Cuticular Structure Proteomics in the Pea Aphid Acyrthosiphon pisum Reveals New Plant Virus Receptor Candidates at the Tip of Maxillary Stylets

Maelle Deshoux,† Victor Masson,‡ Karim Arafah, Sébastien Voisin, Natalia Guschinskaya, Manuella van Munster, Bastien Cayrol, Craig G. Webster, Yvan Rahbé, Stéphane Blanc, Philippe Bulet,*‡ and Marilyne Uzest*‡

Cite This: J. Proteome Res. 2020, 19, 1319−1337

INTRODUCTION

Aphids are phloem-feeding insects known as major pests in agriculture that are able to transmit hundreds of plant viruses. The majority of these viruses, classified as noncirculative, are retained and transported on the inner surface of the cuticle of the needle-like mouthparts while the aphids move from plant to plant. Identification of receptors of viruses within insect vectors is a key challenge because they are promising targets for alternative control strategies. The acrostyle, an organ discovered earlier within the common food/salivary canal at the tip of aphid maxillary stylets, displays proteins at the cuticle−fluid interface, some of which are receptors of noncirculative viruses. To assess the presence of stylet- and acrostyle-specific proteins and identify putative receptors, we have developed a comprehensive comparative analysis of the proteomes of four cuticular anatomical structures of the pea aphid, stylets, antennae, legs, and wings. In addition, we performed systematic immunolabeling detection of the cuticular proteins identified by mass spectrometry in dissected stylets. We thereby establish the first proteome of stylets of an insect and determine the minimal repertoire of the cuticular proteins composing the acrostyle. Most importantly, we propose a short list of plant virus receptor candidates, among which RR-1 proteins are remarkably predominant. The data are available via ProteomeXchange (PXD016517).

KEYWORDS: cuticle, cuticular proteins, stylets, acrostyle, virus receptor, proteomics, aphid

ABSTRACT: Aphids are phloem-feeding insects known as major pests in agriculture that are able to transmit hundreds of plant viruses. The majority of these viruses, classified as noncirculative, are retained and transported on the inner surface of the cuticle of the needle-like mouthparts while the aphids move from plant to plant. Identification of receptors of viruses within insect vectors is a key challenge because they are promising targets for alternative control strategies. The acrostyle, an organ discovered earlier within the common food/salivary canal at the tip of aphid maxillary stylets, displays proteins at the cuticle−fluid interface, some of which are receptors of noncirculative viruses. To assess the presence of stylet- and acrostyle-specific proteins and identify putative receptors, we have developed a comprehensive comparative analysis of the proteomes of four cuticular anatomical structures of the pea aphid, stylets, antennae, legs, and wings. In addition, we performed systematic immunolabeling detection of the cuticular proteins identified by mass spectrometry in dissected stylets. We thereby establish the first proteome of stylets of an insect and determine the minimal repertoire of the cuticular proteins composing the acrostyle. Most importantly, we propose a short list of plant virus receptor candidates, among which RR-1 proteins are remarkably predominant. The data are available via ProteomeXchange (PXD016517).

Received: December 20, 2019
Published: January 28, 2020
mandibular stylets to facilitate penetration into plant tissues.\textsuperscript{15,16} Maxillary stylets are tightly interlocked by a series of longitudinal ridges and grooves, which enclose the food and salivary canals formed by opposing grooves. Food and salivary canals fuse a few micrometers from the tip, leading to the confluent common food/salivary canal, that has long been believed to harbor receptors of noncirculative viruses.\textsuperscript{8,17−20} The existence of such receptors has definitely been evidenced with \textit{Cauliflower mosaic
virus (CaMV), which was shown to bind to cuticular proteins (CPs) accessible at the tip of maxillary stylets, at the surface of a specific organ designated the acrostyle.21 This organ is a distinct anatomical structure restricted to a region of the common canal. Its surface properties are different from the rest of the stylet cuticle.22 Antibodies targeting CPs from the CPR family—the largest CP family widespread among arthropods and containing the Rebers and Buddos S (RR) consensus sequence—allowed the detection of several peptides from RR-1 and RR-2 subgroups in the acrostyle.22–26 More specifically, two highly homologous RR-1 proteins, Stylin-01 and Stylin-02, have been localized in the organ with a peptide corresponding to their common C-terminus sequence directly accessible at the surface. In vitro competition assays and in vivo silencing approaches indicated that Stylin-01 was involved in CaMV transmission.26 However, its role in the transmission of other noncirculative viruses could not be established, and there is currently no evidence for a common receptor of all noncirculative viruses. The acrostyle has not be established, and there is currently no evidence for a common receptor of all noncirculative viruses. The acrostyle has not be established, and there is currently no evidence for a common receptor of all noncirculative viruses.

EXPERIMENTAL PROCEDURES

The experimental design and the organization of the presentation of the results are summarized in the workflow shown in Figure 1.

Chemical Reagents

For sample preparation, all reagents were purchased from Sigma-Aldrich (St. Louis, MO) except RepliGest SF surfactant purchased from Waters (Milford, MA). For LC-MS/MS analysis, formic acid was obtained from Fluka (Sigma-Aldrich, St. Louis, MO), Milli-Q water from Merck (Merck Millipore, Billerica, MA), and all other chemicals were purchased from Carlo-erba Reagents (Val de Reuil, France). Modified sequencing-grade trypsin (Promega Corporation, Madison, WI) was used for protein digestion.

Aphid Clones

A colony of Acrystosiphon pisum (LL01) was maintained on Vicia faba cv Robin Hood in an environmental growth chamber at 23–18 °C (day/night) with a photoperiod of 16/8 h (day/night). Myzus persicae Sulzer was maintained on Solanum melongena cv Barbentane in the same temperature/photoperiod conditions.

Tissue Collection for Liquid Chromatography Coupled to Tandem Mass Spectrometry (LC-MS/MS) Analyses

Alate A. pisum adults collected from Vicia faba plants were starved for 1 h and stored at −20 °C for several hours before dissection. Collection of antennae, legs, and wings did not present any technical difficulties. These appendages were collected in triplicate for each structure, each replicate comprising antennae, wings, or legs from 37 insects. Stylet bundles, hereafter named stylets, are approximately 700 μm in length and 3 μm in diameter for A. pisum adults. They are anchored in aphid heads in the glands from which they are secreted, the retort organs (described by Ponsen and Gustninskaya and colleagues).14,28 To extract exploitable information from MS analyses, hundreds of stylets had to be pooled for each trial as free as possible of other contaminating aphid tissues. In addition, we had to reduce electrostatic charges that too often induced stylets fly off and loss. For these reasons, a “clean-stylet” dissection protocol compatible with MS standards had to be specifically designed (Figure 1A). Frozen aphids were first glued on their back onto a microscope slide using double-sided bonding tape (Figure 1A1). Stylets were then pulled out from the labium under a stereomicroscope using tungsten insect pins (Figure 1A2), separated from the head with microscissors (Figure 1A3), and then transferred onto a conductive glass slide (ITO, Bruker Daltonik, Germany), allowing secured accumulation of stylets (Figure 1A4). Dissecting tools were carefully washed with water, followed by 70% ethanol, and finally dried on lint-free paper after each stylet collection. Once 500 stylet bundles had been accumulated onto the ITO-slide, they were carefully transferred with a fine needle into a single glass microtube, avoiding static charge effects, allowing control of the number of appendages deposited per tube (Figure 1A5). This protocol was systematically used to collect three independent batches of 500 stylets each.

Sample Preparation

All cuticular structures were prepared according to the same protocol. Samples were first washed for 15 min with Milli-Q water and 50% acetonitrile (ACN), both acidified with 0.1% trifluoroacetic acid (TFA) (v/v, final concentration). Supernatants were then removed after centrifugation, and the cuticular structures dried out by centrifugation under vacuum (Labconco, Kansas City, USA). All anatomical structures were then treated as described previously.27 Briefly, proteins were extracted by successive incubations in pure hexafluorosilicopha-
nol (HFIP) and in 50 mM ammonium bicarbonate supplemented with 0.1% RapiGest. After reduction and alkylation, proteins were submitted to trypsin digestion. Finally, digested samples were dried out and resuspended in 2% ACN/0.1% TFA (v/v) prior to mass spectrometry (MS) analysis.

**Nano Liquid Chromatography Coupled to Tandem Mass Spectrometry (nanoLC-MS/MS) Analysis**

NanoLC-MS/MS was carried out using an Ultimate 3000 nanoHPLC (Thermo Scientific, Germany) for the separation, hyphenated to a Q-Exactive Orbitrap mass spectrometer (Thermo Scientific). For chromatography, the digested samples were loaded, concentrated, and washed at 10 μL/min for 6 min with 2% ACN and 0.05% TFA on a microconcentrating column (300 μm × 5 mm PepMap 100, C18, 5 μm, Thermo Scientific). The separation was performed on a reversed-phase column (75 μm × 250 mm Acclaim PepMap 100, C18, particle size 3 μm nanowiper column from Thermo Scientific). The LC mobile phases for the separation were water (A) and ACN (B), each supplemented with 0.1% formic acid (v/v). Separation was achieved at a flow rate of 300 nL/min using a biphasic linear gradient from 2% to 32% B in 100 min and from 32% to 65% B in 5 min. MS analysis was carried out in positive ion and data-dependent modes. The voltage applied to the nanotips (Nano Objective, USA) was approximately 2.0 kV, and the header was at 300 °C. Full scan (MS) spectra were obtained from 380 to 2,000 m/z (70,000 resolution, AGC target 3 × 10^6, maximum IT 200 ms), and for each full-scan the most intense ions (Top 10) were fragmented in MS² (17,500 resolution, AGC target 2 × 10^4, maximum IT 100 ms, intensity threshold 4 × 10^4, excluding charge-unassigned ions, Normalized Collision Energy selected at 27). Parent ions were then excluded from MS/MS for the next 15 s. The softwares Chromeleon Xpress (Thermo Fisher Scientific) was used to control the HPLC, and Xcalibur 2.2 (Thermo Fisher Scientific) to control the mass spectrometer.

**Database Searching and Protein Identification**

The Sequest HT searching algorithm was run by Proteome Discoverer 1.4 (Thermo Fisher Scientific) to match the acquired MS/MS spectra to a protein database, with the following settings: trypsin digest with two maximum missed cleavages, 6 and 144 amino acids as minimum and maximum peptide length, respectively, a tolerance of 10 ppm/0.02 Da for precursors and fragment ions, respectively. Cysteine carbamidomethylation was set as a fixed modification; C-terminal protein amidation and methionine and tryptophan oxidation were set as variable modifications.

Searches were performed against combined proteome databases (Table S1) including: (i) the pea aphid proteome (http://aphidbase.com, version 2.0 containing 33,291 protein-coding sequences); (ii) an eubacterial reference proteome setup for putative symbionts and aphid microbiome-related sequences (116,983 sequences); (iii) a set of sequences of entomopathogenic fungi (28,431 sequences); (iv) a set of viral sequences including aphid viruses and aphid-vectored broad-bean viruses were also selected (6,614 sequences); (v) a mixed plant proteome assembled from available *Fabaceae* proteomes (66,777 sequences); and (vi) a set of common contaminants (116 sequences).

A homemade *A. pisum* CP database comprising 150 annotated CPs was constructed using the Cuticular Protein Family Prediction Tool CutProtFam-Pred (http://aiais.biol.oua.gr/CutProtFam-Pred/home.php) to accurately detect and classify putative CPs present in the pea aphid proteome v2.0, thanks to known consensus and semiconensus sequences (Table S2). Identification of CPs in cuticular structures was carried out with this homemade database.

**Data Presentation**

To identify the proteins that are present in pea aphid cuticular anatomical structures, pools of antennae, wings, legs, and stylo bundles were collected in triplicate on *A. pisum* adults. Proteins consistently identified in all three biological replicates with at least a same peptide in each replicate, regardless of being “Unique” (peptide present in only one protein) or “Shared” (peptide present in multiple different proteins), were considered as confidently identified and constituted whole proteomes of the cuticular structures. In these first data sets, proteins sharing a same identified peptide were grouped under a single protein identifier (accession number corresponding to the protein with the highest score and highest percentage of coverage of the protein. These parameters can differ from one sample to another, and a same “shared” peptide may be assigned to different protein identifiers by the algorithm). Comparative analyses of whole proteomes were carried out using these data sets.

Minimal lists of CPs were then retrieved from whole proteomes using homemade CP database (Table S2). Minimal lists represent the smallest number of CPs that can explain all identified MS peptides and will be used to design peptides for antibodies production and immunolabeling of dissected styles. However, CPs can share high sequence similarity, particularly proteins of the CPR family. Several proteins can be identified with a same tryptic fragment (“shared” peptide). In this case, we cannot tell how many of these proteins are actually incorporated into the cuticle of the anatomical structures. To be fully exhaustive in our quest for virus receptor candidates, lists of CPs potentially present in cuticular structures were extended to all proteins containing these “shared” peptides.

**Functional Classification**

Gene Ontology (GO) terms were assigned to AphidBase 2.0 (http://bipaa.genouest.org/is/aphidbase/acyrthosiphon_pism/downloads/) via Blast2GO software version 5.1.12 using the InterProScan with the Hidden Markov Models (HMMs) present in the PFAM protein family database. The number of proteins identified in each cuticular structure was compared to all corresponding functionally GO annotated terms in the entire pea aphid (*A. pisum*). GO term enrichment was determined via Fisher’s Exact test at an FDR, p-value ≤0.05; the reduced enrichment was determined at an FDR, p-value ≤0.01.

**Antibodies**

A total of 26 antibodies were used to immunolabel aphid styles. Eleven antibodies were already available from previous studies.7,22,23,25,26 In addition, 15 peptides of 13–16 amino acids length with sequences originating from peptides retrieved from MS analyses (either “Unique” or “Shared” peptides), or originating from the full-length sequences of the proteins identified in this study were synthesized by Eurogentec (http://www.eurogentec.com). Eurogentec also carried out the production of antibodies in rabbit and the affinity purification of the antisera. Alexa Fluor 488-conjugated antirabbit IgG (A11070, Thermo Fisher Scientific, Waltham, MA) were used as secondary antibodies.

**Immunolabeling of Dissected Styles**

*A. pisum* and *M. persicae* styles were dissected and immunolabeled according to Webster and colleagues, using
primary antibodies at dilutions of 1:200 and secondary Alexa fluor-conjugated antibody at a dilution of 1:400.25 stylers were either untreated or treated with 2 U/mL of Chitinase from *Streptococcus gрисеus* (Sigma-Aldrich, St. Louis, MO) before their incubation with primary antibodies, as described by Uzest and colleagues, to eventually reveal epitopes that could be localized under the surface layers of the sty. cutica.21 When no labeling was observed after 15 min of digestion, the Chitinase treatment was extended to 30 and 60 min. Two independent repeats were performed per condition and per antibody.

**Quantification of Stylin Transcripts in Various Tissues and at Different Developmental Stages**

Transcripts of stylins were quantified in the heads of all *A. pisum* nymphal instars (N1 to N4) and adult stages and in different tissues. Insect tissues (antennae, legs, wings, and antenna-free heads as a proxy for sty. synthesizing glands) were collected from adults under a stereo.scope. Six pools of 10 aphids/tissues were tested per condition. Total RNA was extracted from whole aphids or dissected tissues using an RNase. ily minikit (Qiagen, Hilden, Germany). A total of 80–100 ng of RNA was treated with RQ1 RNase-free DNase I (Promega Corporation, Madison, WI). First-strand cDNA was synthesized using Moloney murine leukemia virus (MMLV) reverse transcriptase (Promega Corporation, Madison, WI) according to the manufacturer’s instructions, with oligo(dT) as primers. All RT-qPCRs were performed in duplicates on a LightCycler 480 instrument (Roche, Basel, Switzerland) using 1:4 diluted cDNAs and a LightCycler 480 SYBR green I master mix (Roche, Penzberg, Germany) according to the manufacturer’s recommendations with gene-specific primers (Table S5). Two internal reference genes encoding actin and elongation factor 1α (EF1α) from *A. pisum* were used for normalization. Amplification efficiencies were analyzed with LinRegPCR free software (v. 2014.5). Relative expressions were calculated using the threshold cycle (2−ΔΔCt) method.25

**Alignments and Phylogenetic Analyses**

Sequences encoding stylin proteins were retrieved from the seven genomes available on AphidBase (http://bipaa.genouest.org/is/aphidbase/)—*Myzus persicae* (Mp; clone G006), *Myzus cerasi* (Mc), *Rhopalosiphum padi* (Rp), *Diuraphis noxia* (Dn), *A. pisum* (Ap), *Daktulosphaira vitifoliae* (Dv), *Aphis glycines* (Ag; Biotype 4)—using CutprotFam-Pred with standard settings.29,36–39 Among these sequences, 17 were manually curated for the present study.

After removal of the predicted signal peptides using the SignalP-5.0 server, CPR_RR-1 mature protein sequences were used in phylogenetic analyses.40 For CPAP3 proteins, full sequences were used in phylogenetic analyses. Sequences were aligned using MUSCLE (v3.8.31) configured for highest accuracy (MUSCLE with default settings).41 Ambiguous regions (i.e., containing gaps and/or poorly aligned) were removed with Gblocks (v0.91b) using the following parameters: minimum length of a block after gap cleaning: 10, no gap positions were allowed in the final alignment and all segments with contiguous nonconserved positions bigger than eight were rejected, minimum number of sequences for a flank position: 85%. Phylogenetic trees were reconstructed using the maximum likelihood method implemented in the PhyML program (v3.1/3.0.lR). The WAG matrix, which works as empirical aminoacid substitution model to simulate the biological sequence evolution with flexibility, was selected assuming an estimated proportion of invariant sites (of 0.009) and four gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma = 3.517). Reliability for internal branch was assessed using the aLRT test (SH-Like). Graphical representation and edition of the phylogenetic trees and alignments were performed with TREEDYN (v. 198.3) and the T-Coffee software, respectively.

## RESULTS

**Comparative Analyses of the Proteome of Four Cuticular Structures of the Pea Aphid *A. pisum***

To identify sty. specific CPs, we first characterized and compared the proteomes of four cuticular structures: two muscular articulated segmented tissues, antennae and legs; wings which are outgrowths of the exoskeleton; and sty. bundles comprising two mandibular sty. harboring dendrites and two maxillary sty. that are cell-free cuticular structures.

6,574 peptides were identified in antennae, 6,025 in legs, and 7,242 in wings, which were specifically assigned to 971, 749, and 910 proteins. In sty. a smaller number of 1,118 peptides were identified and assigned to 141 proteins (Table 1, Tables S4–S7). Proteins identified in legs, antennae, and wings represent 2.2–2.9% of the whole *A. pisum* proteome, while proteins identified in sty. represent only 0.4%. Of the 1,273 proteins identified in this study, 90 were found in all four cuticular structures (7%) and constitute what we hereafter refer to as the “core proteome” of cuticular structures. 426 proteins were found to be structure-specific (33%), with 162, 59, 165, and 40 proteins solely identified in antennae, legs, wings, and sty. respectively (Figure 2A, Table S8).

### Table 1. Summary of Proteins Identified in *Acyrthosiphon pisum* Cuticular Structures: (A) All Peptides and Proteins Recovered from LC-MS/MS Analyses; (B) Peptides and Proteins Assigned to *A. pisum* Proteome (*A. pisum* ID) Recovered from LC-MS/MS Analysis; (C) Cuticular Proteins (CPs) Assigned to *A. pisum* Recovered from LC-MS/MS Analysis

|         | Antenna | Legs   | Wings  | Stylets |
|---------|---------|--------|--------|---------|
| A Total # of peptides identified | 6590 | 6040 | 7260 | 1119 |
| B Total # of peptides identified (A. pisum ID) | 6574 | 6025 | 7242 | 1118 |
| C Total # of CPs identified—minimal list | 39 | 36 | 41 | 15 |
| Ratio CPs/Total # of proteins (A. pisum ID) % | 4.0 | 4.8 | 4.5 | 10.6 |

*# number; CPs, Cuticular Proteins; ID, identifier. Smallest number of cuticular proteins that can explain all observed peptides recovered from LC-MS/MS analysis. All cuticular proteins recovered from LC-MS/MS analysis identified with "Unique" peptides and "Shared" peptides.
As expected for cuticle-containing insect body parts, the GO terms overrepresented in the core proteome were related to the extracellular region and were significantly enriched in cuticle-related functions and the sclerotization process. Hence, categories such as structural constituent of cuticle, chitin binding, carbohydrate metabolic process, chitin metabolic process, oxidoreductase activity, and oxidation−reduction process were overrepresented (Figure 2B, Table S9). Other categories related to unfolded protein binding or small molecules binding, metabolic processes, proton transmembrane transport, transferase or GTPase activities were also identified.

Proteomes of antennae, legs, and wings were enriched in functions related not only to the extracellular region but also to categories associated with intracellular compartments such as ribosome, proteasome, mitochondria, and vesicles (Figure 3A−C, Tables S10−S12). More categories were overrepresented in these three structures compared to aphid stylets (31−41 categories vs 16 categories, respectively; see Figure 3A−D). Notably, the chitin metabolic process category is not specially enriched in stylets (Figure 3A−D), emphasizing the strictly extracellular origin of stylet cuticle, while the other appendages contain the chitin-synthesizing epithelia and its intracellular functions. Apart from cuticle-related functions, antennae, legs, and wings contained proteins with commonly enriched functions associated with protein synthesis, protein folding, intracellular trafficking, and small molecules binding. In comparison, categories enriched in stylets were mostly related to cuticle synthesis or sclerotization pathways (Figure 3D, Table S13). Not surprisingly, our data sets highlighted differences in proteomic composition between nearly exclusively cuticle-based structures (i.e., stylets) and only partly cuticle-based tissues also including cellular machineries (i.e., antennae, legs, and wings) (Figure 3E).

**Styllet-Specific Proteins**

Forty styllet-specific proteins have been identified in our MS data sets (Figure 2A, Table S8), most of which were not previously annotated. Although not possible to ascertain, at least 13 of these styllet-specific proteins (32.5%) likely came from the saliva remaining in the salivary canal after dissection of styllet bundles. Consistent with this hypothesis, 12 out of these 13 proteins have already been reported in the pea aphid saliome (Table S8), and one is a predicted secreted peroxidase, an enzyme found in aphid saliva. Moreover, two cuticle-related proteins were only present in the stylets data sets, one trehalase (ACYPI007462) reported to play a role in several processes, including chitin biosynthesis in other insects, and one RR-2 protein (ACYPI006670). Whereas the saliva-related proteins and trehalase were identified with unique peptides not detected in the three other cuticular structures, the RR-2 protein was solely identified from two shared peptides (GSYSSELADGSTR and TVEYTTADYNFNAVVK), also identified in antennae, legs, and wings. Because these peptides have been assigned to ACYPI006670 in the stylets, but to other RR-2 proteins in antennae, legs, and wings by the data search program (Tables S14−S17), we cannot strictly conclude that this protein (ACYPI006670) is stylet-specific.

**Distribution of the Identified CPs in the Different Investigated Cuticular Anatomical Structures**

Minimal lists of CPs recovered by MS analyses comprise 39 CPs in antennae, 38 CPs in legs, 41 CPs in wings, and 15 CPs in stylets (Tables 1 and 2, Tables S14−S17). No peptides from the CPF and CPCFC families were identified in any cuticular structure. The representation of distinct CP families was similar in antennae, legs, and wings (Figure 4A, Table 3), with proteins of the CPR (RR-1 and RR-2), CPAP1, CPAP3, and TWDL families. In stylets, only two protein families were represented, CPR (RR-1 and RR-2) and CPAP3.

RR-2 proteins were the most abundant in all structures and were often identified with shared peptides (Table 2). RR-1 and TWDL proteins were identified either with shared or with unique peptides. Finally, CPAP1 and CPAP3 were only identified with unique peptides (Table 3). The number of CPs unequivocally identified with unique peptides was 26, 26, 30, and 9 in antennae, legs, wings, and antennae, respectively. Comparative analysis of CPs potentially present in cuticular structures indicated that only a few proteins were restricted to a single structure: one CPAP3 in antennae (ACYPI001579), one CPAP1 (ACYPI004632), and two RR-2 in legs (ACYPI006712 and ACYPI003698), one CPAP1
Figure 3. Comparative analysis of GO terms enriched by identified proteins in antennae, legs, wings, and stylets of A. pisum. (A–D) Comparison of identified proteins in the four cuticular structures for Molecular Function (MF, green), Cellular Component (CC, orange), and Biological Process (BP, blue). Histograms represent significantly enriched functional terms in antennae (A), legs (B), wings (C), and stylets (D). Bars represent the proportion of enriched genes encoding the identified proteins compared to all genes involved in the specified GO terms (%Gene/Term). Numbers of genes enriched for each functional group are displayed on the bars. *, p < 0.05; **, p < 0.01; ***, p < 0.001. (E) WordClouds summarize the functional profiles of each cuticular anatomical structure. See also Tables S10–S13.
To confirm the presence of the 38 CPs identified by MS in aphid stylets and to give information on their localization and accessibility, *A. pisum* stylets were dissected and immunolabeled with antibodies targeting peptides from these proteins (Table 4). All CPs located in the acrostyle with peptides directly accessible at its surface will be renamed stylins.

Four to five RR-1 proteins were identified in our MS analyses (Table 4) including Stylin-01 (ACYPI009006) and Stylin-02 (ACYPI003649). These two stylins have previously been detected in *A. pisum* stylets using specific antibodies.26

A third RR-1 protein, ACYPI001610, was identified with a unique peptide of 13 AA covering 10.8% of the mature protein (Table 4, Table S17). This peptide could not be detected with an anti-1-03 specific antibody in our experimental conditions despite extensive Chitinase digestion treatment (Figure 5A). Three additional antibodies targeting distinct peptides of ACYPI001610, anti-1-04, anti-1-15, and anti-1-17 were then used to detect this protein in aphid stylets (Table 4). Anti-1-04 and anti-1-17 antibodies revealed their corresponding peptides at the tip of aphid maxillary stylets in the common canal and on the lateral edge of the stylet. Labeling was only observed after preincubation of stylets with Chitinase, indicating that these two peptides were embedded in the chitin matrix (Figure 5A). A stronger labeling was observed when using anti-1-15 antibody, which labeled evenly the acrostyle without the need for Chitinase treatment. Noticeably, the targeted peptide VEGGY-SYTAPDGTPI is part of the RR-1 chitin-binding domain, and its labeling decreased upon Chitinase digestion (Figure 5A). The protein ACYPI001610 is thus accessible at the surface of the acrostyle and was renamed Stylin-03.

The fourth RR-1 protein, ACYPI002877, was identified by MS analyses with two shared peptides also found in ACYPI000308 (Table 4, Table S17). These two proteins are 96.5% identical and cannot be distinguished by immunolabeling. Anti-1-14 antibody, which targets the peptide GSYTFGYQ-SADGTQR, labeled the acrostyle and the lateral edge tip of maxillary stylets of the stylet tip after a Chitinase treatment. The other identified peptide detected in our MS data sets (Table 4, Table S17) and present within the N-terminus of the two proteins was directly accessible at the surface of the upper part of the acrostyle in close vicinity of the food canal and on a lateral edge by anti-1-16 antibody (Figure 5A). At least one of these two RR-1 proteins is thus present in the acrostyle, and they have consequently been renamed Stylin-04 and Stylin-04bis.

Localization of RR-2 Proteins in the Stylets

Eight to 30 RR-2 proteins have been identified in stylets, among which only three were identified by unique peptides. Due to high sequence similarities, even unique peptides were sometimes nearly identical to others present in distinct RR-2 proteins and could not be distinguished by immunological approaches.

Solely two RR-2 proteins were unambiguously detected in the acrostyle. ACYPI066095, identified with four peptides, was detected with anti-2-16 specific antibody. The unique peptide targeted by this antibody was found embedded in the acrostyle, only revealed after Chitinase treatment (Figure 5A, Table 4). The second RR-2 protein, ACYPI007858, was identified by MS with a single peptide of 11 AA covering 2% of the mature protein (ACYPI45536), one CPAP3 (ACYPI000583), one RR-1 (ACYPI001775), three RR-2 in wings (ACYPI000670, ACYPI001681, ACYPI006015), and none in the stylets (Figure 4B, Table 2, Table S18).

### Table 2. Exhaustive List of Cuticular Proteins Identified in the Four Cuticular Structures

| Protein ID  | Antigen E  | Tags | Mass | Stylen | Stylen |
|-------------|------------|------|------|--------|--------|
| ACYPI003094 | U          | U    | U    | Stylen | Stylen |
| ACYPI003160 | U          | U    | U    | Stylen | Stylen |
| ACYPI003210 | U          | U    | U    | Stylen | Stylen |
| ACYPI003248 | U          | U    | U    | Stylen | Stylen |
| ACYPI003251 | U          | U    | U    | Stylen | Stylen |
| ACYPI003265 | U          | U    | U    | Stylen | Stylen |

*U = unique peptide found in protein. S = peptide shared among different proteins. The 15 CPs identified by mass spectrometry in aphid stylets are indicated in red (see also Table S17). RR-1 proteins are highlighted in green, RR-2 in grey, CPAP1 in yellow, CPAP3 in orange, and Tweedle in blue. ID, identifier.*
and targeted by anti-2-13 antibody. This peptide was also embedded in the acrostyle, as well as in the lateral edge of the stylets (Figure 5B, Table 4). Attempts to detect another unique peptide of this CP with anti-2-14 antibody failed whatever the conditions used. A third RR-2 protein, ACYPI001599 was detected by MS with a unique peptide, but no specific antibody could be produced.

Four peptides included in the chitin-binding domain and common to several identified RR-2 proteins have earlier been reported as embedded within the acrostyle—pepL, pepS, Ap2-05, and Ap2-08.\textsuperscript{22,25} We here extend the listing of RR-2 protein peptides detectable in stylets with four additional antibodies. Anti-Ap2-04, anti-Ap2-06, and anti-Ap2-07 are directed against peptides of the conserved RR-2 chitin-binding domain, whereas anti-2-10 targets one peptide present in the C-terminus of a few proteins. All four antibodies labeled maxillary stylets after extended Chitinase treatments. Strong fluorescent labeling appeared as dots along maxillary stylets when using anti-2-06 and anti-2-10 antibodies, indicating that RR-2 proteins are widely present under the surface (Figure 5B, Table 4).

ACYPI066095 and ACYI007858 formally detected within the acrostyle with antibodies targeting unique peptides lack several shared peptides detected in the organ, among which pepL, pepS, and Ap2-08. Therefore, another RR-2 protein containing these shared peptides may be present in the organ. We can conclude that at least three RR-2 proteins are detected in the acrostyle (Figure 5B).

### Localization of CPAP3 Proteins in the Stylets

Three CPAP3 have been identified in the stylet data sets. With the set of antibodies produced and used against these proteins, the labeling was generally weak and visible as dots at the tip of maxillary stylets (Figure 5C, Table 4). ACYPI007860 was only barely revealed in the acrostyle after Chitinase treatment with a single antibody, anti-3-01, targeting a peptide that was not detected by MS. No labeling could be visualized when using anti-3−01 antibody targeting one unique MS-identified peptide of this protein. Its location within aphid stylets could therefore not be definitely stated. ACYPI006031 and ACYPI007911 were detected in the acrostyle, each with two distinct antibodies. Two peptides of ACYPI006031, respectively targeted by anti-3-06 and anti-3-07 were found embedded within the cuticle. For ACYPI007911, the peptide targeted by anti-3-03 was detected only after Chitinase treatment, whereas that targeted by anti-3-09 antibody was detected directly at the surface of the acrostyle (Figure 5C). This protein was renamed Stylin-05.

### Repertoire of Cuticular Proteins Present in the Acrostyle

Of the 15 CP groups identified in stylet bundles, nine proteins were unambiguously detected in the acrostyle. These proteins belong to two CP families, including seven CPR proteins (4 RR-1 and 3 RR-2) and two CPAP3 (Figure 6). Four peptides have been shown to be directly accessible at the surface: three of them are found in RR-1 proteins and are either part of the chitin-binding domain (Stylin-03) or present at the N-terminus and C-terminus of the proteins (Stylin-04/-04bis and Stylin-01/-02, respectively), and one belongs to a CPAP3 protein and is part of...
one of its chitin-binding domain (Figure 7). Noticeably, no RR-2 protein (the most numerous CP class) was detected at the untreated surface of the acrostyle.

Spatial and Temporal Expression of Stylins in *A. pisum*

None of the five proteins detected at the surface of the acrostyle was found to be stylet-specific. They have been identified in all four cuticular structures characterized in this study, except for Stylin-04/-04bis which was absent from the wing’s proteome (Table 4). To characterize possible difference in stylin genes expression in different tissues, we compared their transcript levels in antennaless-heads containing the stylet-synthesizing glands, and in antennae, legs, and wings by real-time RT-qPCR analyses using stylin-specific primers for Stylin-01, -02, -03, and -05 (Table S3). For Stylin-04/-04bis, we could not detect any expression.

| CP ID | Peptides identified by MS or present in the CP | Peptides used for immunization | Antibody ID | Labeling / Accessibility |
|-------|-----------------------------------------------|--------------------------------|-------------|--------------------------|
| ACYP001599 | - VYETTADYGFPNAVYK | n.d. | N/A | N/A |
| ACYP001610 | - YDNQGFDGSRK | n.d. | N/A | N/A |
| ACYP001644 | - GPTSYSAAPYK | n.d. | N/A | N/A |
| ACYP002877 | - FYNNQNPYQGFPQILQSF | n.d. | N/A | N/A |
| ACYP003649 | - AVLSQFGVGFNFK | n.d. | N/A | N/A |
| ACYP004943 | - APYSAAPAYK | n.d. | N/A | N/A |
| ACYP006031 | - DQGPGVFGNFPDGNGGR | n.d. | N/A | N/A |
| ACYP006670 | - GSYLLEADSGT | n.d. | N/A | N/A |
| ACYP007858 | - LPGGATAPAR | n.d. | N/A | N/A |
| ACYP007860 | - ENCDYLVDDCGR | n.d. | N/A | N/A |
| ACYP007911 | - SELQPNGPQSYLPR | n.d. | N/A | N/A |
| ACYP009006 | - AAILQDAPSADGSKF | n.d. | N/A | N/A |
| ACYP006095 | - APAYAAPAPAYK | n.d. | N/A | N/A |
| ACYP071307 | - SQNQADANYGK | n.d. | N/A | N/A |
| ACYP26371 | - GSYLLEADGSTR | n.d. | N/A | N/A |

Note: n.d. = not done; N/A = not applicable; +: labeling detected; (−): no labeling observed under our experimental conditions; Ac, epitope exposed at the surface and directly accessible; PAc, epitope poorly accessible (labeling visible as dots without Chitinase treatment); Em, epitope embedded, not accessible at the surface of the cuticles. CP identified with unique peptides are in red; peptides not identified by LC-MS/MS analysis for which antibodies were already available in our laboratory are in italics. Differences in the AA sequence are indicated in blue. Antibody ID is followed by a reference number when the corresponding antibody was described in a previous study.

**Table 4. Correspondence between Peptides/Proteins Identified in *A. pisum* Stylets by MS Analysis and Peptides Used for Antibody Production**

AA, amino acid; CP, Cuticular protein; ID, identifier; n.d., not done; N/A, not applicable; (+): labeling detected; (−): no labeling observed under our experimental conditions; Ac, epitope exposed at the surface and directly accessible; PAc, epitope poorly accessible (labeling visible as dots without Chitinase treatment); Em, epitope embedded, not accessible at the surface of the cuticles. CP identified with unique peptides are in red; peptides not identified by LC-MS/MS analysis for which antibodies were already available in our laboratory are in italics. Differences in the AA sequence are indicated in blue. Antibody ID is followed by a reference number when the corresponding antibody was described in a previous study.

**Spatial and Temporal Expression of Stylins in *A. pisum***

None of the five proteins detected at the surface of the acrostyle was found to be stylet-specific. They have been identified in all four cuticular structures characterized in this study, except for Stylin-04/-04bis which was/were absent from the wing’s proteome (Table 4). To characterize possible difference in stylin genes expression in different tissues, we compared their transcript levels in antennae-heads containing the stylet-synthesizing glands, and in antennae, legs, and wings by real-time RT-qPCR analyses using stylin-specific primers for Stylin-01, -02, -03, and -05 (Table S3). For Stylin-04/-04bis, we could not detect any expression.
not design primers able to specifically amplify a single transcript. The observed patterns are thus difficult to interpret and provided for information purposes only.

Consistent with the proteomic data, *stylin*-01, -02, -03, and -05 were expressed not only in the head but also in other body parts. They exhibited different expression patterns, with *stylin*-01 and *stylin*-02 transcripts being more expressed in the head, whereas *stylin*-03 was less expressed and *stylin*-05 highly variably but equally expressed in all four tissues (Figure 8A).

Expression profiles were also analyzed during *A. pisum* development at five different time-points: during the four larval stages and in adults. *rr-1 stylin* genes showed similar expression patterns with an increase in successive larval stages peaking at the fourth instar (8 to 20 times higher than the expression level in the first instar) and then decreasing in adults. Noticeably, expression levels were highly variable in the fourth instar, the developmental stage of longest duration in *A. pisum*. Expression levels of *stylin*-05 gene were comparable in the four larval stages and were significantly lower in adults (Figure 8B).
Stylin conserved among aphid species

We constructed two phylogenetic trees for RR-1 and CPAP3 protein sequences identified in the six genomes of aphids and in the phylloxera Daktulosphaira vitifoliae, available in AphidBase (Figure 9A,B). The phylloxera belongs to Phylloxeroidea, a superfamily considered to be the nearest sister taxon of the Aphidoidea, which probably diverged 250 My ago.52 The homologues of each Stylin grouped in a separated clade, except for homologues of Stylin-01 and Stylin-02 which were already shown to group in a single clade.26 Noticeably, only one protein of D. vitifoliae grouped with Stylin-01 and Stylin-02 homologues, probably reflecting gene duplication in aphids after their divergence with phylloxerids. Regarding Stylin-04 and Stylin-04bis found in A. pisum genome, phylogenetic analyses showed that only one of these two proteins was present in other aphid species and in D. vitifoliae. A gene duplication of stylin-04 gene likely occurred only in A. pisum, as previously described in other conserved gene families in this species.36

Stylin are conserved among aphid species. Remarkably, the domains exposed at the surface of the acrostyle are nearly identical in the six aphid species (Figure 9C). The most divergent sequences were D. vitifoliae homologues of Stylin-04/04bis and Stylin-05. In addition, as previously shown in A. pisum, and so suggesting a high conservation of their function, all five stylin were detected at the tip of the maxillary stylets of M. persicae, the most important vector of plant viruses (Figure 6A).

## DISCUSSION

The aphid stylets are composed of a biomaterial with unique surface properties ensuring binding, retention, and release of plant viruses during their journey from one host to another. This transport of viruses is driven at least by cuticular proteins emerging at the surface of the cuticle, in direct contact with endogenous and exogenous compounds flowing in and out the aphid stylets.
maxillary stylets. Although not characterized when we started our study, we could speculate that various processes might specifically functionalize the cuticular surface at the tip of maxillary stylets in the region described as the acrostyle. These speculated processes could be, for example, the local protein composition of the cuticle or the degree of protein sclerotization. We thus initially presumed that some cuticular proteins would be specific to the stylets, even to the tip of the maxillary stylets, and absent from other anatomical structures. Our comparative proteomics analysis was designed to identify these stylet-specific CPs if there were any, because they represent prime candidate receptors of plant viruses.

Common Set of CPs to Form Aphid Cuticles

We have here experimentally generated the first proteome of four cuticular structures (antennae, legs, wings, and stylets) of the pea aphid. These four proteomes contain multiple CPs of which a shared subset likely represents the building blocks of all aphid cuticles, as suggested for the common subset of CP genes identified in a comparative analysis of seven anatomical structures of the mosquito Anopheles gambiae.33,53 The pea aphid proteome profiles were globally similar in antennae, legs, and wings, with higher complexity/depth than the stylets proteome. However, a few CPs were specifically found in antennae, legs, or wings, and whether or how they contribute to specific cuticular properties/functions in these respective anatomic structures has not been investigated.

Fewer CPs Identified in Stylets

The repertoire of CPs identified in aphid cuticles was half in stylets compared to in other structures, reflecting either the absence of some CPs, their significantly reduced accumulation, or extractability. It is notable that the cuticle of antennae, legs, and wings originates from epithelial cells, while that of the stylets is synthesized by the highly specialized retort glands entirely dedicated to cuticle production.13,14,54 This sole ontology difference may lead to cuticles with different degrees of protein complexity. However, a recent characterization of transcripts in the retort glands revealed a more complex set of expressed CP genes than in other previously characterized cuticular transcriptomes.28 Our MS data show that over 50% of the stylet CPs were identified by a single peptide, sometimes covering less than 2% of the mature protein, suggesting that CPs are hardly extractable from this organ and perhaps explaining the low complexity of the corresponding proteome.

CPs Are Not Evenly Distributed in Aphid Stylets

Our labeling experiments showed a variable distribution of CPs within maxillary stylets. Several RR-2 peptides were displayed under the superficial layers of the cuticle all along maxillary

Figure 8. Stylin relative expression patterns in different A. pisum body parts and in different nymphal instars. (A) Stylin genes expression quantified by qRT-PCR in head, antenna, leg, and wing relative to stylin expression in Heads. (B) Stylin genes expression quantified by qRT-PCR in first (1st), second (2nd), third (3rd), and fourth (4th) nymphal stages and in adults (Ad) relative to stylin expression in first instars. actin and EF1α genes were used for data normalization. Results are reported as means ± SD for 3 independent biological replicates. Lowercase letters indicate significant differences between samples (TukeyHSD, p < 0.005).
Figure 9. Stylin and acrostyle surface peptides are conserved in aphid vectors. Phylogenetic relationships of (A) CPR_RR-1 proteins with trimmed signal peptides and (B) full CPAP3 proteins of *Acyrthosiphon pisum* (Ap), *Aphis glycines* (Ag), *Daktulosphaira vitifoliae* (Dv), *Durophus noxia* (Dn), *Myzus cerasi* (Mce), *Myzus persicae* (Mp), *Rhopalosiphum padi* (Rp). AphidBase accession numbers for each gene are indicated on the right of the species abbreviation. It is noteworthy that to retrieve *M. persicae* sequences from AphidBase, numbers should be preceded by MYZIP1364_0001_v1.0_. Branch support values are indicated at the node, and the scale bar represents probabilities of change from one amino acid to another in terms of a unit, which is an expected 1% change between two amino acid sequences. Shaded portions represent stylin clusters and are numbered from 01 to 05 according to corresponding stylin names. (C) Conservation of exposed domains of the fivestylins identified in *M. persicae* (Mp), *A. pisum* (Ap), *M. cerasi* (Mce), *R. padi* (Rp), *A. glycines* (Ag), *D. noxia* (Dn), and *Daktulosphaira vitifoliae* (Dv). Alignments were performed.
styles. Some of them were also detected under the surface of the acrostyle. With the antibodies used here and in previous studies, we could not evidence their direct accessibility at the surface of the cuticle. However, we cannot strictly rule out such a possibility. Actually, repeat-rich sequences containing alanine, proline, tyrosine, and serine residues commonly found in aphid RR-2 proteins and present in several MS-detected peptides are poorly immunogenic and could not be considered for peptide synthesis and antibody production. Therefore, their accessibility in stylet cuticle could not be assessed. By contrast, all RR-1 and CPAP3 have been exclusively detected at the distal extremity of maxillary stylets (Figure 6), with the strongest labeling being observed for RR-1, that likely decorates the entire surface of the acrostyle. Despite their apparent specific localization, we must consider that RR-1 and CPAP3 peptides may also be evenly distributed all over the stylets because these protein families do not seem to be overrepresented in the core stylet material when compared to other aphid cuticular anatomical structures (Figure 4, Table 3), and so are likely generally present in any cuticle. In this hypothesis, what would appear specific would be their accessibility at the surface of the acrostyle, and this could reveal a general property of the cuticle that would be locally functionalized by surface modifications allowing access to normally embedded CPs.

Intriguingly, no labeling was ever observed on mandibular stylets in our experimental conditions. This striking observation suggests that proteins in this structure may be tightly cross-linked and no longer detectable with our antibodies or, as hypothesized above, that these stylets have surface properties that do not give access to the embedded cuticular proteins. Both hypotheses assume a comparable protein constitution for maxillary and mandibular stylets, consistent with their undistinguishable transcript profiles reported earlier.

Receptor Candidates

As expected for structural components of arthropod cuticle, stylins are highly conserved in all aphid species where sequence data are available. Receptors of plant viruses may also be conserved in vectors because the same viral species can sometimes be transmitted by dozens of different aphid species. Our results indicate that the protein domains accessible at the surface of the acrostyle share high degrees of identity with their homologues in other aphid species (Figure 9C). Interestingly, the most divergent protein sequences were those of D. vitifoliae, the grapevine pest species, for which transmission of plant viruses or of other pathogens has not been reported so far.

RR proteins are largely represented at the surface of the acrostyle. Proteins from RR-1 subgroup are generally found in soft cuticle, but a few of them have also been described in hard cuticle. Their role in virus transmission has recently emerged in the literature, and they are associated with both circulative and noncirculative virus transmission. CPR1, a RR-1 protein of the small brown planthopper Laodelphax striatellus, was first shown to interact with pc3, a viral nucleocapsid protein of rice stripe virus (RSV) (62). RSV is a tenuivirus transmitted in a circulative propagative manner. By contrast with noncirculative viruses, these viruses are not retained on their vector mouthparts. They are internalized in insect body, transit through the hemolymph to different internal tissues, and reach the salivary glands from which they can be inoculated together with egestion of saliva into a new host plant. Knockdown of CPR1 transcripts resulted in a decrease in RSV transmission. The authors proposed that CPR1 could assist viral movement within the insect body, a totally unexpected function for cuticular protein. For noncirculative viruses, in vitro competition experiments and transmission phenotypes associated with stylin genes knockdown demonstrated the role of Stylin-01 in CaMV transmission by its aphid vector. Decreasing Stylin-01 transcripts in M. persicae resulted in a reduced CaMV transmission capacity. In an independent study, Stylin-01, named previously Mpcp4 in M. persicae, was shown to interact in yeast with the coat protein of Cucumber mosaic virus (CMV), another aphid-transmitted noncirculative virus. However, its role in CMV transmission still lacks direct evidence. In addition to Stylin-01, Stylin-02, -03, and -04/-04bis, distributed over the surface of the acrostyle, now stand as prime candidate receptors of plant viruses.

CPAP3 are Cuticular Proteins Analogous to Peritrophins, previously known as “gasp” or “obstructor” family. They have been described in all insect, and possess three type 2 chitin-binding domains (ChtBD2). They have been detected in different tissues. However, they were missing from the list of CPs identified in the proboscis of An. gambiae. They have never been reported associated with virus transmission so far and are described here for the first time in aphid stylets. They are involved in cuticle formation and in structural integrity of cuticles. Antibodies produced in our study target less than 7% of their amino acid sequences and further effort will be required to determine if some CPAP3 domains could be better exposed at the surface of aphid stylets. Determining if they play a role in virus transmission would also warrant further investigation. As they belong to a distinct structural class of chitin-binding proteins relative to CPR (harboring cysteine-bridged Chitin-Binding Domain vs cysteine-free Chitin-Binding Domains for CPR), their functional properties might be distinct from the canonical CPR proteins identified so far as stylins and as active players in plant virus transmission.

The RR-2 proteins represent the great majority of CPs identified in A. pisum styles. These proteins are generally found in hard and rigid cuticles, and some of them have been shown to be essential for cuticle integrity, integument structure, insect development and could be involved in cuticle formation. RR-2 distribution within aphid styles suggests that they likely constitute the main building blocks of this type of cuticle. Evidence of their accessibility at the surface of aphid stylets is still lacking. So, although not prime candidates, the role of RR-2 proteins in virus transmission cannot be discarded. They have been cited in several studies as interacting molecules of both noncirculative and circulative viruses. For noncirculative viruses, three RR-2 of M. persicae were shown to interact in vitro with HC-Pro, the ligand protein of a potyvirus, Zucchini yellow mosaic virus. One of these proteins is homologous to ACYPI006670 that contains four peptides detected in maxillary styles (Table 4). Another one has been characterized recently which reduction of transcripts correlated with a decrease in...
Potyvirus Y transmission. However, its putative ortholog in A. pisum was not in our MS data set, and its presence within M. persicae stylets, embedded in the chitin or at the surface of the cuticle, remains to be confirmed.

■ CONCLUSIONS

We provide through this study the first comparative proteomics analysis of four aphid cuticular anatomical structures, namely antennae, legs, wings, and stylets. Our data gives preliminary evidence that a great number of CP proteins are common to antennae, legs, and wings, while a few CP proteins seem specific to each appendage. The stylet, which was of main interest to better understand the vector/virus interaction was found to be distinct in composition compared to the other three studied appendages. We determined the repertoire of CPs of aphid stylets and precisely mapped their accessibility in maxillae at the surface of the acrostyle. Further characterization for this short list of proteins showed that they are highly conserved in aphid species and thus all represent good candidate receptors of plant viruses. These data contribute to a better characterization of aphid mouthparts, a crucial insect feeding appendage, but also point out at the surface specificities of the cuticle and at the distribution of cuticular protein accessibility, which may be relevant for local functionalization of this tissue. Beyond feeding appendage and virus receptor candidates, our proteomic data sets may contribute to future investigations of other important physiological functions in aphids such as chemoreception and sensory system.

■ ASSOCIATED CONTENT

* Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.9b00851.

Table S1. Database generated for protein identification including 252,212 entries. Table S2. Home-made A. pisum CP database identified using CutProtFamPred. Table S3. List of oligonucleotides used in this study. Table S4. A. pisum proteins identified in antennae. Table S5. A. pisum proteins identified in legs. Table S6. A. pisum proteins identified in wings. Table S7. A. pisum proteins identified in stylets. Table S8. Proteins distribution in A. pisum cuticular structures. Table S9. Analysis of GO terms enrichment for proteins identified in the Core Proteome of antennae, legs, wings, and stylets. Table S10. GO terms enrichment in antennae. Table S11. GO terms enrichment in legs. Table S12. GO terms enrichment in wings. Table S13. GO terms enrichment in stylets. Table S14. Cuticular proteins identified in antennae. Table S15. Cuticular proteins identified in legs. Table S16. Cuticular proteins identified in wings. Table S17. Cuticular proteins identified in stylets. Table S18. Cuticular protein distribution in A. pisum cuticular structures. (ZIP)

Accession Codes

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD016517.

■ AUTHOR INFORMATION

Corresponding Authors

Philippe Bulet — Plateforme BioPark d’Archamps, 74160 Archamps, France; CR University of Grenoble-Alpes, Institute for Advances Biosciences, Inserm U1209, CNRS UMR S309, 38058 Grenoble, France; Phone: +(33) 450 432 521; Email: philippe.bulet@univ-grenoble-alpes.fr

Marilyne U zest — BGPI, University of Montpellier, INRAE, 34000 Montpellier, France; orcid.org/0000-0002-7459-8434; Phone: +(33) 499 624 857; Email: marilyne.uzest@inra.fr

Authors

Macle Deshoux — BGPI, University of Montpellier, INRAE, 34000 Montpellier, France

Victor Masson — Plateforme BioPark d’Archamps, 74160 Archamps, France; CR University of Grenoble-Alpes, Institute for Advances Biosciences, Inserm U1209, CNRS UMR S309, 38058 Grenoble, France

Karim Arrah — Plateforme BioPark d’Archamps, 74160 Archamps, France

Sébastien Voisin — Plateforme BioPark d’Archamps, 74160 Archamps, France

Natalia Guschinskaya — INRAE, INS Lyon, UMR5240 MAP CNRS-UCBL, 69622 Villeurbanne, France

Manuella van Munster — BGPI, University of Montpellier, INRAE, 34000 Montpellier, France

Bastien Carel — BGPI, University of Montpellier, INRAE, 34000 Montpellier, France

Craig G. Webster — BGPI, University of Montpellier, INRAE, 34000 Montpellier, France

Yvan Rahbe — BGPI, University of Montpellier, INRAE, 34000 Montpellier, France; INSA Lyon, UMR240 MAP CNRS-UCBL, 69622 Villeurbanne, France; University of Lyon, 69007 Lyon, France

Stéphane Blanc — BGPI, University of Montpellier, INRAE, 34000 Montpellier, France

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.9b00851

Author Contributions

M.D. and V. M. contributed equally to this study. P. B. and M. U. jointly supervised this work. All authors contributed to the work and the manuscript presented here. All authors have approved the final version of the manuscript.

Funding

This work was supported by grants from the Bill and Melinda Gates Foundation (GCeg, OPP1130147) and the French National Research Agency (ANR15-CE20-0011). Funding for the sequencing of Myzus persicae and Rhopalosiphum padi was provided by ERC Starting Grant APHIDHOST-310190 awarded to Jorunn Bos at the James Hutton Institute, United Kingdom. Funding for Daktulosphaira vitifoliae clone PcC genomic sequencing was provided by Inra (AIP Bioresources) and BGI Biotech in the frame of i5k initiative. Parts of the transcriptomic resources were obtained within the 1KITE projects (Bernhard Misof, Bonn, Germany).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Association Plateforme BioPark d’Archamps (France) for its technical facilities through its Research & Development program and Mr. Hervé Breton from the society Chromoptic for providing us with the glass microtubes.
ABBREVIATIONS

AA, amino acid(s); ACN, acetonitrile; CaMV, cauliflower mosaic virus; ChtBD2, type 2 chitin-binding domains; CMV, cucumber mosaic virus; CP, cuticular proteins; CPR, cuticular proteins with the Rebers and Riddiford consensus sequence; CPAP3, cuticular protein analogous to peritrophins 3; FDR, false discovery rate; GO, gene ontology; h, hour; HFIP, hexafluoroisopropanol; HMMs, Hidden Markov Models; HPLC, high-performance liquid chromatography; ID, identifier; ITO, Indium–Tin Oxide; LC-MS/MS, Liquid chromatography coupled to tandem mass spectrometry; min, minutes; MS, mass spectrometry; RR, Rebers and Riddiford consensus sequence; RSV, rice stripe virus; TFA, trifluoroacetic acid; v/v, volume/volume; U, Unit

REFERENCES

1. Dixon, A. F. G. K.; Leps, P. J.; Holman, J. Why there are so few species of aphids, especially in the tropics. Ann. Nat. 1987, 129 (4), 580–592.
2. Favret, C. 2017, Aphid species file. http://aphid.speciesfile.org.
3. Ogawa, K.; Miura, T. Aphid polyphenisms: trans-generational developmental regulation through viviparity. Front. Physiol. 2014, 5, 1.
4. Hogenhout, S. A.; Ammar; el, D.; Whitfield, A. E.; Redinbaugh, M. G. Insect vector interactions with persistently transmitted viruses. Annu. Rev. Phytopathol. 2008, 46, 327–359.
5. Katis, N. I.; Tsitsipis, J. A.; Stevens, M.; Powell, G. Transmission of Plant Viruses. In Aphids as Crops pests of; van Emden, H. F., Harrington, R. E.; CABl, Wallingford, United Kingdom, 2007; pp 353–390.
6. Nault, L. R. Arthropod transmission of plant viruses: a new synthesis. Ann. Entomol. Soc. Am. 1997, 90, 521–541.
7. Whitfield, A. E.; Falk, B. W.; Rotenberg, D. Insect vector-mediated transmission of plant viruses. Virology 2015, 479–480, 278–289.
8. Martin, B.; Collar, J. L.; Tjallingi, W. F.; Fereres, A. Intracellular ingestion and salivation by aphids may cause the acquisition and inoculation of non-persistently transmitted plant viruses. J. Gen. Virol. 1997, 78 (10), 2701–2705.
9. Perry, K. L.; Zhang, L.; Palukaitis, P. Amino acid changes in the coat protein of cucumber mosaic virus differentially affect transmission by the aphids Myzus persicae and Aphis gossypii. Virology 1998, 242 (1), 204–210.
10. Moreno, A.; Hebrard, E.; Uzest, M.; Blanc, S.; Fereres, A. A single amino acid position in the helper component of cauliflower mosaic virus is determined solely by the viral coat protein. Virology 1998, 257 (1), 1387–1393.
11. Chen, B.; Franci, R. I. B. Cucumovirus transmission by the aphid Myzus persicae is determined solely by the viral coat protein. J. Gen. Virol. 1990, 71, 939–944.
12. Pirone, T. P.; Blanc, S. Helper-dependent vector transmission of plant viruses. Annu. Rev. Phytopathol. 1996, 34, 227–247.
13. Davidson, J. The structure and biology of Schizoneura lanigera Hausmann or Woolly Aphis of the Apple tree. I. The Apterous viviparous female. Q. Jl. microsc. Sci. 1913, 58, 653–702.
14. Ponsen, M. B. The site of potato leafroll virus multiplication in its adult Anopheles gambiae. Insect Biochem. Mol. Biol. 2014, 52, 51–59.
15. He, N.; Botelho, J. M. C.; McNall, R. J.; Belozerov, V.; Dunn, W. A.; Mize, T.; Orlando, R.; Willis, J. H. Proteomic analysis of cast cuticles from Anopheles gambiae by tandem mass spectrometry. Insect Biochem. Mol. Biol. 2007, 37 (2), 135–146.
16. Cornel, R. S.; Togawa, T.; Dunn, W. A.; He, N.; Emmons, A. C.; Willis, J. H. Annotation and analysis of a large cuticular protein family with the R&R Consensus in Anopheles gambiae. BMC Genomics 2008, 9, 22.
17. Cornel, R. S.; Willis, J. H. Extensive gene amplification and concerted evolution within the CPR family of cuticular proteins in mosquitoes. Insect Biochem. Mol. Biol. 2008, 38 (6), 661–676.
18. Zhou, Y.; Badgett, M. J.; Bowen, J. H.; Vannini, L.; Orlando, R.; Willis, J. H. Distribution of cuticular proteins in different structures of adult Anopheles gambiae. Insect Biochem. Mol. Biol. 2016, 75, 45–57.
19. Conesa, A.; Götz, S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int. J. Plant Genomics 2008, 2008, 169832.
20. Livak, K. J.; Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2–ΔΔCT Method. Methods 2001, 25 (4), 402–408.
21. Mathers, T. C.; Chen, Y.; Raithakottig, G.; Legeai, F.; Mugford, S. T.; Bae-Puyoulet, P.; Bretaudeau, A.; Clavijo, B.; Colella, S.; Collin, O.; et al. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. Genome Biol. 2017, 18 (1), 27.
22. Nicholson, S. J.; Nickerson, M. L.; Dean, M.; Song, Y.; Hoyt, P. R.; Rhee, H.; Kim, C.; Puterka, G. J. The genome of Diuraphis noxia, a global aphid pest of small grains. BMC Genomics 2015, 16 (1), 429.
(38) Gallot, A.; Rispe, C.; Leterme, M.; Gauthier, J. P.; Jaubert-Paussamaï, S.; Tagu, D. Cytolic proteins and seasonal photoperiodism in aphids. *Insect Biochem. Mol. Biol.* 2010, 40, 235–240.

(39) Wenger, J. A.; Cassone, B. J.; Legai, F.; Johnston, J. S.; Bansal, R.; Yates, A. D.; Coates, B. S.; Pavanato, V. A. C.; Michel, A. Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect Biochem. Mol. Biol.* 2017, 18, (1), 27.

(40) Almagro Armenteros, J. J.; Tsigiros, K. D.; Sonderby, C. K.; Petersen, T. N.; Winther, O.; Brunak, S.; von Hejne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 2019, 37 (4), 420–423.

(41) Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004, 32 (5), 1792–1797.

(42) Chevenet, F.; Brun, C.; Banuls, A. L.; Jacqu, B.; Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinf.* 2006, 7, 439.

(43) Notredame, C.; Higgins, D. G.; Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment11Edited by J. Thornton. *J. Mol. Biol.* 2000, 302 (1), 205–217.

(44) Di Tommaso, P.; Moratti, S.; Xenarios, I.; Orobitg, M.; Montanyola, A.; Chang, J. M.; Taly, J. F.; Notredame, C. T-coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 2011, 39 (suppl. 2), W13–W17.

(45) Carolan, J. C.; Fitzroy, C. T. J.; Ashton, P. D.; Douglas, A. E.; Wilkinson, T. L. The secreted salivary proteome of the pea aphid *Acyrthosiphon pisum* characterised by mass spectrometry. *Proteomics* 2009, 9 (9), 2457–2467.

(46) Carolan, J. C.; Caragea, D.; Reardon, K. T.; Muthukrishnan, S.; Kramer, K. J.; Arakane, Y. Tribolium castaneum RR1-cytolic protein TcCPR4 is required for formation of pore canals in rigid cuticle. *PLoS Genet.* 2015, 11 (2), e1004963.

(47) Vandermosten, S.; Harmel, N.; Mazzucchelli, G.; De Paauw, E.; Haubruege, E.; Francis, F. Comparative analyses of salivary proteins from three aphid species. *Insect Biochem. Mol. Biol.* 2014, 23 (1), 67–77.

(48) Boulain, H.; Legeai, F.; Guy, E.; Morlière, S.; Douglas, N. E.; Oh, J.; Murugan, M.; Smith, M.; Jaquiey, J.; Peccoud, J.; White, F. F.; Carolan, J. C.; Simon, J. C.; Sugio, A. Fast Evolution and Lineage-Specific Gene Family Expansions of Aphid Salivary Effectors Driven by Interactions with Host-Plants. *Genome Biol. Evol.* 2018, 10 (6), 1547–1572.

(49) Madhusudhan, V. V.; Miles, P. W. Mobility of salivary components as a possible reason for differences in response of alfalfa to the spotted alfalfa aphid and pea aphid. *Entomol. Exp. Appl.* 1998, 86, 25–39.

(50) Nardelli, A.; Vecchi, M.; Mandrioli, M.; Manicardi, G. C. The Evolutionary History and Functional Divergence of Trehalase (treh) Genes in Insects. *Front. Physiol.* 2019, 10, 62.

(51) Fevay, G.; Delobel, B.; Rahbé, Y. Influence of the amino acid balance on the improvement of an artificial diet for a biotype of *Acyrthosiphon pisum* (Homoptera: Aphididae). *Can. J. Zool.* 1988, 66, 2449–2453.

(52) Heie, O. E. Paleontology and phylogeny. In *Aphids: their biology, natural enemies, and control*; Minks, A. A. K., Harrewijn, P., Eds.; Elsevier: Amsterdam, The Netherlands, 1987; Vol. 2a, pp 367–391.

(53) Zhou, Y.; Badgett, M. J.; Orlando, R.; Willis, J. H. Proteomics reveals localization of cytolic proteins in *Anopheles gambiae* genome. *Insect Biochem. Mol. Biol.* 2019, 104, 91–105.

(54) Moussian, B. Molecular model of skeletal organisation and differentiation. In *Extracellular composite matrices in arthropods*; Cohen, E., Moussian, B.; Eds.; Springer: Switzerland, 2016; pp 67–87.

(55) Blackman, L. M.; Eastop, V. F. Taxonomic Issues. In *Aphids as Crop Pests*; van Emden, H. F., Harrington, R., Eds.; CABI: Wallingford, United Kingdom, 2007; pp 1–29.

(56) U zest, M.; Blanc, S. Non circulative virus-vector interactions. In *Microbe-Arthropode vector interactions*; Brown, J., Ed.; USA: Academic Press: San Diego, 2016; pp 59–72.
(75) Noh, M. Y.; Kramer, K. J.; Muthukrishnan, S.; Kanost, M. R.; Beeman, R. W.; Arakane, Y. Two major cuticular proteins are required for assembly of horizontal laminae and vertical pore canals in rigid cuticle of Tribolium castaneum. Insect Biochem. Mol. Biol. 2014, 53, 22–29.

(76) Qiao, L.; Xiong, G.; Wang, R. X.; He, S. Z.; Chen, J.; Tong, X. L.; Hu, H.; Li, C. L.; Gai, T. T.; Xin, Y. Q.; Liu, X. F.; Chen, B.; Xiang, Z. H.; Lu, C.; Dai, F. Y. Mutation of a cuticular protein, BmorCPR2, alters larval body shape and adaptability in silkworm, Bombyx mori. Genetics 2014, 196 (4), 1103–1115.

(77) Pan, P. L.; Ye, Y. X.; Lou, Y. H.; Lu, J. B.; Cheng, C.; Shen, Y.; Moussian, B.; Zhang, C. X. A comprehensive omics analysis and functional survey of cuticular proteins in the brown plant hopper. Proc. Natl. Acad. Sci. U. S. A. 2018, 115 (20), 5175–5180.

(78) Deshoux, M.; Monsion, B.; Uzest, M. Insect cuticular proteins and their role in transmission of phytophages. Curr. Opin. Virol. 2018, 33, 137–143.

(79) Bahrami-Kamangar, S.; Christiaens, O.; Taning, C.; De Jonghe, K.; Smagghe, G. The cuticle protein MPCP2 is involved in Potato virus Y transmission in the green peach aphid Myzus persicae. J. Plant Dis. Prot. 2019, 126, 1–7.

(80) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 2019, 47 (D1), D442–D450.