* 24:115–128.
- Giron D, Dedeine F, Dubreuil G, Huguet E, Mouton L, Outreman Y, Vavre F, Simon J-C. 2017. Chapter Seven-Influence of microbial symbionts on plant–insect interactions. *Adv. Bot. Res.* 81:225–257.
- Girousse C, Moulia B, Silk W, Bonnemain J-L. 2005. Aphid infestation causes different changes in carbon and nitrogen allocation in Alfalfa stems as well as different inhibitions of longitudinal and radial expansion. *Plant Physiol.* 137:1474–1484.

- Gouin A, Legeai F, Nouhaud P, Whibley A, Simon J-C, Lemaitre C. 2015. Whole-genome re-sequencing of non-model organisms: lessons from unmapped reads. *Heredity* 114:494–501.
- Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, Osborne EJ, Dermauw W, Thi Ngoc PC, Ortego F, et al. 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479:487–492.
- Guo K, Wang W, Luo L, Chen J, Guo Y, Cui F. 2014. Characterization of an aphid-specific, cysteine-rich protein enriched in salivary glands. *Biophys. Chem.* 189:25–32.
- Guy E, Boulain H, Aigu Y, Le Penne C, Chawki K, Morlière S, Schädel K, Kunert G, Simon J-C, Sugio A. 2016. Optimization of agroinfiltration in *Pisum sativum* provides a new tool for studying the salivary protein functions in the pea aphid complex. *Front. Plant Sci.* 7.
- Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461:393–398.
- Hales DF, Wilson AC, Sloane MA, Simon J-C, Legallic J-F, Sunnucks P. 2002. Lack of detectable genetic recombination on the X chromosome during the parthenogenetic production of female and male aphids. *Genet. Res.* 79:203–209.
- Haley SD, Pears FB, Walker CB, Rudolph JB, Randolph TL. 2004. Occurrence of a new russian wheat aphid biotype in Colorado. *Crop Sci.* 44:1589–1592.
- Ham RCHJ van, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernández JM, Jiménez L, Postigo M, Silva FJ, et al. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl. Acad. Sci.* 100:581–586.
- Hansen AK, Moran NA. 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol. Ecol.* 23:1473–1496.
- Hardy NB, Peterson DA, von Dohlen CD. 2015. The evolution of life cycle complexity in aphids: Ecological optimization or historical constraint? *Evolution* 69:1423–1432.
- Harmel N, Létocart E, Cherqui A, Giordanengo P, Mazzucchelli G, Guillonneau F, De Pauw E, Haubruge E, Francis F. 2008. Identification of aphid salivary proteins: a proteomic investigation of *Myzus persicae*. *Insect Mol. Biol.* 17:165–174.
- Harris KF, Maramorosch K. 2014. *Aphids as virus vectors*. Elsevier
- Harris MO, Freeman TP, Rohfritsch O, Anderson KG, Payne SA, Moore JA. 2006. Virulent Hessian Fly (Diptera: Cecidomyiidae) larvae induce a nutritive tissue during compatible interactions with wheat. *Ann. Entomol. Soc. Am.* 99:305–316.
- Hawthorne DJ, Via S. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* 412:904–907.
- Hewer A, Becker A, Bel AJE van. 2011. An aphid's Odyssey – the cortical quest for the vascular bundle. *J. Exp. Biol.* 214:3868–3879.



- Hilker M, Fatouros NE. 2015. Plant responses to insect egg deposition. *Annu. Rev. Entomol.* 60:493–515.
- Hogenhout SA, Bos JI. 2011. Effector proteins that modulate plant–insect interactions. *Curr. Opin. Plant Biol.* 14:422–428.
- Hogenhout SA, Van der Hoorn RA, Terauchi R, Kamoun S. 2009. Emerging concepts in effector biology of plant-associated organisms. *Mol. Plant. Microbe Interact.* 22:115–122.
- Howe GA, Jander G. 2008. Plant immunity to insect herbivores. *Annu. Rev. Plant Biol.* 59:41–66.
- Hu J, Xiao C, He Y. 2016. Recent progress on the genetics and molecular breeding of brown planthopper resistance in rice. *Rice* 9:30.
- Ido Y, Nakahara KS, Uyeda I. 2012. White clover mosaic virus-induced gene silencing in pea. *J. Gen. Plant Pathol.* 78:127–132.
- Ighil ET-A, Dedryver C-A, Hullé M, Chaubet B. 2011. Les pucerons des grandes cultures: cycles biologiques et activités de vol. Editions Quae
- Janssen B-J, Gardner RC. 1990. Localized transient expression of GUS in leaf discs following cocultivation with *Agrobacterium*. *Plant Mol. Biol.* 14:61–72.
- Janz N. 2011. Ehrlich and Raven Revisited: Mechanisms underlying codiversification of plants and enemies. *Annu. Rev. Ecol. Evol. Syst.* 42:71–89.
- Janzen DH. 1980. When is it coevolution. *Evolution* 34:611–612.
- Jaouannet M, Rodriguez PA, Thorpe P, Lenoir CJ, MacLeod R, Escudero-Martinez C, Bos JI. 2014. Plant immunity in plant–aphid interactions. *Front. Plant Sci.* 5.
- Jaquiery J, Peccoud J, Ouisse T, Legeai F, Prunier-Leterme N, Gouin A, Nouhaud P, Brisson J, Bickel R, Purandare S, et al. 2017. Disentangling The Causes For Faster-X Evolution In Aphids. *bioRxiv*.
- Jaquiéry J, Rispe C, Roze D, Legeai F, Trionnaire GL, Stoeckel S, Mieuzet L, Silva CD, Poulain J, Prunier-Leterme N, et al. 2013. Masculinization of the X chromosome in the pea aphid. *PLOS Genet.* 9:e1003690.
- Jaquiéry J, Stoeckel S, Larose C, Nouhaud P, Rispe C, Mieuzet L, Bonhomme J, Mahéo F, Legeai F, Gauthier J-P, et al. 2014. Genetic control of contagious asexuality in the pea aphid. *PLOS Genet.* 10:e1004838.
- Jaquiéry J, Stoeckel S, Nouhaud P, Mieuzet L, Mahéo F, Legeai F, Bernard N, Bonvoisin A, Vitalis R, Simon J-C. 2012. Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex. *Mol. Ecol.* 21:5251–5264.
- Jaubert-Possamai S, Le Trionnaire G, Bonhomme J, Christophides GK, Rispe C, Tagu D. 2007. Gene knockdown by RNAi in the pea aphid *Acyrtosiphon pisum*. *BMC Biotechnol.* 7:63.

- Jaubert-Possamai S, Rispe C, Tanguy S, Gordon K, Walsh T, Edwards O, Tagu D. 2010. Expansion of the miRNA pathway in the Hemipteran insect *Acyrtosiphon pisum*. *Mol. Biol. Evol.* 27:979–987.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in adaptive bacterial immunity. *Science* 337:816–821.
- Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444:323–329.
- Kaloshian I, Walling LL. 2016. Hemipteran and dipteran pests: Effectors and plant host immune regulators. *J. Integr. Plant Biol.* 58:350–361.
- Kanvil S, Collins CM, Powell G, Turnbull CGN. 2015. Cryptic virulence and avirulence alleles revealed by controlled sexual recombination in pea aphids. *Genetics* 199:581–593.
- Kanvil S, Powell G, Turnbull C. 2014. Pea aphid biotype performance on diverse *Medicago* host genotypes indicates highly specific virulence and resistance functions. *Bull. Entomol. Res.* 104:689–701.
- Karimi M, Inzé D, Depicker A. 2002. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7:193–195.
- Kazan K, Lyons R. 2014. Intervention of phytohormone pathways by pathogen effectors. *Plant Cell Online* 26:2285–2309.
- Kerchev PI, Karpińska B, Morris JA, Hussain A, Verrall SR, Hedley PE, Fenton B, Foyer CH, Hancock RD. 2013. Vitamin C and the Abscisic Acid-Insensitive 4 Transcription factor are important determinants of aphid resistance in *Arabidopsis*. *Antioxid. Redox Signal.* 18:2091–2105.
- Kerppola TK. 2006. Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. *Nat. Protoc.* 1:1278–1286.
- Kerr A. 1987. The impact of molecular genetics on plant pathology. *Annu. Rev. Phytopathol.* 25:87–110.
- Kettles GJ, Drurey C, Schoonbeek H, Maule AJ, Hogenhout SA. 2013. Resistance of *Arabidopsis thaliana* to the green peach aphid, *Myzus persicae*, involves camalexin and is regulated by microRNAs. *New Phytol.* 198:1178–1190.
- Kettles GJ, Kaloshian I. 2016. The potato aphid salivary effector Me47 is a Glutathione-S-Transferase involved in modifying plant responses to aphid infestation. *Front. Plant Sci.* 7.
- Kim H, Lee S, Jang Y. 2011. Macroevolutionary patterns in the Aphidini aphids (Hemiptera: Aphididae): Diversification, host association, and biogeographic origins. *PLOS ONE* 6:e24749.
- Kim JH, Jander G. 2007. *Myzus persicae* (green peach aphid) feeding on *Arabidopsis* induces the formation of a deterrent indole glucosinolate. *Plant J. Cell Mol. Biol.* 49:1008–1019.

- Kim JH, Lee BW, Schroeder FC, Jander G. 2008. Identification of indole glucosinolate breakdown products with antifeedant effects on *Myzus persicae* (green peach aphid). *Plant J. Cell Mol. Biol.* 54:1015–1026.
- Kim K-S, Hill CB, Hartman GL, Mian MAR, Diers BW. 2008. Discovery of soybean aphid biotypes. *Crop Sci.* 48:923–928.
- Kurth EG, Peremyslov VV, Prokhnevsky AI, Kasschau KD, Miller M, Carrington JC, Dolja VV. 2012. Virus-derived gene expression and RNA Interference vector for grapevine. *J. Virol.* 86:6002–6009.
- Lacomme C, Chapman S. 2005. Use of potato virus X (PVX)-based vectors for gene expression and virus-induced gene silencing (VIGS). In: *Current Protocols in Microbiology*. John Wiley & Sons, Inc.
- Lange M, Yellina AL, Orashakova S, Becker A. 2013. Virus-induced gene silencing (vigs) in plants: an overview of target species and the virus-derived vector systems. in: *virus-induced gene silencing. Methods in Molecular Biology*. Humana Press, Totowa, NJ. p. 1–14.
- Lee C. 2007. Coimmunoprecipitation assay. *Methods Mol. Biol.* Clifton NJ 362:401–406.
- Leonardo TE. 2004. Removal of a specialization-associated symbiont does not affect aphid fitness. *Ecol. Lett.* 7:461–468.
- Li F, Han Z. 2004. Mutations in acetylcholinesterase associated with insecticide resistance in the cotton aphid, *Aphis gossypii* Glover. *Insect Biochem. Mol. Biol.* 34:397–405.
- Loxdale HD, Harvey JA. 2016. The ‘generalism’ debate: misinterpreting the term in the empirical literature focusing on dietary breadth in insects. *Biol. J. Linn. Soc.* 119:265–282.
- Ma W, Dong FFT, Stavrinos J, Guttman DS. 2006. Type III Effector diversification via both pathoadaptation and horizontal transfer in response to a coevolutionary arms race. *PLOS Genet.* 2:e209.
- Mao Y-B, Cai W-J, Wang J-W, Hong G-J, Tao X-Y, Wang L-J, Huang Y-P, Chen X-Y. 2007. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat. Biotechnol.* 25:1307–1313.
- Mathers TC, Chen Y, Kaithakottil G, Legeai F, Mugford ST, Baa-Puyoulet P, Bretaudeau A, Clavijo B, Colella S, Collin O, et al. 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol.* 18.
- Mayhew PJ. 2007. Why are there so many insect species? Perspectives from fossils and phylogenies. *Biol. Rev.* 82:425–454.
- McLean AHC, van Asch M, Ferrari J, Godfray HCJ. 2011. Effects of bacterial secondary symbionts on host plant use in pea aphids. *Proc. Biol. Sci.* 278:760–766.
- McPherson BA, Smith DC, Berlocher SH. 1988. Genetic differences between host races of *Rhagoletis pomonella*. *Nature* 336:64–66.

- Medina-Ortega KJ, Walker GP. 2013. Does aphid salivation affect phloem sieve element occlusion in vivo? *J. Exp. Bot.* 64:5525–5535.
- Meihls LN, Handrick V, Glauser G, Barbier H, Kaur H, Haribal MM, Lipka AE, Gershenzon J, Buckler ES, Erb M, et al. 2013. Natural variation in maize aphid resistance is associated with 2,4-Dihydroxy-7-Methoxy-1,4-Benzoxazin-3-One glucoside methyltransferase activity. *Plant Cell:tpc.113.112409*.
- Meziadi C, Blanchet S, Geffroy V, Pflieger S. 2017. Virus-Induced Gene Silencing (VIGS) and foreign gene expression in *Pisum sativum* L. using the “One-Step” Bean pod mottle virus (BPMV) viral vector. *Methods Mol. Biol. Clifton NJ* 1654:311–319.
- Meziadi C, Blanchet S, Richard M, Pilet-Nayel M-L, Geffroy V, Pflieger S. 2016. Bean pod mottle virus: a new powerful tool for functional genomics studies in *Pisum sativum*. *Plant Biotechnol. J.* 14:1777–1787.
- Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346:763–767.
- Missbach C, Dweck HK, Vogel H, Vilcinskas A, Stensmyr MC, Hansson BS, Grosse-Wilde E. 2014. Evolution of insect olfactory receptors. *eLife.* 3.
- Mitter C, Farrell B, Wiegmann B. 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *Am. Nat.* 132:107–128.
- Moeller FW. 1973. The host plants of the potato aphid *Macrosiphum euphorbiae* and of closely related species.
- Montllor CB, Maxmen A, Purcell AH. 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrtosiphon pisum* under heat stress. *Ecol. Entomol.* 27:189–195.
- Moran NA, Degnan PH, Santos SR, Dunbar HE, Ochman H. 2005. The players in a mutualistic symbiosis: Insects, bacteria, viruses, and virulence genes. *Proc. Natl. Acad. Sci. U. S. A.* 102:16919–16926.
- Moran NA, Munson MA, Baumann P, Ishikawa H. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond. B Biol. Sci.* 253:167–171.
- Moran PJ, Thompson GA. 2001. Molecular responses to aphid feeding in *Arabidopsis* in relation to plant defense pathways. *Plant Physiol.* 125:1074–1085.
- Mutti NS, Louis J, Pappan LK, Pappan K, Begum K, Chen M-S, Park Y, Dittmer N, Marshall J, Reese JC, et al. 2008. A protein from the salivary glands of the pea aphid, *Acyrtosiphon pisum*, is essential in feeding on a host plant. *Proc. Natl. Acad. Sci.* 105:9965–9969.
- Mutti NS, Park Y, Reese JC, Reeck GR. 2006. RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrtosiphon pisum*. *J. Insect Sci.* 6:1–7.
- Naessens E, Dubreuil G, Giordanengo P, Baron OL, Minet-Kebdani N, Keller H, Coustau C. 2015. A secreted MIF cytokine enables aphid feeding and represses plant immune responses. *Curr. Biol. CB* 25:1898–1903.

- Ng JC, Perry KL. 2004. Transmission of plant viruses by aphid vectors. *Mol. Plant Pathol.* 5:505–511.
- Nicholson SJ, Hartson SD, Puterka GJ. 2012. Proteomic analysis of secreted saliva from Russian wheat aphid (*Diuraphis noxia* Kurd.) biotypes that differ in virulence to wheat. *J. Proteomics* 75:2252–2268.
- Nicholson SJ, Nickerson ML, Dean M, Song Y, Hoyt PR, Rhee H, Kim C, Puterka GJ. 2015. The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC Genomics* 16:429.
- Nombela G, Williamson VM, Muñiz M. 2003. The root-knot nematode resistance gene Mi-1.2 of tomato is responsible for resistance against the whitefly *Bemisia tabaci*. *Mol. Plant. Microbe Interact.* 16:645–649.
- Noon JB, Qi M, Sill DN, Muppirala U, Eves-van den Akker S, Maier TR, Dobbs D, Mitchum MG, Hewezi T, Baum TJ. 2016. A *Plasmodium*-like virulence effector of the soybean cyst nematode suppresses plant innate immunity. *New Phytol.* 212:444–460.
- Nouhaud P, Gautier M, Gouin A, Jaquiéry J, Peccoud J, Legeai F, Mieuze L, Smadja C, Lemaitre C, Vitalis R, et al. 2017. Identifying genomic hotspots of differentiation and candidate genes involved in the adaptive divergence of pea aphid host races. *Mol Ecol.*
- Nouhaud P, Peccoud J, Mahéo F, Mieuze L, Jaquiéry J, Simon J-C. 2014. Genomic regions repeatedly involved in divergence among plant-specialized pea aphid biotypes. *J. Evol. Biol.* 27:2013–2020.
- Oliver KM, Degnan PH, Burke GR, Moran NA. 2010. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu. Rev. Entomol.* 55:247–266.
- Oliver KM, Degnan PH, Hunter MS, Moran NA. 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science* 325:992–994.
- Oliver KM, Moran NA, Hunter MS. 2006. Costs and benefits of a superinfection of facultative symbionts in aphids. *Proc. R. Soc. Lond. B Biol. Sci.* 273:1273–1280.
- Oliver KM, Russell JA, Moran NA, Hunter MS. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc. Natl. Acad. Sci.* 100:1803–1807.
- Pan Y, Zhu J, Luo L, Kang L, Cui F. 2015. High expression of a unique aphid protein in the salivary glands of *Acyrtosiphon pisum*. *Physiol. Mol. Plant Pathol.* 92:175–180.
- Park S-H, Choi H, Kim S, Cho WK, Kim K-H. 2016. Development of virus-induced gene expression and silencing vector derived from grapevine algerian latent virus. *Plant Pathol. J.* 32:371–376.
- Peccoud J, Bonhomme J, Mahéo F, de la Huerta M, Cosson O, Simon J-C. 2014. Inheritance patterns of secondary symbionts during sexual reproduction of pea aphid biotypes. *Insect Sci.* 21:291–300.
- Peccoud J, Mahéo F, de la Huerta M, Laurence C, Simon J-C. 2015. Genetic characterisation of new host-specialised biotypes and novel associations with bacterial symbionts in the pea aphid complex. *Insect Conserv. Divers.* 8:484–492.

- Peccoud J, Ollivier A, Plantegenest M, Simon J-C. 2009a. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc. Natl. Acad. Sci.* 106:7495–7500.
- Peccoud J, Simon J-C. 2010. The pea aphid complex as a model of ecological speciation. *Ecol. Entomol.* 35:119–130.
- Peccoud J, Simon J-C, McLaughlin HJ, Moran NA. 2009b. Post-Pleistocene radiation of the pea aphid complex revealed by rapidly evolving endosymbionts. *Proc. Natl. Acad. Sci.* 106:16315–16320.
- Pegadaraju V, Louis J, Singh V, Reese JC, Bautor J, Feys BJ, Cook G, Parker JE, Shah J. 2007. Phloem-based resistance to green peach aphid is controlled by *Arabidopsis* PHYTOALEXIN DEFICIENT4 without its signaling partner ENHANCED DISEASE SUSCEPTIBILITY1. *Plant J. Cell Mol. Biol.* 52:332–341.
- Pereira LF, Erickson L. 1995. Stable transformation of alfalfa (*Medicago sativa* L.) by particle bombardment. *Plant Cell Rep.* 14:290–293.
- Pickett J, Wadhams L, Woodcock C, Hardie J. 1992. The chemical ecology of aphids. *Annu. Rev. Entomol.* 37:67–90.
- Pieterse CMJ, Does DV der, Zamioudis C, Leon-Reyes A, Wees SCMV. 2012. Hormonal modulation of plant immunity. *Annu. Rev. Cell Dev. Biol.* 28:489–521.
- Pitino M, Coleman AD, Maffei ME, Ridout CJ, Hogenhout SA. 2011. Silencing of aphid genes by dsRNA feeding from plants. *PLOS ONE* 6:e25709.
- Pitino M, Hogenhout SA. 2012. Aphid protein effectors promote aphid colonization in a plant species-specific manner. *Mol. Plant. Microbe Interact.* 26:130–139.
- Powell G, Tosh C, Hardie J. 2004. Parturition by colonizing aphids: no correlation with phloem ingestion. *Simon JC Dedryer CA Risper C Hull M*:485–489.
- Powell G, Tosh CR, Hardie J. 2006. Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. *Annu Rev Entomol* 51:309–330.
- Powell THQ, Hood GR, Murphy MO, Heilveil JS, Berlocher SH, Nosil P, Feder JL. 2013. Genetic Divergence along the speciation continuum: the transition from host race to species in *Rhagoletis* (diptera: Tephritidae). *Evolution* 67:2561–2576.
- Price DRG, Duncan RP, Shigenobu S, Wilson ACC. 2011. Genome expansion and differential expression of amino acid transporters at the aphid/*Buchnera* symbiotic interface. *Mol. Biol. Evol.* 28:3113–3126.
- Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, Williamson MS, Bass C. 2010. Amplification of a cytochrome p450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLOS Genet.* 6:e1000999.
- Qu F. 2001. Plant viruses as gene expression and silencing vectors. In: eLS. John Wiley & Sons, Ltd.

- Ramasamy S, Ometto L, Crava CM, Revadi S, Kaur R, Horner DS, Pisani D, Dekker T, Anfora G, Rota-Stabelli O. 2016. The evolution of olfactory gene families in *Drosophila* and the genomic basis of chemical-ecological adaptation in *Drosophila suzukii*. *Genome Biol. Evol.* 8:2297–2311.
- Ramsey JS, Wilson ACC, de Vos M, Sun Q, Tamborindeguy C, Winfield A, Malloch G, Smith DM, Fenton B, Gray SM, et al. 2007. Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. *BMC Genomics* 8:423.
- Rao SAK, Carolan JC, Wilkinson TL. 2013. Proteomic profiling of cereal aphid saliva reveals both ubiquitous and adaptive secreted proteins. *PLOS ONE* 8:e57413.
- Rentel MC, Leonelli L, Dahlbeck D, Zhao B, Staskawicz BJ. 2008. Recognition of the *Hyaloperonospora parasitica* effector ATR13 triggers resistance against oomycete, bacterial, and viral pathogens. *Proc. Natl. Acad. Sci.* 105:1091–1096.
- Rispe C, Kutsukake M, Doublet V, Hudaverdian S, Legeai F, Simon J-C, Tagu D, Fukatsu T. 2007. Large gene family expansion and variable selective pressures for cathepsin B in aphids. *Mol. Biol. Evol.* 25:5–17.
- Robert-Seilaniantz A, Grant M, Jones JDG. 2011. Hormone crosstalk in plant disease and defense: more than just JASMONATE-SALICYLATE antagonism. *Annu. Rev. Phytopathol.* 49:317–343.
- Rodriguez PA, Bos JIB. 2012. Toward understanding the role of aphid effectors in plant infestation. *Mol. Plant. Microbe Interact.* 26:25–30.
- Rodriguez PA, Escudero-Martinez C, Bos JIB. 2017. An aphid effector targets trafficking protein VPS52 in a host-specific manner to promote virulence. *Plant Physiol.* 173:1892–1903.
- Rodriguez PA, Stam R, Warbroek T, Bos JIB. 2013. Mp10 and Mp42 from the aphid species *Myzus persicae* trigger plant defenses in *Nicotiana benthamiana* through different activities. *Mol. Plant. Microbe Interact.* 27:30–39.
- Rossi M, Goggin FL, Milligan SB, Kaloshian I, Ullman DE, Williamson VM. 1998. The nematode resistance gene Mi of tomato confers resistance against the potato aphid. *Proc. Natl. Acad. Sci.* 95:9750–9754.
- Russell JA, Moran NA. 2006. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proc. R. Soc. Lond. B Biol. Sci.* 273:603–610.
- Sanchez-Arcos C, Reichelt M, Gershenzon J, Kunert G. 2016. Modulation of legume defense signaling pathways by native and non-native pea aphid clones. *Front. Plant Sci.* 7.
- Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nat. Rev. Genet.* 14:807–820.
- Scarborough CL, Ferrari J, Godfray HCJ. 2005. Aphid protected from pathogen by endosymbiont. *Science* 310:1781–1781.
- Schadt EE, Turner S, Kasarskis A. 2010. A window into third-generation sequencing. *Hum. Mol. Genet.* 19:R227–R240.

- Schell J, Montagu MV. 1977. The Ti-Plasmid of *Agrobacterium tumefaciens*, a natural vector for the introduction of nif genes in plants? In: Genetic engineering for nitrogen fixation. Basic Life Sciences. Springer, Boston, MA. p. 159–179.
- Schmelz EA. 2015. Impacts of insect oral secretions on defoliation-induced plant defense. *Curr. Opin. Insect Sci.* 9:7–15.
- Schmid M, Sieber R, Zimmermann Y-S, Vorburger C. 2012. Development, specificity and sublethal effects of symbiont-conferred resistance to parasitoids in aphids. *Funct. Ecol.* 26:207–215.
- Schwarzkopf A, Rosenberger D, Niebergall M, Gershenson J, Kunert G. 2013. To feed or not to feed: plant factors located in the epidermis, mesophyll, and sieve elements influence pea aphid's ability to feed on legume species. *PLOS ONE* 8:e75298.
- Shan L, He P, Li J, Heese A, Peck SC, Nürnberger T, Martin GB, Sheen J. 2008. Bacterial effectors target the common signaling partner BAK1 to disrupt multiple MAMP receptor-signaling complexes and impede plant immunity. *Cell Host Microbe* 4:17–27.
- Shikano I, Rosa C, Tan C-W, Felton GW. 2017. Tritrophic interactions: microbe-mediated plant effects on insect herbivores. *Annu. Rev. Phytopathol.* 55:313–331.
- Simon J-C, d'Alençon E, Guy E, Jacquin-Joly E, Jaquiéry J, Nouhaud P, Peccoud J, Sugio A, Streiff R. 2015. Genomics of adaptation to host-plants in herbivorous insects. *Brief. Funct. Genomics* 14:413–423.
- Simon J-C, Biere A, Sugio A. 2017. The promises and challenges of research on plant–insect–microbe interactions. *Insect Sci.*
- Simon J-C, Carré S, Boutin M, Prunier-Leterme N, Sabater-Muñoz B, Latorre A, Bournoville R. 2003. Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. *Proc. R. Soc. Lond. B Biol. Sci.* 270:1703–1712.
- Simon J-C, Stoeckel S, Tagu D. 2010. Evolutionary and functional insights into reproductive strategies of aphids. *C. R. Biol.* 333:488–496.
- Simpson GG. 1953. *The Major Features of Evolution* (Columbia Univ. Press, New York)
- Smadja C, Shi P, Butlin RK, Robertson HM. 2009. Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Mol. Biol. Evol.* 26:2073–2086.
- Smadja CM, Canbäck B, Vitalis R, Gautier M, Ferrari J, Zhou J-J, Butlin RK. 2012. Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution* 66:2723–2738.
- Sohn KH, Lei R, Nemri A, Jones JDG. 2007. The downy mildew effector proteins ATR1 and ATR13 promote disease susceptibility in *Arabidopsis thaliana*. *Plant Cell Online* 19:4077–4090.
- Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J. 2010. Adaptation genomics: the next generation. *Trends Ecol. Evol.* 25:705–712.



- Stewart SA, Hodge S, Ismail N, Mansfield JW, Feys BJ, Prospéri J-M, Huguet T, Ben C, Gentzbittel L, Powell G. 2009. The RAP1 gene confers effective, race-specific resistance to the pea aphid in *Medicago truncatula* independent of the hypersensitive reaction. *Mol. Plant. Microbe Interact.* 22:1645–1655.
- Stuart J. 2015. Insect effectors and gene-for-gene interactions with host plants. *Curr. Opin. Insect Sci.* 9:56–61.
- Stuart JJ, Chen M-S, Shukle R, Harris MO. 2012. Gall midges (hessian flies) as plant pathogens. *Annu. Rev. Phytopathol.* 50:339–357.
- Sugio A, Dubreuil G, Giron D, Simon J-C. 2015. Plant-insect interactions under bacterial influence: ecological implications and underlying mechanisms. *J. Exp. Bot.* 66:467–478.
- Sugio A, Kingdom HN, MacLean AM, Grieve VM, Hogenhout SA. 2011. *Phytoplasma* protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. *Proc. Natl. Acad. Sci.* 108:E1254–E1263.
- Svabova L, Smykal P, Griga M, Ondrej V. 2005. *Agrobacterium*-mediated transformation of *Pisum sativum* in vitro and in vivo. *Biol. Plant.* 49:361–370.
- Tagu D, Le Trionnaire G, Tanguy S, Gauthier J-P, Huynh J-R. 2014. EMS mutagenesis in the pea aphid *Acyrtosiphon pisum*. *G3 GenesGenomesGenetics* 4:657–667.
- Team RC. 2017. R: A language and environment for statistical computing.
- The International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8:e1000313.
- Thomas WJ, Thireault CA, Kimbrel JA, Chang JH. 2009. Recombineering and stable integration of the *Pseudomonas syringae* pv. *syringae* 61 hrp/hrc cluster into the genome of the soil bacterium *Pseudomonas fluorescens* Pf0-1. *Plant J. Cell Mol. Biol.* 60:919–928.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- Thorpe P, Mantelin S, Cock PJ, Blok VC, Coke MC, Eves-van den Akker S, Guzeeva E, Lilley CJ, Smant G, Reid AJ, et al. 2014. Genomic characterisation of the effector complement of the potato cyst nematode *Globodera pallida*. *BMC Genomics* 15.
- Tjallingii WF. 2006. Salivary secretions by aphids interacting with proteins of phloem wound responses. *J. Exp. Bot.* 57:739–745.
- Tjallingii WF, Esch TH. 1993. Fine structure of aphid stylet routes in plant tissues in correlation with EPG signals. *Physiol. Entomol.* 18:317–328.
- Toda S, Komazaki S, Tomita T, Kono Y. 2004. Two amino acid substitutions in acetylcholinesterase associated with pirimicarb and organophosphorous insecticide resistance in the cotton aphid, *Aphis gossypii* Glover (Homoptera: Aphididae). *Insect Mol. Biol.* 13:549–553.

- Trieu AT, Burleigh SH, Kardailsky IV, Maldonado-Mendoza IE, Versaw WK, Blaylock LA, Shin H, Chiou T-J, Katagi H, Dewbre GR, et al. 2000. Transformation of *Medicago truncatula* via infiltration of seedlings or flowering plants with *Agrobacterium*. *Plant J.* 22:531–541.
- Tsuchida T, Koga R, Fukatsu T. 2004. Host plant specialization governed by facultative symbiont. *Science* 303:1989–1989.
- Tsuchida T, Koga R, Horikawa M, Tsunoda T, Maoka T, Matsumoto S, Simon J-C, Fukatsu T. 2010. Symbiotic bacterium modifies aphid body color. *Science* 330:1102–1104.
- Vachon V, Laprade R, Schwartz J-L. 2012. Current models of the mode of action of *Bacillus thuringiensis* insecticidal crystal proteins: a critical review. *J. Invertebr. Pathol.* 111:1–12.
- Van Emden HF, Harrington R. 2007. Aphids as crop pests. Cabi
- Van Valen L. 1977. The Red Queen. *Am. Nat.* 111:809–810.
- Vertacnik KL, Linnen CR. 2017. Evolutionary genetics of host shifts in herbivorous insects: insights from the age of genomics. *Ann. N. Y. Acad. Sci.* 1389:186–212.
- Via S. 1991. The genetic structure of host plant adaptation in a spatial patchwork: demographic variability among reciprocally transplanted pea aphid clones. *Evolution* 45:827–852.
- Via S, Hawthorne D. 2005. Back to the future: genetic correlations, adaptation and speciation. *Genet. Adapt.*:147–156.
- Via S, Hawthorne DJ. 2002. The genetic architecture of ecological specialization: correlated gene effects on host use and habitat choice in pea aphids. *Am. Nat.* 159:S76–S88.
- Vieira FG, Sánchez-Gracia A, Rozas J. 2007. Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biol.* 8:R235.
- Villada ES, González EG, López-Sesé AI, Castiel AF, Gómez-Guillamón ML. 2009. Hypersensitive response to *Aphis gossypii* Glover in melon genotypes carrying the Vat gene. *J. Exp. Bot.* 60:3269–3277.
- Vitalis R, Gautier M, Dawson KJ, Beaumont MA. 2013. Detecting and Measuring Selection from Gene Frequency Data. *Genetics:genetics*.113.152991.
- de Vos M, Kim JH, Jander G. 2007. Biochemistry and molecular biology of *Arabidopsis*-aphid interactions. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 29:871–883.
- Wagner SM, Martinez AJ, Ruan Y-M, Kim KL, Lenhart PA, Dehnel AC, Oliver KM, White JA. 2015. Facultative endosymbionts mediate dietary breadth in a polyphagous herbivore. *Funct. Ecol.* 29:1402–1410.
- Wang W, Dai H, Zhang Y, Chandrasekar R, Luo L, Hiromasa Y, Sheng C, Peng G, Chen S, Tomich JM, et al. 2015a. Armet is an effector protein mediating aphid–plant interactions. *FASEB J.* 29:2032–2045.

- Wang W, Luo L, Lu H, Chen S, Kang L, Cui F. 2015b. Angiotensin-converting enzymes modulate aphid–plant interactions. *Sci. Rep.* 5.
- War AR, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, Sharma HC. 2012. Mechanisms of plant defense against insect herbivores. *Plant Signal. Behav.* 7:1306–1320.
- Wenger JA, Cassone BJ, Legeai F, Johnston JS, Bansal R, Yates AD, Coates BS, Pavinato VAC, Michel A. 2017. Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect Biochem. Mol. Biol.*
- Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. 2007. The genetic basis of a plant–insect coevolutionary key innovation. *Proc. Natl. Acad. Sci.* 104:20427–20431.
- Whitham SA, Lincoln LM, Chowda-Reddy RV, Dittman JD, O’Rourke JA, Graham MA. 2016. Virus-induced gene silencing and transient gene expression in soybean (*Glycine max*) using Bean Pod Mottle Virus infectious clones. In: *Current Protocols in Plant Biology*. John Wiley & Sons, Inc.
- Wiens JJ, Lapoint RT, Whiteman NK. 2015. Herbivory increases diversification across insect clades. *Nat. Commun.* 6:ncomms9370.
- Will T, Furch ACU, Zimmermann MR. 2013. How phloem-feeding insects face the challenge of phloem-located defenses. *Front. Plant Sci.* 4.
- Will T, Tjallingii WF, Thönnessen A, Bel AJE van. 2007. Molecular sabotage of plant defense by aphid saliva. *Proc. Natl. Acad. Sci.* 104:10536–10541.
- Will T, Vilcinskis A. 2015. The structural sheath protein of aphids is required for phloem feeding. *Insect Biochem. Mol. Biol.* 57:34–40.
- Win J, Morgan W, Bos J, Krasileva KV, Cano LM, Chaparro-Garcia A, Ammar R, Staskawicz BJ, Kamoun S. 2007. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* 19:2349–2369.
- Wroblewski T, Piskurewicz U, Tomczak A, Ochoa O, Michelmore RW. 2007. Silencing of the major family of NBS–LRR-encoding genes in lettuce results in the loss of multiple resistance specificities. *Plant J.* 51:803–818.
- Wu J, Baldwin IT. 2010. New insights into plant responses to the attack from insect herbivores. *Annu. Rev. Genet.* 44:1–24.
- Wulff JA, White JA. 2015. The endosymbiont *Arsenophonus* provides a general benefit to soybean aphid (Hemiptera: Aphididae) regardless of host plant resistance (Rag). *Environ. Entomol.* 44:574–581.
- Yassin A, Debat V, Bastide H, Gidaszewski N, David JR, Pool JE. 2016. Recurrent specialization on a toxic fruit in an island *Drosophila* population. *Proc. Natl. Acad. Sci.* 113:4771–4776.
- Yun HS, Kwon C. 2017. Vesicle trafficking in plant immunity. *Curr. Opin. Plant Biol.* 40:34–42.
- Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18:292–298.

- Zhang Z, Thomma BPHJ. 2014. Virus-induced gene silencing and *Agrobacterium tumefaciens*-mediated transient expression in *Nicotiana tabacum*. *Methods Mol. Biol.* Clifton NJ 1127:173–181.
- Zhao C, Escalante LN, Chen H, Benatti TR, Qu J, Chellapilla S, Waterhouse RM, Wheeler D, Andersson MN, Bao R, et al. 2015. A massive expansion of effector genes underlies gall-formation in the wheat pest *Mayetiola destructor*. *Curr. Biol.* 25:613–620.
- Zhao C, Shukle R, Navarro-Escalante L, Chen M, Richards S, Stuart JJ. 2016. Avirulence gene mapping in the Hessian fly (*Mayetiola destructor*) reveals a protein phosphatase 2C effector gene family. *J. Insect Physiol.* 84:22–31.

## RÉSUMÉ

### Diversité, caractéristiques évolutives et rôles des effecteurs salivaires du puceron du pois dans l'interaction avec ses plantes hôtes

Les effecteurs jouent un rôle fondamental lors des interactions antagonistes plantes-pathogènes en supprimant les défenses de la plante, permettant ainsi aux parasites de se développer. De tels effecteurs ont été caractérisés chez les insectes herbivores mais leur rôle dans la spécialisation à la plante reste méconnu. Les pucerons se nourrissent de la sève du phloème et injectent dans la plante des effecteurs salivaires. L'étude des patrons d'évolution des effecteurs, ainsi que la caractérisation de leurs fonctions sont nécessaires à la compréhension des mécanismes de spécialisation chez les pucerons. Au cours de ces travaux, nous avons cherché à identifier les effecteurs salivaires impliqués dans l'adaptation du puceron du pois, *Acyrtosiphon pisum*, à ses hôtes. Des approches évolutives, basées sur un nouveau catalogue de 740 effecteurs candidats surexprimés dans les glandes salivaires de *A. pisum*, ont révélé que certains d'entre eux évoluent rapidement et que l'expansion de familles multigéniques apparaît comme une source importante de diversité des effecteurs. En parallèle, ces travaux ont permis d'optimiser l'expression transitoire médiée par *Agrobacterium* dans le pois. Ce nouvel outil d'analyse fonctionnelle permet maintenant l'étude des effecteurs candidats afin d'identifier les effecteurs du puceron du pois impliqués dans l'adaptation à la plante hôte.

**Mots clés :** effecteurs salivaires, puceron du pois, interactions plantes-pucerons

## ABSTRACT

### *Diversity, evolutionary characteristics and role of pea aphid salivary effectors in the interaction with host plants*

*Effectors play fundamental roles in antagonistic plant-pathogen interactions mainly by suppressing plant defense and allow parasites to multiply on the plant. Some effectors have been characterized in herbivorous insects; however, their role to the evolution in plant specialization remains unknown. Aphids feed from phloem sap and inject salivary effectors into the host plant. Studying evolutionary patterns and characterizing functions of effectors appear as important steps toward unveiling the mechanisms of host plant specialization in aphids. This work sought to identify salivary effectors that are involved in plant specialization of the pea aphid, *Acyrtosiphon pisum*. Evolutionary approaches based on a new catalogue of 740 putative effectors that are up-regulated in salivary glands of *A. pisum* revealed that some of them evolve rapidly. Moreover, gene family expansion appear as an important source of novel effectors. In parallel, this work optimized *Agrobacterium*-mediated transient gene expression in pea to provide a new tool for functional analyses of pea aphid effectors. The construction of a comprehensive catalogue of *A. pisum* salivary effectors and evolutionary analysis of them provide new candidates in host plant adaptation. By using the gene expression tool now available in pea, functional characterization of candidates will help to identify the effectors that are involved in plant specialization of the pea aphid.*

**Keywords:** salivary effectors, pea aphid, plant-aphid interactions



AGROCAMPUS OUEST • Institut supérieur des sciences agronomiques, agroalimentaires, horticoles et du paysage  
65 rue de Saint-Brieuc - CS84215 - F-35042 Rennes Cedex  
Tél. : 02 23 48 50 00  
[www.agrocampus-ouest.fr](http://www.agrocampus-ouest.fr)

