HAL open science

# Insect mouthpart transcriptome unveils extension of cuticular protein repertoire and complex organization

Natalia Guschinskaya, Denis Ressnikoff, Karim Arafah, Sebastien Voisin, Philippe Bulet, Marilyne Uzest, Yvan Rahbé

▶ **To cite this version:**

Natalia Guschinskaya, Denis Ressnikoff, Karim Arafah, Sebastien Voisin, Philippe Bulet, et al.. Insect mouthpart transcriptome unveils extension of cuticular protein repertoire and complex organization. iScience, 2020, 23 (2), pp.100828. 10.1016/j.isci.2020.100828 . hal-02427505v2

**HAL Id: hal-02427505**

**https://hal.science/hal-02427505v2**

Submitted on 13 Nov 2021

**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.

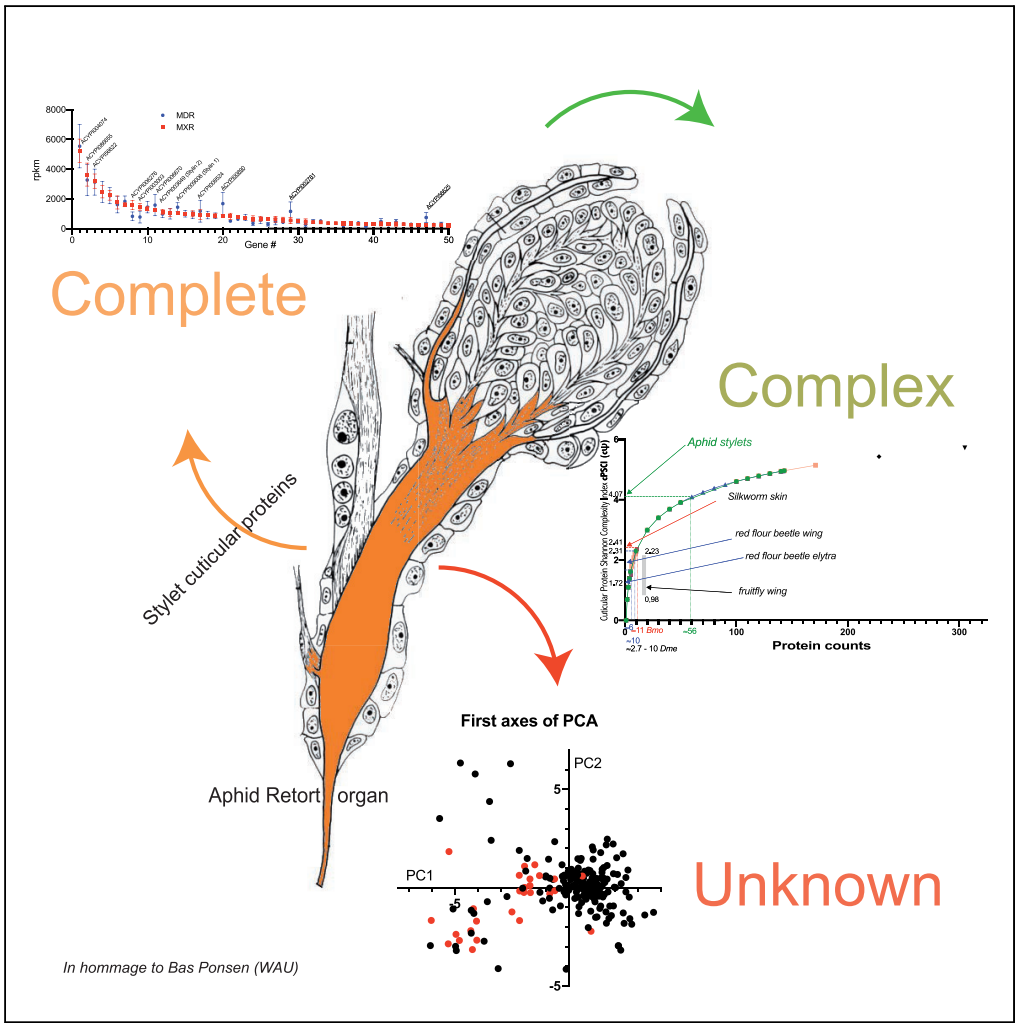
Distributed under a Creative Commons Attribution 4.0 International License

iScience

CellPress

Article

# Insect Mouthpart Transcriptome Unveils Extension of Cuticular Protein Repertoire and Complex Organization




Natalia Guschinskaya, Denis Ressnikoff, Karim Arafah, Sébastien Voisin, Philippe Bulet, Marilyne Uzest, Yvan Rahbé

yvan.rahbe@inra.fr

## HIGHLIGHTS

First transcriptome of aphid retort glands and stylet cuticular protein composition

A pyrokinin transcript is mandibular gland specific at the onset of adult moult

Stylet cuticle is of higher protein complexity than other insect cuticles

A new class of low-complexity cuticular proteins is predicted

Guschinskaya et al., iScience *23*, 100828
February 21, 2020 © 2020 The Author(s).
https://doi.org/10.1016/j.isci.2020.100828

iScience

CellPress

Article

# Insect Mouthpart Transcriptome Unveils Extension of Cuticular Protein Repertoire and Complex Organization

Natalia Guschinskaya,[1,6] Denis Ressnikoff,[2,6] Karim Arafah,[3] Sébastien Voisin,[3] Philippe Bulet,[3,4] Marilyne Uzest,[5] and Yvan Rahbé[1,5,6,7,8,*]

## SUMMARY

**Insects have developed intriguing cuticles with very specific structures and functions, including microstructures governing their interactions with transmitted microbes, such as in aphid mouthparts harboring virus receptors within such microstructures. Here, we provide the first transcriptome analysis of an insect mouthpart cuticle ("retort organs" [ROs], the stylets' precursors). This analysis defined stylets as a complex composite material. The retort transcriptome also allowed us to propose an algorithmic definition of a new cuticular protein (CP) family with low complexity and biased amino acid composition. Finally, we identified a differentially expressed gene encoding a pyrokinin (PK) neuropeptide precursor and characterizing the mandibular glands. Injection of three predicted synthetic peptides PK1/2/3 into aphids prior to ecdysis caused a molt-specific phenotype with altered head formation. Our study provides the most complete description to date of the potential protein composition of aphid stylets, which should improve the understanding of the transmission of stylet-borne viruses.**

## INTRODUCTION

The insect cuticle is a solid biomaterial composed of four structural/biochemical compartments: chitin, proteins, lipids, and an aromatic-based cross-linking matrix (Andersen, 1979). With its fundamental participation in the lightweight exoskeleton of arthropods, it is one of the main structural constituents that allowed insects to colonize early terrestrial environments (Garrouste et al., 2012) and develop flight (Lease and Wolf, 2010). Chitin is a linear polysaccharide, second only to cellulose in its participation in terrestrial biomass and is specific to a majority of fungal species, a limited number of animal lineages, and some protists (Goncalves et al., 2016). In contrast to cellulose in plant cell walls, chitin in cephalopod plumes and arthropod cuticles is heavily loaded with proteins to develop the full mechanical, structural, and functional properties of these structures (Andersen, 1979). Despite this seminal role, due to technical limitations, the complete characterization of the protein composition of insect cuticles is still partial (Willis, 2010). The technical limitations include mass spectrometry blindness at highly polymerized moieties, resulting from either extractability issues or the difficulty of identifying intrinsic post-translational modifications of covalent sclerotization adducts (Andersen, 2010; Willis, 2018; Zhou et al., 2016).

Among cuticular structures, the most studied are the bulky surface covers layered over the ectodermal epithelia of insects, but some represent more specialized regions, mostly associated with the segmental organization of insects (metameric appendages, legs, wings). The biochemistry of the most specialized cuticular structures, those of head appendages such as the antennae and mouthparts, has scarcely been studied (Zhao et al., 2018; Awuoche et al., 2017; Zhou et al., 2014; Oliveira et al., 2017). A single recent publication involving mouthpart transcriptomics did not address the time and organs strictly responsible for cuticular biogenesis *stricto sensu* (tsetse fly proboscis organ) and was not focused on biomaterial characterization (Awuoche et al., 2017). As part of the feeding specialization process, mouthparts are crucial players with sensory and morphological structures that shape the front line of insect/host coevolutionary processes (Futuyma and Agrawal, 2009; Nel et al., 2018). In Hemiptera, a primarily plant-feeding order, the evolution of one of the 6-9 piercing-sucking type mouthparts of insects (Garrouste et al., 2012; Nel et al., 2018; Huang et al., 2016) has profoundly shaped the ecology of almost the entire order toward a dominant parasitic/predatory lifestyle (Weirauch and Schuh, 2011). In aphids, piercing-sucking mouthparts are composed of (1) a short and triangular labrum, which covers the base of the stylet bundle, (2) the labium,

[1]Insa de Lyon, UMR5240 MAP CNRS-UCBL, 69622 Villeurbanne, France

[2]CIQLE, Centre d'imagerie Quantitative Lyon-Est, UCB Lyon 1, Lyon, France

[3]Platform BioPark Archamps, Archamps, France

[4]CR University of Grenoble Alpes, Institute for Advanced Biosciences, Inserm U1209, CNRS UMR 5309, La Tronche, France

[5]BGPI, Univ Montpellier, INRA, CIRAD, Montpellier SupAgro, Montpellier, France

[6]Université de Lyon

[7]Lead Contact

[8]Twitter: @yrahbe

*Correspondence: yvan.rahbe@inra.fr

https://doi.org/10.1016/j.isci.2020.100828


This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

which is a segmented and tubular organ with complex musculature that contracts and shortens during insertion of the stylet into plant tissue, and (3) the stylet bundle, which is inserted in a groove dug along the length of the anterior surface of the labium (Forbes, 1966). The basic morphology of the stylet bundle dates back to more than 300 My ago (Misof et al., 2014). It comprises two external mandibular (*mdr*) stylets that surround and protect two inner and deeply asymmetrical maxillary (*mxr*) stylets, all transformed into four long needle-like cuticle structures. Only the mandibular stylets are innervated by two dendrites that lie in a small internal duct; their distal extremity is curved inward and sharpens at the tip, and the outer surface harbors a series of barb-like ridges that participate in the insertion of the bundle into the plant (Forbes, 1966). The maxillary stylets are also sharply pointed at the distal extremity but display a much more complex architecture; the whole length of the inner face of both maxillary stylets interlocks to enable the formation of two ducts, a large food canal and a small salivary canal carved in the left maxillary stylet. The food and salivary canals are fused into a single common duct at the tip (Forbes, 1966).

One of the most striking features of the order Hemiptera, directly linked to their needle-mouthparts, is their ability to transmit to their hosts an impressive array of viruses and bacteria. From the initial trivial view of a "*dirty syringe*," our progress in understanding plant disease transmission mechanisms has led to more sophisticated models. For example, in aphids, receptors of the *Cauliflower mosaic virus* (CaMV), a noncirculative stylet-borne virus, were recently characterized. They were first demonstrated to be present and accessible solely at the internal surface of the maxillary stylets (Uzest et al., 2007; Webster et al., 2018), and virus binding sites were associated with very specific cuticular regions at the tip of the stylet's common canal (Uzest et al., 2007), the acrostyle (Uzest et al., 2010). Moreover, the molecular partners of CaMV at the cuticular surface were demonstrated to be proteins (Uzest et al., 2007). More recently, two cuticular proteins (CPs) were identified at the surface of the acrostyle (Webster et al., 2017), among which Stylin-01 was confirmed to be involved in CaMV transmission (Webster et al., 2018). These two proteins were the first to be identified in arthropod mouthparts and are both prime candidate receptors for other noncirculative viruses. However, the acrostyle was also shown to have a more complex proteomic composition, which has been only recently characterized by a proteomic approach (Webster et al., 2018). In this context, the full transcriptomic characterization of cuticular polymeric materials is a complementary approach to proteomic studies in cases where biogenetic tissues are available (Awuoche et al., 2017).

In our quest for a full identification of nonpersistent virus receptors, as well as a first complete definition of the protein composition of a cuticle's polymeric matrix, we undertook an RNA-Seq analysis of the cuticular glands secreting the four aphid stylets at each molt, a set of glands hitherto known as the retort organ (RO), or stylet glands, of macrosiphine aphids (Ponsen, 1972; Davidson, 1913). This organ was not studied in aphids since early in the previous century (Pinet, 1968; Heriot, 1934; Davidson, 1913; Ponsen, 1972) and characterized morphologically in elegant works on the potato psyllid, a crippling disease vector (Cicero, 2016). We present an updated description of this organ in the Supplemental Information. The goals of our present work were as follows: (1) determination of the technical and temporal features of stylet biogenesis in the preimaginal stage of our model aphid *Acyrthosiphon pisum*, (2) differential characterization of the presumptive protein composition of both *mxr* and *mdr* stylets, and (3) establishment of the putative quantitative formulation of an insect cuticular material, the aphid stylet(s), in terms of the structural protein composition. Surprisingly, we found very few differences in the structural protein composition of the *mxr* and *mdr* stylets and identified a single gene that discriminates between the functioning of the stylets at molting time. The retort transcriptomes of the aphid stylets revealed a complex set of expressed genes, although 20 genes account for 65% of the total expression and three genes account for 22%. This information paves the way for experiments on cuticle-biomimetic polymer assemblages that mix chitin and selected CPs through a rational analytical process (Vaclaw et al., 2018; Faivre et al., 2018).

## RESULTS

### Aphid Genes Coding for Cuticular Proteins Have Expression Peaks at L4-48 h

To determine the time point at which the retort transcriptome should be carried out, a synchronization procedure was designed, resulting in a sampling scheme of up to 16 times within the preimaginal stage (see Figure S1). Sampling of whole insects from multiple biological batches or aphid heads (restricting the analyzed RNA pools "around" the RO tissue) allowed us to refine the time-slot to the L4-46/48 stage or L4_e (Figures 1 and S2). This transect clearly shows that all target genes peak at this time point, including those encoding (1) the two CPs already identified in stylets (ACYPI009006 and ACYPI003649, termed *stylin-01* and *-02* (Webster et al., 2018)), (2) other CPs highly expressed in *A.pisum* head ESTs (ACYPI006250,

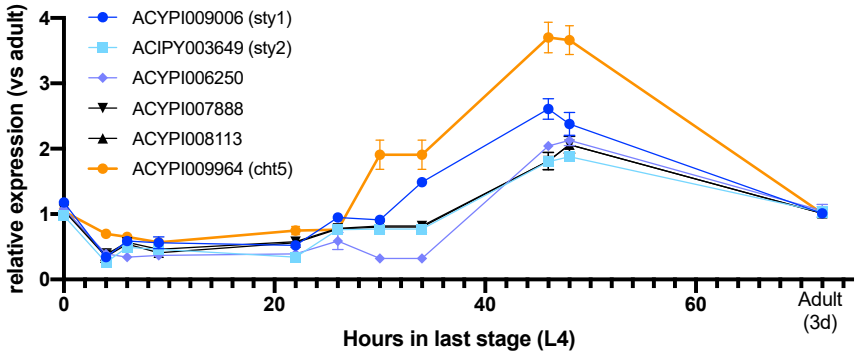



**Figure 1. mRNA Expression Profile of Six Genes of Interest during Aphid Development**

5 genes coding for cuticular proteins, among which two already identified in stylets (ACYPI009006, Stylin-01, ACYPI003649, Stylin-02); and three highly expressed in head ESTs (ACYPI006250, ACYPI007888, ACYPI008113) were used in qRT-PCR analysis together with the gene ACYPI009964, coding for an imaginal moult-associated chitinase Cht5 (J9KAI2). Gene-expression levels at different life stages are expressed relative to adult aphid. The EF1$\alpha$ and actin genes were used for data normalization. Results are reported as means $\pm$ SD; n = 3 independent biological replicates per stage. Data were analyzed by one-way ANOVA followed by a post hoc multiple comparisons test (Tukey's HSD test). (A) Expression level in aphid's heads was analyzed using three pooled heads from three independent biological replicates per stage.

ACYPI007888 and ACYPI008113), and (3) the aphid ecdysial chitinase ACYPI009964 (*cht5*) (Nakabachi et al., 2010). To obtain access to the largest array of genes encoding potential stylet CPs, we decided to sample the retort glands by head dissection and next-generation sequencing (NGS) analysis at point L4-46/48, enabling reliable distinction between *mxr* and *mdr* RO glands, as controlled by SEM (see later and Supplemental Information).

## Retort Organ as a Unique Preimaginal Cuticular Gland

Our initial attempt to sample RNA from aphid ROs was performed through laser microdissection, according to a protocol developed for the recovery of neurons from *Drosophila* brains (Iyer and Cox, 2010). This approach proved unfeasible due to the lack of anatomic integrity of these organs upon cryo-cutting (Figure S3). However, dissection of aphid heads from the rear, allowing us to safely avoid salivary gland sampling, gave us access to the target tissue (Figure S3). As aphid ROs have never been studied since the seminal work by Ponsen on *Myzus persicae* (Ponsen, 1972), we initiated a histological investigation of the anatomy of these glands by both serial sections and laser confocal scanning microscopy (LCSM) reconstructions (Figure 2 and Video S1). Full serial stacks and LCSM z-stacks are published as Data S1 (data.inra.fr). This tedious experimental procedure allowed us to secure RO samples free of (1) any confusion between the maxillary and mandibular glands, which were sampled separately (Figure S4) and checked by scanning electron microscopy (SEM) for *mxr*/*mdr* stylet recognition and post hoc by target transcript checks, and (2) any salivary gland contamination (Figure S3). The general quality of the six NGS libraries was good (three replicas of *mxr* ROs and three *mdr* ROs libraries; see Transparent Methods) and they displayed up to 11% of CP transcripts on rpkm basis (143 genes identified *a priori* as coding for CPs, from Tables S1 and S3) (Ioannidou et al., 2014). The depth of the libraries ensured the first full description of the potential composition of the CPs of any insect cuticular material (reads from *all* 143 genes were identified; a few reads were indeed present but did not pass the statistical thresholds of the differential analyses; see further).

## An Anti-stylin Antibody Extensively Labels the ROs

To confirm that ROs secrete stylet CPs at our sampling stage, both glands, but not salivary gland controls, were immunolabeled with an antibody specific to a CP previously identified in aphid stylets (Stylin-02, ACYPI003649) (Webster et al., 2017, 2018) and observed with CLSM (Figure 2C). Stylin-02 readily extends to the whole gland content but is distributed heterogeneously within different regions, including the extensively labeled secretory canals that secrete the presumptive (future) stylets. The same distribution was observed in *mxr* and *mdr* ROs (Figure S5 and Videos S2 and S3). More antibodies were used but only the Stylin-02 antibody is displayed in this work; a minimum of five glands were observed per antibody (Figure S5).

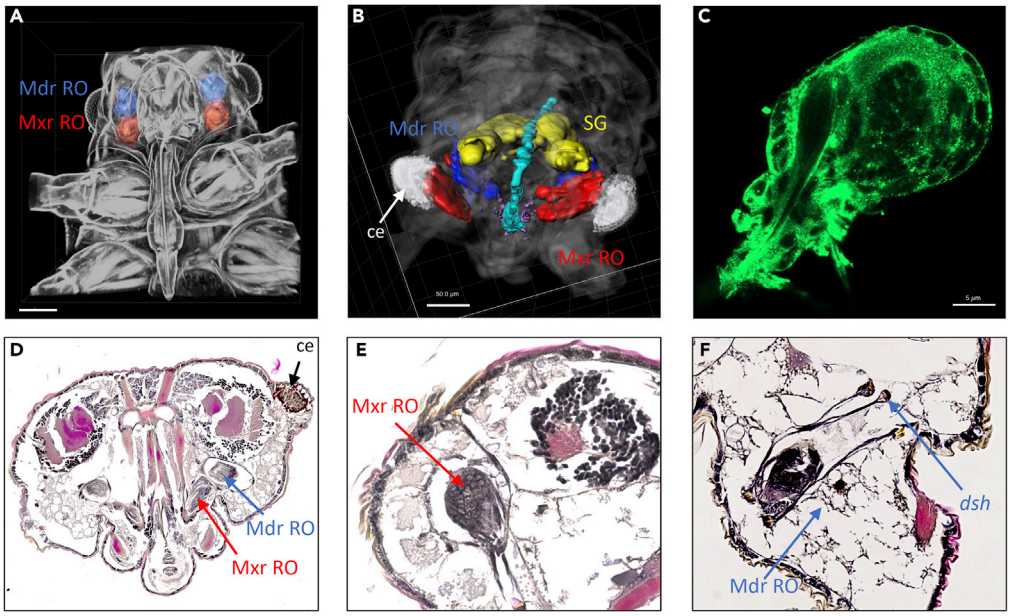



**Figure 2. Retort Organs of the Aphid *Acyrthosiphon pisum***

(A) 3D reconstruction of *A. pisum* head based on *in situ* CLSM imaging after cuticle clearing. Maxillary (*mxr*) and mandibular (*mdr*) retort organs are shown in red and blue, respectively.

(B) 3D reconstruction of retort organs based on formalin-fixed paraffin-embedded serial sections stained with hematoxylin/eosin (HE). The *mxr* (red) and *mdr* (blue) retort organs, salivary glands (yellow), digestive tract (light blue), and stylet basis (purple) have been artificially colored.

(C) *In vitro* immunohistological analysis of *mxr* retort organs with antibody specifically targeting Stylin-02 (ACYPI003649) (Webster et al., 2018).

(D–F) Formalin-fixed paraffin-embedded serial sections stained with HE, giving details of the positioning and anatomy of (D) both the *mdr* and *mxr* retort organs; (E) the *mxr* retort organ and (F) the *mdr* retort organ, showing characteristic association with dumbbell-shaped cells. Bars correspond to 5 (C) or 50 μm (A and B). ***mxr*** RO, maxillary retort organs; ***mdr*** RO, mandibular retort organs; ***CLSM***, confocal laser scanning microscopy; ***dsh***, dumbbell-shaped cells, ***ce***, compound eye.

## Maxillary and Mandibular ROs Transcriptomes Share Similar CP Transcript Profiles

Our experiments were designed to compare the transcriptomes of *mdr* and *mxr* ROs and therefore highlight differences in the presumptive CP composition of the two stylet types (maxillary and mandibular), among which candidate receptor proteins were supposed to be expressed exclusively in the maxillary stylets (Uzest et al., 2007; Webster et al., 2018). Surprisingly, the mapping of RO transcripts to the 143 known *A. pisum* CPs (Ioannidou et al., 2014) (see Transparent Methods, and Table S1) for a differential expression analysis showed no significant difference between the two glands. The ratio of CP expression fold-changes between *mxr* and *mdr* ROs ranged from 1 to 5 (ACYPI008570, see Figure 3A), and none were statistically significant (FDR = 0.99 in *edgeR* analysis after Benjamini-Hochberg correction, as was also the case with *DE-Seq2* adjusted p values; Table S2). The noncorrected p value was significant for only two genes overexpressed in the *mdr* RO, ACYPI009491 (FC = 3.6; p value 0.0072) and ACYPI007152 (FC = 2.9; p value 0.0196), both of which have low expression (Figures 3A, not 3C; Table S2). Notwithstanding any differences that might occur in the timing of expression between the two glands, we unexpectedly conclude that the global stylet composition should be common to the two aphid stylets, despite their very different shapes (Ponsen, 1972) and functions (Angelini and Kaufman, 2004; Forbes, 1966).

## Deep Sequencing of a Premolting Gland Suggests Presumptive Stylet CP Complexity

As no CP gene expression differentiates *mxr* from *mdr* ROs, we analyzed the spectrum of the protein transcripts present in both ROs (Figure 3). A total of 129 CP genes were expressed above the threshold in both the *mxr* and *mdr* organs (Table S2). Three major genes are top expressors, accounting for ≈22% of the total CPs (ACYPI004074, 10%; ACYPI086655, 6%; and ACYPI56622, 6%, members of the CPR-RR2 CP family, Figures 3B and 3C). Another 20 genes accounted for a cumulative total of ≈65% (12 CPR-RR2, 48%; 7 CPR-RR1,

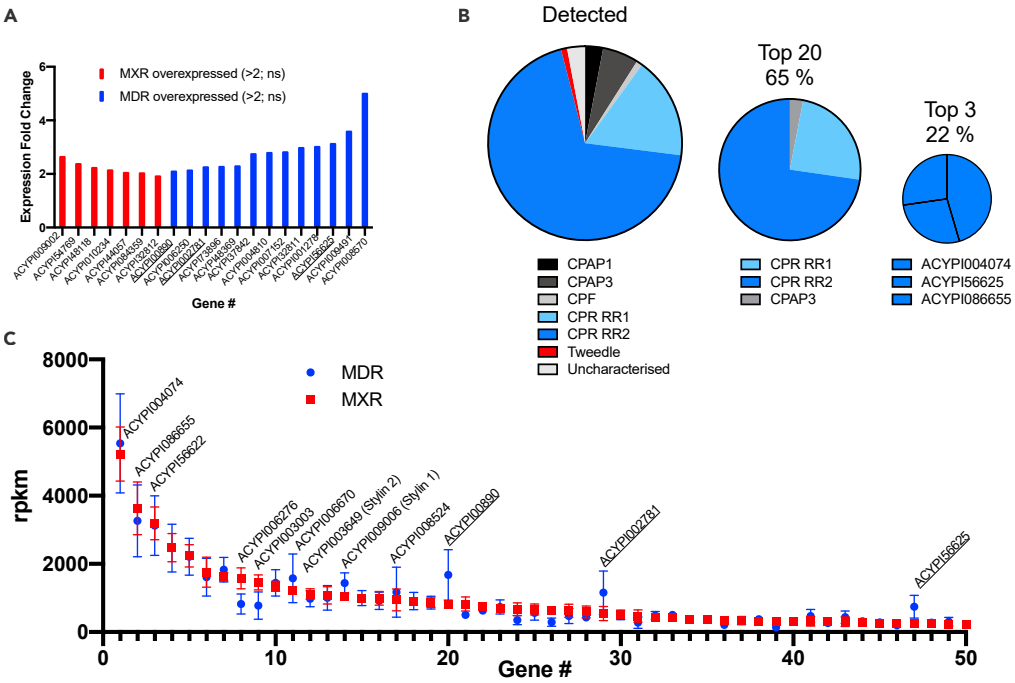



**Figure 3. Cuticular Protein Expression in Mandibular and Maxillary Retort Organs**

(A) Differential expression analysis of cuticular protein transcripts (CPs) between *mxr* and *mdr* glands (fold-change, FC > 2); underlined genes are the highest expressors, also highlighted in 2C.

(B) Distribution of CPs within CP families in the pea aphid genome, in the whole set of CP genes expressed in the retort organs, and in a subset of the Top20 and Top3 most abundant expressors.

(C) Expression of the 50 most expressed genes coding for known CP, ranked by decreasing mean rpkm (reads per kilo base per million mapped reads) of *mxr* expression; data are represented as mean ± SEM. At the low end of the distribution (not shown), 14 ACYPIs (14/143, 10%) are designated as nonexpressed (rpkm<0.5; see Table S2, *1 CPCFC, 11 RR2, and 2 unclassified*), and 79 CP genes have mean rpkm values less than 250. (*13 CP genes represent less than 0.01% of total CP expression each* (*10 RR2, 3 RR1*), *31 represent approximately 0.1-0.5% each* (*2 CPAP1, 2 CPAP3, 1 RR1, 22 RR2, 2 Tweedle, 2 unclassified) and 29 ACYPIs represent more than 0.5% each* (*2 CPAP1, 4 CPAP3, 2 CPF, 1 RR1, 2 RR2, 2 unclassified*)).

16%; and 1 CPAP3, 2%; Figure 3B). This group contains two proteins previously identified in stylets as potential virus receptors (Webster et al., 2018): Stylin-01, ACYPI009006, (2.36% of CP expression; rpkm, reads per kilobase per million mapped reads) and Stylin-02, ACYPI003649 (1.97%, see Table S2). The *rpkm* values of the 50 most highly expressed CP genes are presented in Figure 3C. Figure 3B graphically illustrates the inhomogeneity of presumptive CP expression in stylets across CP classes (RR2, high; RR1/CPAP3, medium; and all the rest form the distribution tail). As a basis for comparison with published and future work, we introduce the Shannon Index of the protein component distribution ($-\sum_{n=1}^{143}(p_i.\ln(p_i))$, see Methods) as a measure of *polymer alloy complexity*. The *Shannon protein complexity index* of aphid stylets is calculated to be 4.1, almost twice the index we calculated for *Bombyx* epidermal cuticle (Wang et al., 2017) and two to four times that of *Drosophila* wing (Sobala and Adler, 2016), and indicates that the stylet polymeric biomaterial is of rather high complexity (see Discussion and Figure S9).

## Pyrokinin (pban, capa) Transcript and Its Mandibular Gland Expression

In addition to the results on CPs, whole-genome analysis of the differential gene expression between the maxillary and mandibular stylets revealed statistically significant differential expression of only two genes between the two ROs (*deseq2* p-adj<0.05). The most significantly differentially expressed gene, ACYPI48299 (FC = 6, *deseq2* p-adj = 0.006; Table S3), encoding a pyrokinin-like set of three neuropeptides (pyrokinin [PK]-like), is specifically expressed in the *mdr* glands. Another peptide of unknown nature, coded for by ACYPI43401 (FC = 14; *deseq2* p-adj = 0.004; Table S3), was also detected among the most differentially expressed genes and found to be expressed in the *mxr* glands. We tried to detect these different peptides by MALDI mass spectrometry in dissected RO glands. Peptide ACYPI43401 was initially targeted to a

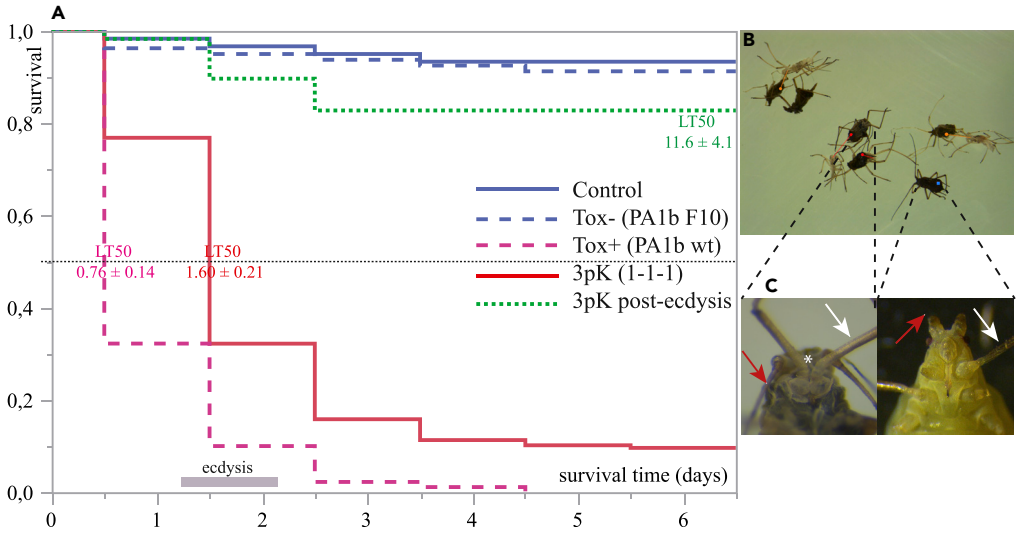



**Figure 4. Injection of Pyrokinin Peptides Induces Molt-associated Mortality and Phenotype**

(A) Aphid survival curves (n = 30) after the injection of pyrokinin peptides (3PK (1-1-1)), PA1b toxin used as a positive control (Tox+, wt, PA1b), PA1b F10 inactivated mutant used as a negative control (Tox-, PA1b F10). Pyrokinins were also injected after the imaginal molt (3 PK post-ecdysis control). Lethal times 50% (LT50) are given as means.

(B) Lethal phenotypes of *A. pisum* injected with an equimolar mixture of three pyrokinins before ecdysis; exuviae were often recovered attached to aphid remains (red dots/lines); green-dot aphids were not headless.

(C) Enlarged image of headless phenotype (left) compared with normal aphid head (right); white and red arrows indicate the left anterior leg and right antenna, respectively, and the asterisk indicates the missing head.

potential candidate at less than Δ 4 Da (predicted peptide at 4865,76 m/z, average mass). Further calibration checks and LC-ESI-MS analyses ruled out the identity to our target sequence (see Methods MS section).

A full functional annotation analysis of retort organ transcripts was performed, including a differential analysis with whole genome gene set, as well as a comparison with another secretory gland system, the salivary glands, reanalyzed from published data (Boulain et al., 2018). The latter analysis resulted in more than 13,000 differential genes and a functional (GO or Kegg) comparison too complex to fit in the present report. The plain comparison of retort organ functions with full genome is shown in Figure S6, resulting in many expected (animal organ morphogenesis, cell redox, chitin metabolism, extracellular location, chitin binding, hexosyl transfer, structural constituents…) and some unexpected annotation enrichments (growth factors and neuropeptide signaling, proteolysis of which serine and metallopeptidases types, neuron and synaptic elements, protein kinase activities) at FDR <0.001 levels.

Having identified a single well-defined differentially expressed gene (ACYPI48299) encoding small predictable and detected peptides (PK1, PK2, PK3, detected by MALDI-TOF in central nervous system but not in RO glands, see Figure S7), we decided to produce and test these peptides to characterize their role in stylet cuticle biogenesis. Their sequences are similar to that of the *Drosophila* CG15520 gene (the two genes *capability* and *hugin* encode PKs in *Drosophila*, *capa*/CG15520 being the closest to ACYPI48299, although orthology is uncertain). Their mature peptide sequences could be predicted by homology with other known PKs (Altstein et al., 2013; Choi et al., 2015). The PK/PBAN bioassay is also well described and routinely used, mainly in Lepidoptera (Choi et al., 2015). Synthetic peptides were delivered by the injection of an equimolar mixture of PK1/PK2/PK3 synthetic peptides 24 h before the L4-46/48 time point. In parallel, a peptide toxin (PA1b wt, Pea Albumin 1, subunit b) and an inactive mutant (PA1b F10A) were injected as positive and negative controls, respectively (Da Silva et al., 2010). The impact of these treatments on the aphid developmental cycle (molting capability) and stylet cuticle formation was observed until the adult stage. Figure 4 shows the result of this survival analysis. The same PK cocktail (PK1/2/3) was injected after the imaginal molt as an additional negative control. The PKs strongly interfered with aphid survival at the imaginal molt. As suspected, no impact was observed when the mixture was injected postecdysis. Stylets collected from dead aphids and observed by scanning electron microscopy did not show altered

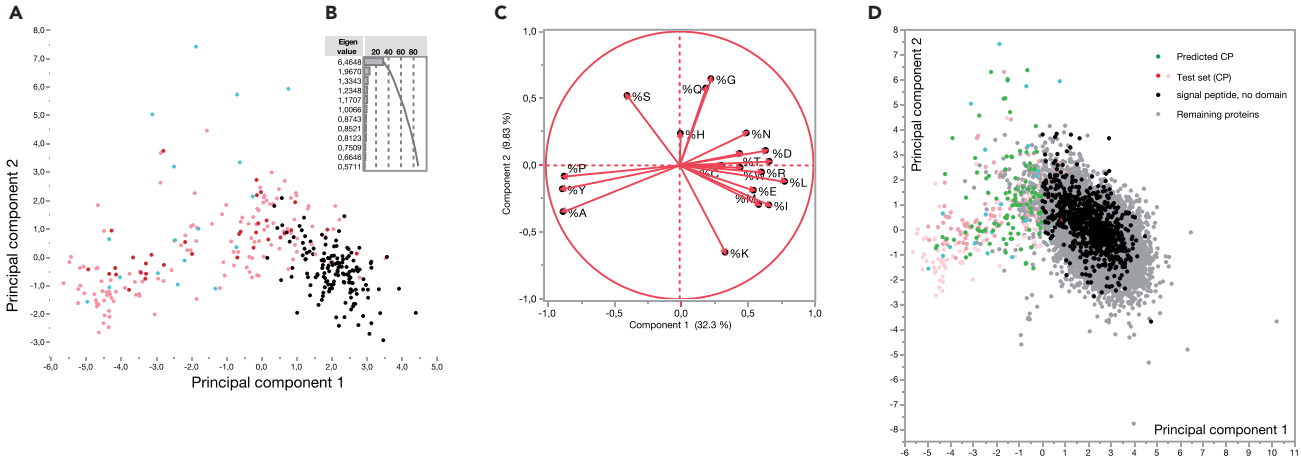



**Figure 5. Identification of New Cuticular Proteins**

(A) PCA on amino acid compositional profiles allowed the identification of new cuticular proteins (CPs) using the test set of the 200 most expressed genes in retort glands, to which a full set of *A. pisum* CPs devoid of their chitin-binding domains was added (Methods; Data S1).

(B) Eigenvalues of the previous PCA showed that the first two axes accounted for more than 42% of cloud variance; the first five axes were considered useful for cuticular protein discrimination and kept for the subsequent analysis.

(C) Variable correlation circle of PCA (first factorial plane).

(D) The full aphid proteome was plotted on the obtained PCA axes, and the first axis served to threshold the prediction of potential new CPs (Methods; Data S1). Color coding: *gray dots*: background proteins either not secreted or with known noncuticular domain annotation; *red dots* (A and D): 143 known CPs; *pink dots* (A and D): the same, with amino acid compositions calculated from extradomain segments (non-chitin-binding segments); *blue dots* (A and D): 18 proteins manually parsed from A for cuticle protein properties; *green dots* (D): 121 proteins predicted from whole proteome compositional analysis to be new potential CPs; *black dots*: noncuticular proteins from the test set (A, annotated as other proteins) and from the whole genome set (D, no signal peptide prediction and no domain annotation, but improper amino acid compositional bias).

morphology. In contrast, many of the PK-treated dead aphids presented a strong phenotype, *headless* individuals, as shown in Figure 4B. We checked the effect of single pyrokinin peptides injected in the same timeframe and concentrations. No single peptide did reproduce the same survival or ecdysis phenotype than the mixture (Figure S8), but each of them displayed different survival effects, with PK1 ≈ control < PK3 ≈ PK2 < PK1+2 + 3 and the three groups showing statistically significant survival phenotypes, LogRank statistics of control/PK1 vs group PK2/3 being $p(>\chi 2) = 0.0010$, whereas the latter vs PK1+2 + 3 being $p(>\chi 2) = 0.0050$. Figure S8 shows lethal times 50 (LT50) of all active treatments and the magnified headless phenotype induced by the mixture injection. The most active single peptide was the proline-rich SPPYSPPFSPRL*amide* but did not induce the full effect of the 3PK mixture, as almost one-third of individual survived the experiment and ecdysis; no headless phenotype was observed in single peptide injections.

## RO Expression Analysis Defines a Potential New Class of CPs

We combined the data on the *mxr* and *mdr* glands and parsed the high-expression genes for structural proteins not yet annotated as CPs through the HMM-based tool CutProtFamPred (Ioannidou et al., 2014). First manually and then computationally through a procedure defined in the Methods section, we identified among the 200 most expressed genes a series of secreted proteins with a striking imbalance in amino acid composition. These proteins may be defined as low-complexity proteins and display an amino acid composition globally similar to the composition of already known (annotated) CPs but contain no identified conserved region (that could be captured by existing HMM-based protein family detection procedures). As such proteins were expressed in a tissue devoted to producing CP, we decided to consider them CPs. Many individual members of this potential new family were already described individually in previous works (Cornman, 2010; Cornman and Willis, 2009; Andersen, 2011; Mun et al., 2015), and we designed our non-homology-based procedure to capture them in any insect genome. In our RO "most expressed 200" list, we found 19 of them (Figure 5A), including the most highly expressed structural protein ACYPI087730 (overall new + old CPs make up 35% of expression of the total expression in this top200 set, whereas old CPs made up "only" 21%). Applying our automated definition to the whole aphid genome captured a total of 121 new CPs (Figure 5D; total CPs making up 18% of the total RO expression, on a whole genome basis, versus 10.6% for the previously known CPs). When comparing the distributions of expression

levels of transcripts from "new" and "old" CPs within the RO gland, it comes that the first moments of these distributions (mean and variance) do not differ significantly: means were 371 ± 67 rpkm and 296 ± 73 rpkm, respectively ("old" and "new," mean ± SE, Anova $p > F = 0.448$); Levene test for variance homogeneity, $p = 0.69$. Such a validation by expression may be taken as an additional cue that the new set of protein behaves as cognate cuticular proteins. If validated by the community, this class could then define ≈ +80% new CPs in addition to the presently known CP classes (Ioannidou et al., 2014; Magkrioti et al., 2004; Willis, 2010).

## DISCUSSION

Among insect cuticle regions, the mouthparts are the most specialized structures and are crucial players in feeding (Busse and Gorb, 2018), sensorial exchanges, and interactions with environmental, gut-resident, and insect-transmitted microbes. In aphids, the transmission of noncirculative viruses is associated with a specific micro-organ (the acrostyle) at the tip of the needle-like mouthparts, the stylets (Uzest et al., 2007; Webster et al., 2018). In this context, the full transcriptomic characterization of the cuticular composition of the stylets was performed to complement ongoing proteomic studies (unpublished data). This approach provides the most extensive catalog to date as a foundation for further work, due to both the high sensitivity of transcriptomics and its precursor status preceding the very high level of posttranslational modifications known to affect cuticular material (Mun et al., 2015; Andersen, 2010).

### Glands, Expression Profiles, and the Spatiotemporal Limit

Our main goal was the complete definition of the (potential) protein composition of a stylet cuticle. We opted for RNA-Seq analysis of the cuticular glands, specifically termed *retort organs* (ROs) in aphids (Ponsen, 1972), which secrete the two mandibular and two maxillary stylets at each molt. Our study provided the only molecular study of retort organs, which have remained completely unexplored since 1972 (Ponsen, 1972). Immunochemical analysis confirmed the main role of this organ in the secretion of stylet CPs, showing the extensive distribution of labels associated with antibodies raised against a main CP, Stylin-02 (ACYPI003649), within the retort glands (Figures 2C and S5). The comparative analysis of *mxr* and *mdr* stylet expression showed unexpected similarity and therefore a common global composition of the two aphid stylets, despite their very different shapes and functions and their ancient evolutionary origin (Angelini and Kaufman, 2005; Cicero, 2016; Faucheux, 1975; Forbes, 1966; Uzest et al., 2010). We are, however, quite aware of the challenging spatiotemporal resolution limit. Even a 3–4 h temporal slot lies far behind potential variations in individual gene expression profiles in such a highly critical pre-ecdysial period (Dittmer et al., 2015; Pan et al., 2015; Charles, 2010; Willis, 1996), as this was clearly shown in the *Drosophila* pupal wing model (Sobala and Adler, 2016). This limitation may also apply to potential individual cell heterogeneity occurring in this complex organ (Ponsen, 1972), composed of more than 50 matrix cells per gland (Figures 2C and 2E and Video S2 and S3). In other cuticular models, such as the elytra and forewings of the red flour beetle *T. castaneum*, the very different shapes and mechanical properties of the two cuticle types were associated with very different CP compositions (Dittmer et al., 2012). We feel that only single-cell single-individual approaches with proper post hoc data analysis methods (Leonavicius et al., 2018) could reasonably overcome such limits and improve our results. It should also be noted that ongoing proteomic work on both retort organ and casted stylets (unpublished) will soon allow us to have an extended temporal view of aphid mouthpart cuticular peptides, ranging from their expression to their secretion as proteins within their biogenetic glands, ending-up to cognate and sclerotized cuticular proteins. This explains that our present work only comprises mass spectrometry confirmation of the peptides corresponding to differentially expressed transcripts in maxillary vs mandibular ROs, and not the identification of cuticular proteins, which need appropriate extraction protocols not used herein (Masson et al., 2018). Actually, our present proteomic results were unable to locate the peptide issued from ACYPI43401 (expected mature peptide IKGRNRRMLANRIPGPEGTFIVGMLPLAIQGAEQKIKGAQDVYR), neither in RO (crude organ MALDI shots or extraction and full LC MS/MS analysis) nor in dissected nervous system extracts. In contrast, were were able to identify in the CNS (not RO glands) the three pyrokinin peptides processed from ACYPI48299's prepro-protein, predicted by orthology prediction from drosophila or lepidopteran genes (see Figure S7 for leucineamide peptide sequences). This is both congruent with previously reviewed pea aphid neuropeptides (Huybrechts et al., 2010), also identifying pyrokinins from different neural tissues (brain, retrocerebral complex, suboesophageal ganglion, and thoracic ganglia), and from general knowledge on insect pyrokinins often displaying divergent transcription (mRNA) and mature peptide storage sites (neuropeptide stores as granular bodies) (Altstein et al., 2013; Predel and Nachman, 2006). We may infer that potential transcriptionally active glands such as those displayed in Figure 2E (*dsh*, dumbbell-shaped cells) might have been associated with mandibular glands, and displayed pyrokinin transcripts,

whereas only nervous system *per se* did contain the active peptides characterized by mass spectrometry. Such an interpretation should however be properly confirmed by adequate neuroimagery immunolabelling, outside the scope of the present work.

### A New Class of CPs: How General?

As we did not identify a significant difference between maxillary and mandibular stylet composition, we focused on analyzing the spectrum of CPs in both stylets. Currently, the only consensus method for identifying CPs in insect genomes is based on homology and HMM-based domain identification (Ioannidou et al., 2014). These domains usually represent identified molecular functions and are present in all classes of CPs as chitin-binding domains (Dong et al., 2016).

From these already identified CP sets, three major expressors (ACYPI004074, ACYPI086655, and ACYPI56622) make together 22% of the total CP composition. They belong to the CPR_RR-2 subfamily and all three to a canonical subtype with a central chitin-binding domain and identifiable N- and C-terminal tails with a so-called "PAYSA" motif (making these low-complexity domains enriched in the corresponding amino acids). The following "top20" group shows, in contrast, that the CPR_RR-1 family is overrepresented in this high-expression category of CPs in the retort organs (Figure 3B). This is to be linked with the overrepresentation of this RR1 family in the virus receptor candidates set (Webster et al., 2018), also characterized by their availability to stylet surface/antibodies. Overall, this CPR_RR-1 subfamily is undoubtedly related to surface interactions between nonpersistently transmitted viruses and is most likely overrepresented at the surface of the acrostyle.

In addition to these well-identified structural CPs, our computational approach is based on both low complexity and specific compositional skewness (i.e., compositional similarity to cognate CPs), which allowed us to identify new CPs that were already "randomly" picked-up by analytical works in some other insect models, such as *Bombyx mori* (Dong et al., 2016), *Manduca sexta* (Tetreau et al., 2015a, 2015b), *Anopheles gambiae* (Zhou et al., 2016, 2017), and *Dendrolimus punctatus* (Yang et al., 2017). The most recently identified family of CPs was the CPAP1/3 family (Pan et al., 2018), but all other transcripts or proteins identified individually from -omic analyses of insect cuticles were classified as "ungrouped CPs" (NlGp21.92 in *N. lugens* (Lu et al., 2018)) or as poorly defined low-complexity classes.

Our attempt to classify this low-complexity group comprised the establishment of our prediction on a test set (the 200 most expressed genes in ROs being parsed for genes coding for unannotated secreted proteins that show both a low-complexity index and a clustering in amino acid composition with standard HMM-predicted CPs; Figure 5) followed by extension to a full set/complete genome (Figures 5B and 5C). This strategy identified 19 new CPs in the RO (in addition to the 36 already known CPs), including the two most expressed structural proteins, ACYPI087730 (with expression twice that of the most expressed CPR_RR-2 CP gene ACYPI004074) and ACYPI001038 (with a level similar to that of ACYPI004074). For the whole genome, we identified a total of 121 new proteins that can be potentially classified as CPs.

In this group, we also identified proline-rich (containing many repeats of P[Y/V] dipeptides, often in PYPV blocks), as well as glycine-rich proteins, a very frequent feature in proteins that occurs in approximately thirty Interpro domains (Mitchell et al., 2019). These two types of proteins are hydrophobic and may contribute to the sclerotization process during the dehydration of cuticle, thereby increasing its rigidity. The predominance of these CP types was already proposed for the elytral cuticle of *T. castaneum* (Dittmer et al., 2012), and stiffness but not brittleness was proposed as a specific feature of such stress-acting cuticle regions (Andersen, 2000).

Introducing a global measure of *CP complexity*, the cψ index (see Methods) based on Shannon entropy (Shannon, 1997), allowed us to compare the stylet data with the two other whole-genome transcriptomic analyses of cuticle-producing tissues (Wang et al., 2017; Dittmer et al., 2012; Sobala and Adler, 2016). One NGS analysis of *Bombyx* larval epithelium (Wang et al., 2017) displayed a complexity index of 2.41 (11.2 protein-equivalents), and another microarray analysis of coleopteran wings in *T. castaneum* revealed very low complexity of the elytran cuticle (cψ of 1.72 and 5.6 protein-equivalents) and moderate complexity of the hind-wings (cψ 2.30; 10.0 protein-equivalents). Such expression data for *Drosophila* wing (Sobala and Adler, 2016) also rendered a rather low complexity, ranging from cψ index of 1–2.3 (≈ 3 to 10 protein-equivalents) for the different developmental stages of pupal wings. The aphid stylet data, with a cψ of 4.07 and

≈56 protein-equivalents, show that at least five times more protein entities (58/11) are needed for this material than for any other known cuticle (Figure S9). How this complexity is reflected in the mechanical and biological properties or the topological organization of the proteins within the stylet cuticle is uncertain. However, it likely correlates with the diversity of functions supported by these specialized mouthparts (microfluidics, proprioception, mechanics, microbial interactions, and functional micromorphology).

### Stylet Cuticular Material: Proteins, Chitin-derived Polymer(s), Lipids, and Sclerotizing Matrix

Quantitative analysis of expression over CP classes (Wilcoxon test over all class pairs) identifies two features: higher expression of RR1 proteins than RR2 (shown in Figure 3B, and highly significant by a Fisher exact test of RR1 vs expression classes) and higher expression of CPAP3 over CPAP1 (Wilcoxon test, $p = 0.046$). Together with the higher surface exposure of stylet RR1 proteins to antibodies, mentioned above, this result may reflect the fact that CPAP3 proteins are larger than both CPR and CPAP1 proteins, containing three cysteine-structured chitin-binding domains (vs one, cysteine-free, for the CPR family). It is likely that these four major and contrasting classes constituting aphid stylets display a variable topological distribution within the stylet (Webster et al., 2018) and between the two stylet types (although not globally significant, many expression ratios are greater than two, as illustrated in Figure 3A); the contrasting sizes, chitin-binding stoichiometry, RR-subtype motif, and the presence or absence of structuring domains that are or are not sensitive to the oxidative stress of sclerotization are potential determinants of the selectivity needed for the extreme morphological variation displayed by stylets at perimicron scales. Adding all (new) low-complexity components that lack any conserved domain, and probably any chitin-binding module, should allow fine-tuning of the cohesive properties of the cuticle: components with high proportions of tyrosine and histidine residues are prone to dense, irreversible covalent cross-linking of the protein chains (Mun et al., 2015; Sugumaran, 2010) and contrast with the reversible nature of either chitin-binding domains or non-covalently linked low complexity domains (Pena-Francesch and Demirel, 2019). This feature may recall the sacrificial bonding properties of fibrous proteins such as silks (Brown et al., 2011) or elastic proteins, which are often displayed at the intramolecular scale but can also be translated to higher (supramolecular) scales in the cuticle network, as in fibrous artificial polymers (Passieux et al., 2015). Finally, it is noteworthy that two proteolytic enzyme classes are enriched over their whole-genome counterparts in the expression pattern of the RO secretome (serine- and metallo-endoproteases; Figure S6), which indicates that protein remodeling in the extracellular matrix is needed for the maturation and full functioning of the stylets; whether this remodeling occurs before or after sclerotization is not known.

The activity of the glands also reflects the nonprotein components of the cuticle (chitin, lipids, and sclerotization). The unique aphid chitin-synthase *chs1* is expressed at significant levels (aphids do not possess a gut-associated peritrophic membrane or its specifically expressed *chs2* gene; ACYPI065097, Table S3). All five aphid chitin-deacetylases are expressed in the glands, and two are high expressors (ACYPI002929 and ACYPI001932 in the top100, Table S3), illustrating the essentiality of these enzymes (Yu et al., 2016, 2018) even in cuticles with noncanonical, i.e., epithelial, organization. Finally, the two-catalytic-domain chitinase recently described as essential to cuticular organization in *T. castaneum* (Noh et al., 2018), but not to the molting cycle as is Ap-*Cht5* (Figures 1 and S2), was also found to be very highly expressed in the RO (ACYPI006403, Ap-*cht4*); the same holds for the chitinase-like ortholog of the *Drosophila* IDGF gene (ACYPI001365, Ap-*cht1* following the aphid chitinase nomenclature (Nakabachi et al., 2010)); both genes are 1-to-1 orthologs of their holometabolous counterparts and may also act as "organizing chitinases" in the aphid RO. The last and most complex component of the cuticle, the tanning-sceloritizing linkers, is indicated by very high expression of key genes in the process: tyrosine hydroxylase, which scavenges tyrosine to form the first specific metabolite in the pathway, DOPA, displays a *rpkm* of more than 2000 and an expression rank of 144, and more than ten of the key genes are in the 500 first expressors (*ebony*, *tan*, dopa decarboxylase, laccase, etc.; Table S3). This phenomenon is unsurprising for a cuticular gland but noteworthy for a structure displaying very specialized mechanical properties in terms of both strength and flexibility.

### Molting Process in the Head and the Mouthparts: A More Complex Situation

Our study identified only two genes with a statistically significant differential expression between the two stylet glands. One gene, ACYPI48299, encodes three PK-like neuropeptides that are specifically expressed in the mandibular glands. We evaluated their role by injection before and after the imaginal molt, which showed clear interference of the PKs with aphid survival at the imaginal molt and strong head-related

altered phenotypes. As with other neuropeptides, PKs have been annotated (Christie, 2008) and even experimentally tested *per os* on aphids (Nachman et al., 2012). The latter study showed acute toxicity and antifeeding activity of both stabilized PK derivatives and agonists of PK receptors, suggesting a neuroendocrine target potentially related to gut motility (Nachman et al., 2012). Our injection results are also compatible with a canonical myotropic target, perhaps related to the anatomical extrusion of stylets during the molt (overexpression of PK/ACYPI48299 in the mandibular glands). One striking feature of the mandibular glands during dissection is their association with a couple of visible dumbbell-shaped organs with a potential neurosecretory nature (histological features; Figure 2F). Energy metabolism was found to be extremely active in the RO (vs salivary glands, for example, not shown), but whether this syndrome is associated with hyperactive secretion or with mechanical movements within the organ, which are apparently devoid of muscular fibers, remains conjectural.

### Conclusions

Our transcriptomic study gives the most extensive list to date of genes expressed in aphid retort glands at the time of casting presumptive adult stylets and therefore constitutes the best available basis for the complete protein composition of an insect mouthpart, usefully completing novel approaches on such materials (Busse and Gorb, 2018). This cuticle is more complex (i.e., more diverse in protein profile) than other "model cuticles" (*Tribolium* elytra, *Bombyx* epithelium, *Drosophila* wings). The work (1) provides a solid comparative basis for aphid stylet proteomics and its crucial importance in deciphering the molecular mechanisms of transmission of plant viruses and (2) with the support from ongoing stylet proteomics results, allows us to initiate our work on the biomimetics of aphid acrostyle cuticles and further on hemipteran stylet cuticles, still in the wide context of vector biology.

### Limitations of the Study

In this work, we analyzed the stylet glands (retort organ) at one time illustrating the expression peak of the organ as a whole. This organ is constituted of a series of more than fifty secretory cells with potential different time-dependent expressions. Therefore, the time (peak within instar) and space (cells, organ, MXR vs MDR) resolution of our study is only a first approach to the biology of this retort/stylet organ. Single cell transcriptomics is presently not reachable on such a tissue, probably requiring *ex-vivo* organ cultivation, out-of-reach of the current knowhow and knowledge of our field and insect model. A first step toward such objectives will be the precise analysis of this organ's evolution within the last larval instar, which is a reachable short-term objective with the techniques developed within the present paper, CLSM for example.

## METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

Supplementary data files at data.inra.fr.

Rahbe, Yvan; Guschinskaya, Natalia; Ressnikoff, 2019, "Confocal cleared L4 aphid head-stylets", https://doi.org/10.15454/VPUGUG, Portail Data Inra, V1.

Rahbe, Yvan; Guschinskaya, Natalia; Ressnikoff, Denis, 2019, "Histology of L4 aphid head-stylets," https://doi.org/10.15454/QUM1LQ, Portail Data Inra, V1.

## DATA AND CODE AVAILABILITY

Sequence data are available at GenBank SRA archive as Genbank: PRJNA510301 and histology data available at data.inra.fr archive, urls: https://doi.org/10.15454/VPUGUG and https://doi.org/10.15454/QUM1LQ.

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.100828.

## ACKNOWLEDGMENTS

PRABI (https://prabi.fr) and BIPAA (https://bipaa.genouest.org) bioinformatics platforms are acknowledged for Galaxy support (Fabrice Legeai, Christine Oger, Vincent Navratil, Stéphanie Robin); the

CIQLE platform (http://ciqle.univ-lyon1.fr) is acknowledged for histology support in aphid head serial sectioning (Annabelle Bouchardon, Bruno Chapuis, BatoulSmatti); Catherine Rey and Joel Lachuer (ProfileExpert) are thanked for their help with NGS processing and initial help on Laser Capture Microdissection. The ANR grant StylHook (ANR-15-CE20-0011) is acknowledged for funding and salary (GN); and Nicolas Parisot and Guy Perrière are warmly thanked for python script help in the compositional analysis of proteomes. FR3728 BioEnviS is acknowledged for its Plant facility (Plateforme Serres et chambres climatiques). We thank AJE (American Journal Experts, Durham NC USA) for their diligent language check.

## AUTHOR CONTRIBUTIONS

N.G. and Y.R. conceived this work, analyzed the data, and wrote the manuscript; N.G. conducted experiments (molecular biology, dissections, NGS analysis); Y.R. conducted experiments (microinjection and survival assays); N.G. conducted analysis of MALDI-TOF and LC MS/MS; K.A. P.B. and S.V. helped analyzing MS data; D.R. performed image analysis; M.U. designed and coordinated the StylHook ANR grant project, and provided immunological tools.

## DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: April 15, 2019
Revised: November 3, 2019
Accepted: January 6, 2020
Published: February 21, 2020

## REFERENCES

Altstein, M., Hariton, A., and Nachman, R.J. (2013). Chapter 37-FXPRLamide (pyrokinin/PBAN) family A2-Kastin. In Handbook of Biologically Active Peptides, Second Edition, J. Abba, ed. (Academic Press), pp. 255–266.

Andersen, S.O. (1979). Biochemistry of insect cuticle. Annu. Rev. Entomol. *24*, 29–61.

Andersen, S.O. (2000). Studies on proteins in post-ecdysial nymphal cuticle of locust, Locusta migratoria, and cockroach, Blaberus craniifer. Insect Biochem. Mol. Biol. *30*, 569–577.

Andersen, S.O. (2010). Insect cuticular sclerotization: a review. Insect Biochem. Mol. Biol. *40*, 166–178.

Andersen, S.O. (2011). Are structural proteins in insect cuticles dominated by intrinsically disordered regions? Insect Biochem. Mol. Biol. *41*, 620–627.

Angelini, D.R., and Kaufman, T.C. (2004). Functional analyses in the hemipteran Oncopeltus fasciatus reveal conserved and derived aspects of appendage patterning in insects. Dev. Biol. *271*, 306–321.

Angelini, D.R., and Kaufman, T.C. (2005). Insect appendages and comparative ontogenetics. Dev. Biol. *286*, 57–77.

Awuoche, E.O., Weiss, B.L., Vigneron, A., Mireji, P.O., Aksoy, E., Nyambega, B., Attardo, G.M., Wu, Y., O'neill, M., Murilla, G., et al. (2017). Molecular characterization of tsetse's proboscis and its response to Trypanosoma congolense infection. PLoS Negl. Trop. Dis. *11*, e0006057.

Boulain, H., Legeai, F., Guy, E., Morlière, S., Douglas, N.E., Oh, J., Murugan, M., Smith, M., Jaquiéry, J., Peccoud, J., et al. (2018). Fast evolution and lineage-specific gene family expansions of aphid salivary effectors driven by interactions with host-plants. Genome Biol. Evol. *10*, 1554–1572.

Brown, C.P., Macleod, J., Amenitsch, H., Cacho-Nerin, F., Gill, H.S., Price, A.J., Traversa, E., Licoccia, S., and Rosei, F. (2011). The critical role of water in spider silk and its consequence for protein mechanics. Nanoscale *3*, 3805–3811.

Busse, S., and Gorb, S.N. (2018). Material composition of the mouthpart cuticle in a damselfly larva (Insecta: Odonata) and its biomechanical significance. R. Soc. Open Sci. *5*, 172117.

Charles, J.P. (2010). The regulation of expression of insect cuticle protein genes. Insect Biochem. Mol. Biol. *40*, 205–213.

Choi, M.-Y., Sanscrainte, N.D., Estep, A.S., Vander Meer, R.K., and Becnel, J.J. (2015). Identification and expression of a new member of the pyrokinin/pban gene family in the sand fly Phlebotomus papatasi. J. Insect Physiol. *79*, 55–62.

Christie, A.E. (2008). In silico analyses of peptide paracrines/hormones in Aphidoidea. Gen. Comp. Endocrinol. *159*, 67–79.

Cicero, J.M. (2016). Stylet biogenesis in bactericera cockerelli (Hemiptera: Triozidae). Arthropod. Struct. Dev. *46*, 644–661.

Cornman, R.S. (2010). The distribution of GYR- and YLP-like motifs in Drosophila suggests a general role in cuticle assembly and other protein-protein interactions. PLoS One *5*, e12536.

Cornman, R.S., and Willis, J.H. (2009). Annotation and analysis of low-complexity protein families of Anopheles gambiae that are associated with cuticle. Insect Mol. Biol. *18*, 607–622.

Da Silva, P., Rahioui, I., Laugier, C., Jouvensal, L., Meudal, H., Chouabe, C., Delmas, A.F., and Gressent, F. (2010). Molecular requirements for the insecticidal activity of the plant peptide pea albumin 1 subunit b (PA1b). J. Biol. Chem. *285*, 32689–32694.

Davidson, J. (1913). Memoirs: the structure and biology of Schizoneura lanigera, Hausmann or woolly Aphis of the apple tree. Q. J. Microscopical Sci. *653*, s2–58.

Dittmer, N.T., Hiromasa Y Fau - Tomich, J.M., Tomich Jm Fau - Lu, N., Lu N Fau - Beeman, R.W., Beeman Rw Fau - Kramer, K.J., Kramer Kj Fau - Kanost, M.R., and Kanost, M.R. (2012). Proteomic and transcriptomic analyses of rigid and membranous cuticles and epidermis from the elytra and hindwings of the red flour beetle, Tribolium castaneum. J. Proteome Res. *11*, 269–278.

Dittmer, N.T., Tetreau, G., Cao, X., Jiang, H., Wang, P., and Kanost, M.R. (2015). Annotation and expression analysis of cuticular proteins from the tobacco hornworm, Manduca sexta. Insect Biochem. Mol. Biol. *62*, 100–113.

Dong, Z., Zhang, W., Zhang, Y., Zhang, X., Zhao, P., and Xia, Q. (2016). Identification and characterization of novel chitin-binding proteins from the larval cuticle of silkworm, Bombyx mori. J. Proteome Res. *15*, 1435–1445.

Faivre, J., Sudre, G., Montembault, A., Benayoun, S., Banquy, X., Delair, T., and David, L. (2018). Bioinspired microstructures of chitosan hydrogel

provide enhanced wear protection. Soft Matter. *14*, 2068–2076.

Faucheux, M.-J. (1975). Relations entre l'ultrastructure des stylets mandibulaires et maxillaires et la prise de nourriture chez les insectes Hémiptères. C.R. Acad. Sci. Paris *281*, 41–44.

Forbes, A.R. (1966). Electron microscope evidence for nerves in the mandibular stylets of the green peach aphid. Nature *212*, 726.

Futuyma, D.J., and Agrawal, A.A. (2009). Macroevolution and the biological diversity of plants and herbivores. Proc. Natl. Acad. Sci. U S A *106*, 18054–18061.

Garrouste, R., Clement, G., Nel, P., Engel, M.S., Grandcolas, P., D'haese, C., Lagebro, L., Denayer, J., Gueriau, P., Lafaite, P., et al. (2012). A complete insect from the Late Devonian period. Nature *488*, 82–85.

Goncalves, I.R., Brouillet, S., Soulie, M.C., Gribaldo, S., Sirven, C., Charron, N., Boccara, M., and Choquer, M. (2016). Genome-wide analyses of chitin synthases identify horizontal gene transfers towards bacteria and allow a robust and unifying classification into fungi. BMC Evol.Biol. *16*, 252.

Heriot, A.D. (1934). The renewal and replacement of the stylets of sucking insects during each stadium, and the method of penetration. Can. J. Res. *11*, 602–612.

Huang, D.Y., Bechly, G., Nel, P., Engel, M.S., Prokop, J., Azar, D., Cai, C.Y., Van De Kamp, T., Staniczek, A.H., Garrouste, R., et al. (2016). New fossil insect order Permopsocida elucidates major radiation and evolution of suction feeding in hemimetabolous insects (Hexapoda: acercaria). Sci. Rep. *6*, 23004.

Huybrechts, J., Bonhomme, J., Minoli, S., Prunier-Leterme, N., Dombrovsky, A., Abdel-Latief, M., Robichon, A., Veenstra, J.A., and Tagu, D. (2010). Neuropeptide and neurohormone precursors in the pea aphid, Acyrthosiphon pisum. Insect Mol. Biol. *19*, 87–95.

Ioannidou, Z.S., Theodoropoulou, M.C., Papandreou, N.C., Willis, J.H., and Hamodrakas, S.J. (2014). CutProtFam-Pred: detection and classification of putative structural cuticular proteins from sequence alone, based on profile Hidden Markov Models. Insect Biochem. Mol. Biol. *52*, 51–59.

Iyer, E.P.R., and Cox, D.N. (2010). Laser capture microdissection of Drosophila peripheral neurons. J. Vis. Exp. *39*, e2016.

Lease, H.M., and Wolf, B.O. (2010). Exoskeletal chitin scales isometrically with body size in terrestrial insects. J. Morphol. *271*, 759–768.

Leonavicius, K., Nainys, J., Kuciauskas, D., and Mazutis, L. (2018). Multi-omics at single-cell resolution: comparison of experimental and data fusion approaches. Curr.Opin.Biotechnol. *55*, 159–166.

Lu, J.-B., Luo, X.-M., Zhang, X.-Y., Pan, P.-L., and Zhang, C.-X. (2018). An ungrouped cuticular protein is essential for normal endocuticle formation in the brown planthopper. Insect Biochem. Mol. Biol. *100*, 1–9.

Magkrioti, C.K., Spyropoulos, I.C., Iconomidou, V.A., Willis, J.H., and Hamodrakas, S.J. (2004). cuticleDB: a relational database of Arthropod cuticular proteins. BMC Bioinformatics *5*, 138.

Masson, V., Arafah, K., Voisin, S., and Bulet, P. (2018). Comparative proteomics studies of insect cuticle by tandem mass spectrometry: application of a novel proteomics approach to the pea aphid cuticular proteins. Proteomics *18*, 1700368.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. Science *346*, 763–767.

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. *47*, D351–D360.

Mun, S., Young Noh, M., Dittmer, N.T., Muthukrishnan, S., Kramer, K.J., Kanost, M.R., and Arakane, Y. (2015). Cuticular protein with a low complexity sequence becomes cross-linked during insect cuticle sclerotization and is required for the adult molt. Sci. Rep. *5*, 10484.

Nachman, R.J., Hamshou, M., Kaczmarek, K., Zabrocki, J., and Smagghe, G. (2012). Biostable and PEG polymer-conjugated insect pyrokinin analogs demonstrate antifeedant activity and induce high mortality in the pea aphid Acyrthosiphon pisum (Hemiptera: aphidae). Peptides *34*, 266–273.

Nakabachi, A., Shigenobu, S., and Miyagishima, S. (2010). Chitinase-like proteins encoded in the genome of the pea aphid, Acyrthosiphon pisum. Insect Mol. Biol. *19* (*Suppl 2*), 175–185.

Nel, P., Bertrand, S., and Nel, A. (2018). Diversification of insects since the Devonian: a new approach based on morphological disparity of mouthparts. Sci. Rep. *8*, 3516.

Noh, M.Y., Muthukrishnan, S., Kramer, K.J., and Arakane, Y. (2018). A chitinase with two catalytic domains is required for organization of the cuticular extracellular matrix of a beetle. PLOS Genet. *14*, e1007307.

Oliveira, D.S., Brito, N.F., Nogueira, F.C.S., Moreira, M.F., Leal, W.S., Soares, M.R., and Melo, A.C.A. (2017). Proteomic analysis of the kissing bug Rhodnius prolixus antenna. J. Insect Physiol. *100*, 108–118.

Pan, H., Yang, X., Siegfried, B.D., and Zhou, X. (2015). A comprehensive selection of reference genes for RT-qPCR analysis in a predatory lady beetle, Hippodamia convergens (Coleoptera: Coccinellidae). PLoS One *10*, e0125868.

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

Passieux, R., Guthrie, L., Rad, S.H., Levesque, M., Therriault, D., and Gosselin, F.P. (2015). Instability-assisted direct writing of microstructured fibers featuring sacrificial bonds. Adv. Mater. *27*, 3676–3680.

Pena-Francesch, A., and Demirel, M.C. (2019). Squid-inspired Tandem repeat proteins: functional fibers and films. Front. Chem. *7*, 69.

Pinet, J.M. (1968). Structure et formation des coaptations des stylets maxillaires de Rhodnius prolixus [Heter.Reduviidae]. Ann. Soc. Entomol. Fr. *4*, 455e475.

Ponsen, M.B. (1972). The Site of Leafroll Virus Multiplication in its Vector, Myzus persicae. An Anatomical Study, Ph.D. Thesis (WAU (Wageningen Agricultural University)).

Predel, R., and Nachman, R.J. (2006). CHAPTER 32-the FXPRLamide (pyrokinin/PBAN) peptide family. In Handbook of Biologically Active Peptides, A.J. Kastin, ed. (Academic Press), pp. 207–212.

Shannon, C.E. (1997). The mathematical theory of communication. 1963. MD.Comput. *14*, 306–317.

Sobala, L.F., and Adler, P.N. (2016). The gene expression program for the formation of wing cuticle in Drosophila. PLoS Genet. *12*, e1006100.

Sugumaran, M. (2010). Chemistry of cuticular sclerotization. Adv. Insect Physiol. *39*, 151–209.

Tetreau, G., Cao, X., Chen, Y.-R., Muthukrishnan, S., Jiang, H., Blissard, G.W., Kanost, M.R., and Wang, P. (2015a). Overview of chitin metabolism enzymes in Manduca sexta: identification, domain organization, phylogenetic analysis and gene expression. Insect Biochem. Mol. Biol. *62*, 114–126.

Tetreau, G., Dittmer, N.T., Cao, X., Agrawal, S., Chen, Y.-R., Muthukrishnan, S., Haobo, J., Blissard, G.W., Kanost, M.R., and Wang, P. (2015b). Analysis of chitin-binding proteins from Manduca sexta provides new insights into evolution of peritrophin A-type chitin-binding domains in insects. Insect Biochem. Mol. Biol. *62*, 127–141.

Uzest, M., Gargani, D., Dombrovsky, A., Cazevieille, C., Cot, D., and Blanc, S. (2010). The "acrostyle": a newly described anatomical structure in aphid stylets. ArthropodStruct. Dev. *39*, 221–229.

Uzest, M., Gargani, D., Drucker, M., Hébrard, E., Garzo, E., Candresse, T., Fereres, A., and Blanc, S. (2007). A protein key to plant virus transmission at the tip of the insect vector stylet. Proc. Natl. Acad. Sci. U S A *104*, 17959–17964.

Vaclaw, M.C., Sprouse, P.A., Dittmer, N.T., Ghazvini, S., Middaugh, C.R., Kanost, M.R., Gehrke, S.H., and Dhar, P. (2018). Self-assembled coacervates of chitosan and an insect cuticle protein containing a Rebers-Riddiford motif. Biomacromolecules *19*, 2391–2400.

Wang, P., Qiu, Z., Xia, D., Tang, S., Shen, X., and Zhao, Q. (2017). Transcriptome analysis of the epidermis of the purple quail-like (q-lp) mutant of silkworm, Bombyx mori. PLoS One *12*, e0175994.

Webster, C.G., Pichon, E., Van Munster, M., Monsion, B., Deshoux, M., Gargani, D., Calevro, F., Jimenez, J., Moreno, A., Krenz, B., et al. (2018). Identification of plant virus receptor candidates in

the stylets of their aphid vectors. J. Virol. *92*, e00432–e00518.

Webster, C.G., Thillier, M., Pirolles, E., Cayrol, B., Blanc, S., and Uzest, M. (2017). Proteomic composition of the acrostyle: novel approaches to identify cuticular proteins involved in virus–insect interactions. Insect Sci. *24*, 990–1002.

Weirauch, C., and Schuh, R.T. (2011). Systematics and evolution of Heteroptera: 25 years of progress. Annu. Rev. Entomol. *56*, 487–510.

Willis, J. (1996). Metamorphosis of the cuticle, its proteins, and their genes. In Metamorphosis. Postembryonic Reprogramming of Gene Expression in Amphibians and Insect Cells, L.I. Gilbert, J.R. Tata, and B.G. Atkinson, eds. (Academic Press), pp. 253–282.

Willis, J.H. (2010). Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. Insect Biochem. Mol. Biol. *40*, 189–204.

Willis, J.H. (2018). The evolution and metamorphosis of arthropod proteomics and genomics. Annu. Rev. Entomol. *63*, 1–13.

Yang, C.-H., Yang, P.-C., Zhang, S.-F., Shi, Z.-Y., Kang, L., and Zhang, A.-B. (2017). Identification, expression pattern, and feature analysis of cuticular protein genes in the pine moth Dendrolimus punctatus (Lepidoptera: lasiocampidae). Insect Biochem. Mol. Biol. *83*, 94–106.

Yu, R., Liu, W., Li, D., Zhao, X., Ding, G., Zhang, M., Ma, E., Zhu, K., Li, S., Moussian, B., et al. (2016). Helicoidal organization of chitin in the cuticle of the migratory locust requires the function of the chitin Deacetylase2 enzyme (LmCDA2). J. Biol. Chem. *291*, 24352–24363.

Yu, R.R., Liu, W.M., Zhao, X.M., Zhang, M., Li, D.Q., Zuber, R., Ma, E.B., Zhu, K.Y., Moussian, B., and Zhang, J.Z. (2018). LmCDA1 organizes the cuticle by chitin deacetylation in Locusta migratoria. Insect Mol. Biol. *28*, 301–312.

Zhao, H., Peng, Z., Du, Y., Xu, K., Guo, L., Yang, S., Ma, W., and Jiang, Y. (2018). Comparative antennal transcriptome of Apis cerana cerana from four developmental stages. Gene *660*, 102–108.

Zhou, S.-S., Sun, Z., Ma, W., Chen, W., and Wang, M.-Q. (2014). De novo analysis of the Nilaparvata lugens (Stål) antenna transcriptome and expression patterns of olfactory genes. Comp. Biochem. Physiol. Part D Genomics Proteomics *9*, 31–39.

Zhou, Y., Badgett, M.J., Billard, L., Bowen, J.H., Orlando, R., and Willis, J.H. (2017). Properties of the cuticular proteins of Anopheles gambiae as revealed by serial extraction of adults. PLoS One *12*, e0175423.

Zhou, Y., Badgett, M.J., Bowen, J.H., Vannini, L., Orlando, R., and Willis, J.H. (2016). Distribution of cuticular proteins in different structures of adult Anopheles gambiae. Insect Biochem. Mol. Biol. *75*, 45–57.

iScience, Volume *23*

## Supplemental Information

# Insect Mouthpart Transcriptome Unveils Extension of Cuticular Protein Repertoire and Complex Organization

**Natalia Guschinskaya, Denis Ressnikoff, Karim Arafah, Sébastien Voisin, Philippe Bulet, Marilyne Uzest, and Yvan Rahbé**

# Supplemental Information

## Supplemental data

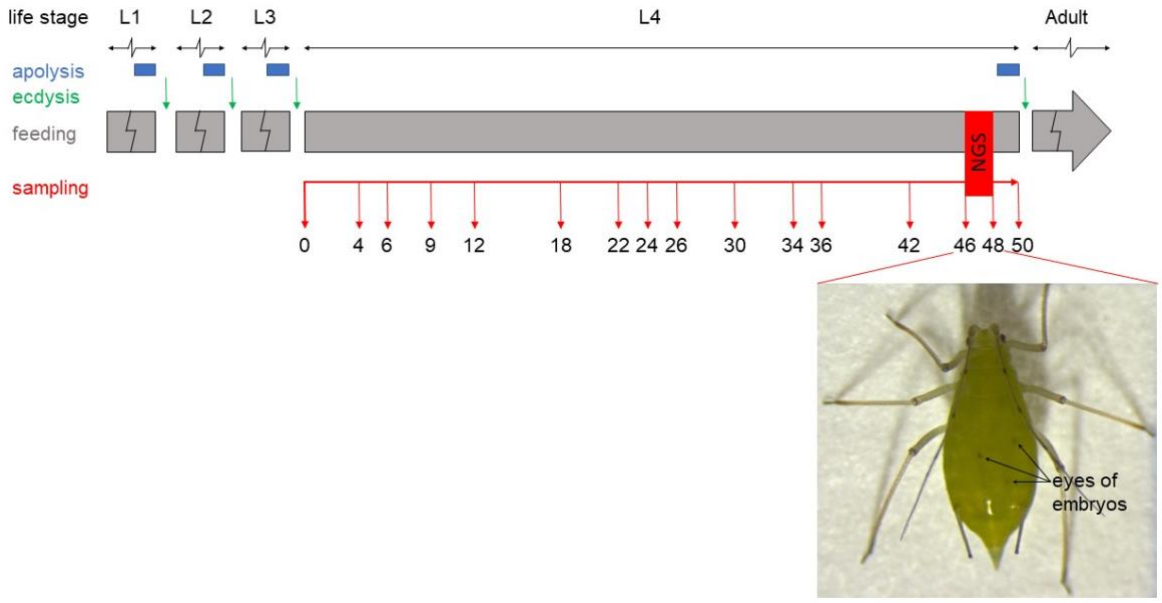



**Figure S1. Schematic representation of life cycle of *A. pisum*.** This diagram illustrates the place of apolysis (the detachment of the old cuticle from the underlying epidermis; shown as blue rectangle), ecdysis (the shedding of the old cuticle; shown with green arrow). Grey arrow represents the life stage (L1, L2, L3, L4 nymphal instars and adult). Red arrow explains the sampling for RT-qPCR analysis and Next-Generation Sequencing (NGS). RNA was extracted from wholes bodies or dissected heads of aphids at different times during L4 nymphal instar (at 0, 4, 6, 9, 12, 18, 22, 24, 26, 30, 34, 36, 42, 46, 48 and 50 (adult) hours from the beginning of L4 instar). For NGS experiments, the retort organs (maxillary, Mxr, and mandibular, Mdr) were dissected from the L4 nymphal instar individuals with visible embryo's eyes (photo of L4-46 individual, three embryo's eyes are visible), this time point correspond to 46-48 hours from the beginning of L4 instar (red rectangle). *Related to qPCR and NGS sampling protocol (Figure 1).*

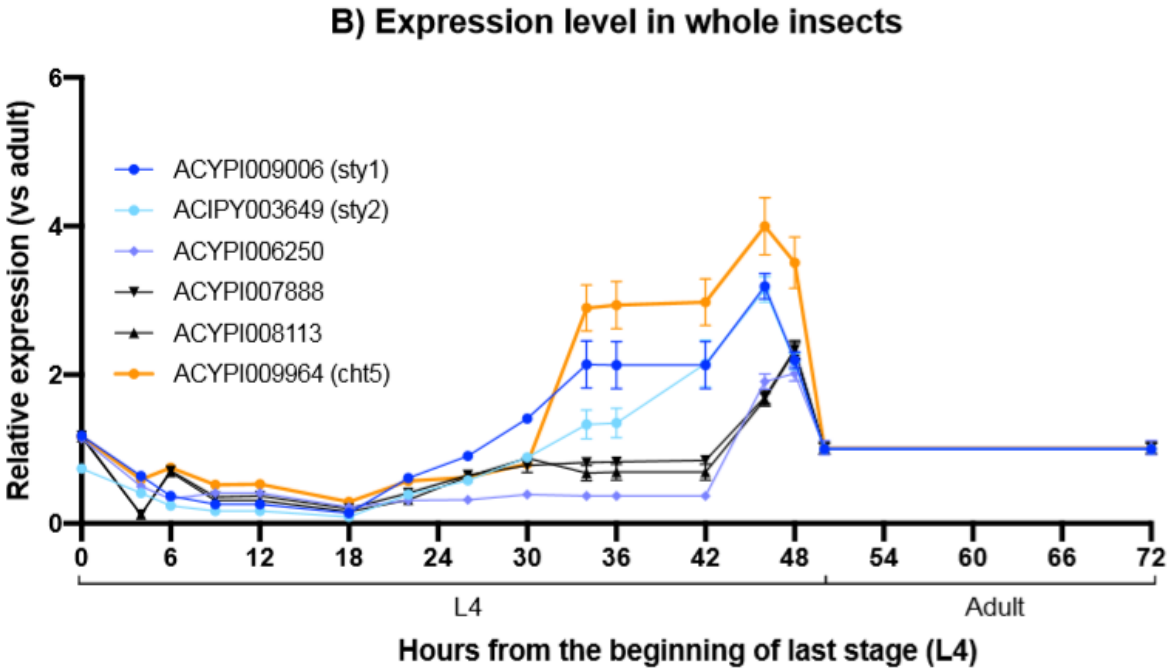



**Figure S2. mRNA expression profile of 6 genes of interest during aphid development.**

5 genes coding for cuticular proteins, among which two already identified in stylets (ACYPI009006, Stylin-01, ACYPI003649, Stylin-02); and (ACYPI006250, ACYPI007888, ACYPI008113), were used in qRT-PCR analysis together with the gene ACYPI009964, coding for an imaginal moult-associated chitinase Cht5 (J9KAI2). Gene-expression levels at different life stages are expressed relative to adult aphid. The EF1a and actin genes were used for data normalization. Results are reported as means ± SD; n=3 independent biological replicates per stage. Data were analyzed by one-way ANOVA followed by a post hoc multiple comparisons test (Tukey's HSD test). B) Expression level in whole aphids was analyzed using individuals from 9 independent biological replicates per stage from 3 biological batches. *Related to Figure 1.*

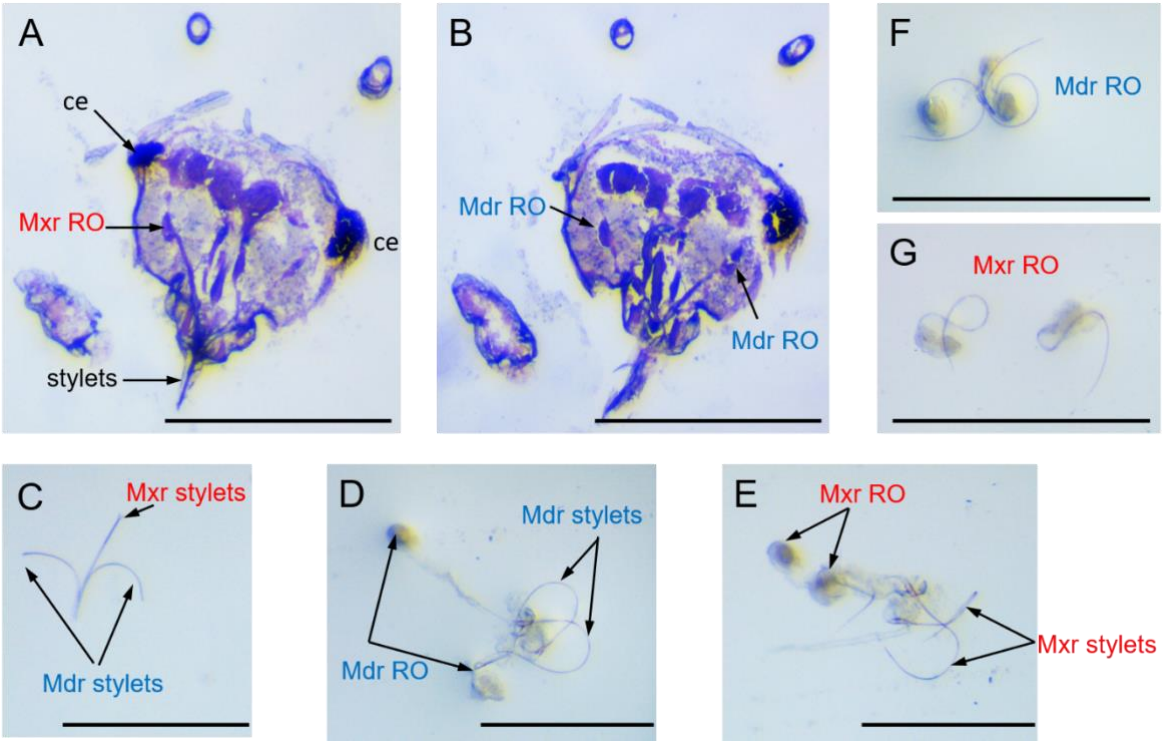



**Figure S3. Imaging of retort organs of *A.pisum*.**

15 µm thickness sections obtained after cryotome cutting of aphid heads and stained with May Grundvald-haematoxylin permitted to observe maxillary (A, Mxr RO) and mandibular retort organs (B, Mdr RO). Scalebar=500µm. The properties of maxillary and mandibular stylets help to distinguish the retort organs during direct dissection of retort organs. C) stylet bundle (two Mdr stylets curve in the same plane, and two Mxr stylets stay together). D) Dissected pair of mandibular retort organs (adult mandibular stylets curve in the same plane). E) Dissected pair of maxillary retort organs (adult maxillary stylets obtain helical form). F) Dissected pair of mandibular retort organs with new-formed stylets (new-formed mandibular stylets curve in the same plane). G) Dissected pair of maxillary retort organs with new-formed stylets (new-formed maxillary stylets obtain helical form). RO, retort organs; Mxr, maxillary; Mdr, mandibular, ce, compound eye. *Related to NGS sampling protocol and gland imaging (Fig. S1 and Fig. 2).*

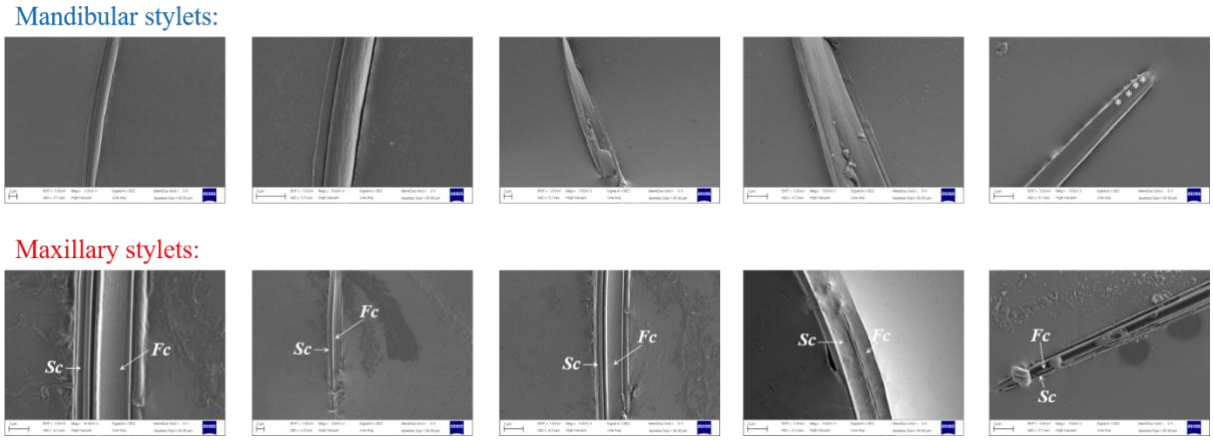



**Figure S4. Control of proper dissection of retort organs by Scanning Electron Microscopy (SEM).** Retort organs were dissected as usually, the form of old stylets helped to distinguish maxillary and mandibular retort organs (see Figure S3). Then, the samples were analysed by SEM to confirm the proper dissection (n=10 for maxillary and mandibular retort organs). Mandibular stylets could be distinguished by their surface and serrated edges (*) on the tip. Maxillary stylets differ by the presence of salivary (Sc) and food (Fc) canals. *Related to NGS experiment sampling (Fig.3)*

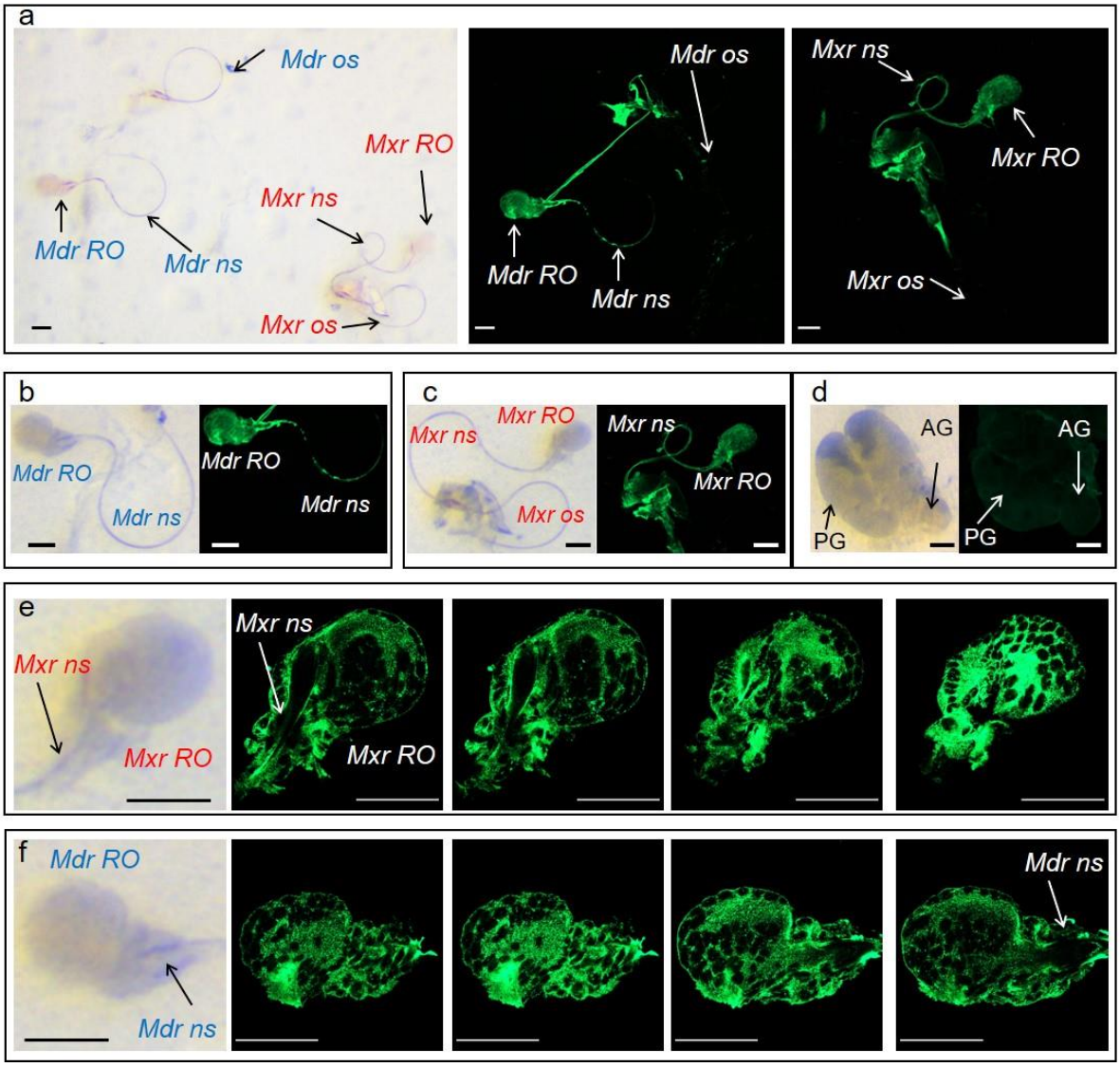



**Figure S5.** ***In vitro*** **immunohistological analysis of mandibular and maxillary retort organs and salivary glands.** The mandibular (Mdr RO) and maxillary (Mxr RO) retort organs and salivary glands (SG) were dissected from L4-48h individuals. The antibodies @1.07 raised against "SQEQEVNFDGNFKNK" peptide present in ACYPI003649, Stylin-02 were used for immunohistological analysis. a) Mandibular and maxillary retort organs; dissected and immunolabelled. b) Mandibular retort organs; dissected (left panel) and immunolabelled (right panel). c) Maxillary retort organs; dissected (left panel) and immunolabelled (right panel). d) salivary glands; dissected (left panel) and immunolabelled (right panel). Heterogeneity of labelling in maxillary (e) and mandibular (f) retort organs. Mdr, mandibular; Mxr, maxillary; RO, retort organ; ns, new stylet; os, old stylet; PG, principal salivary gland; AG, accessory salivary glands. *Related to Fig.2.*

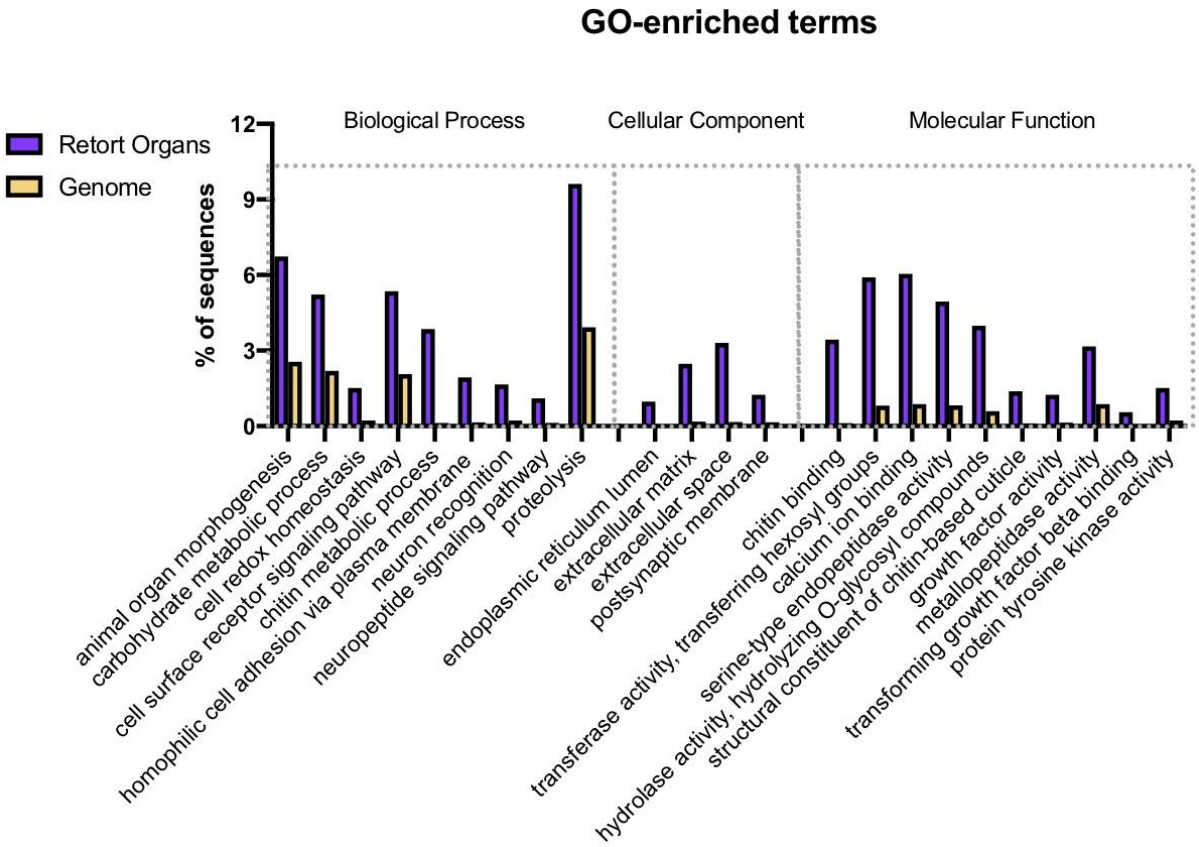



**Figure S6. Gene Ontology term enrichment analysis of genes identified as expressed in maxillary and mandibular retort glands (with FPKM>0,5 in each replicas) and predicted to be secreted by SignalP v4.1.**

The figure shows on the x-axis the most specific GO-terms significantly enriched (FDR<0,001) comparing to A. pisum total annotated unigenes. The GO-terms are presented for three categories, 'Biological Process', 'Cellular Component' and 'Molecular Function'. For each GO-term, the percentages of sequences in test set (genes identified as expressed predicted to be secreted in retort organs) and reference set (total annotated unigenes in *A.pisum*) are reported. *Related to NGS data analysis, Figure 3 and Tables S2 and S3.*

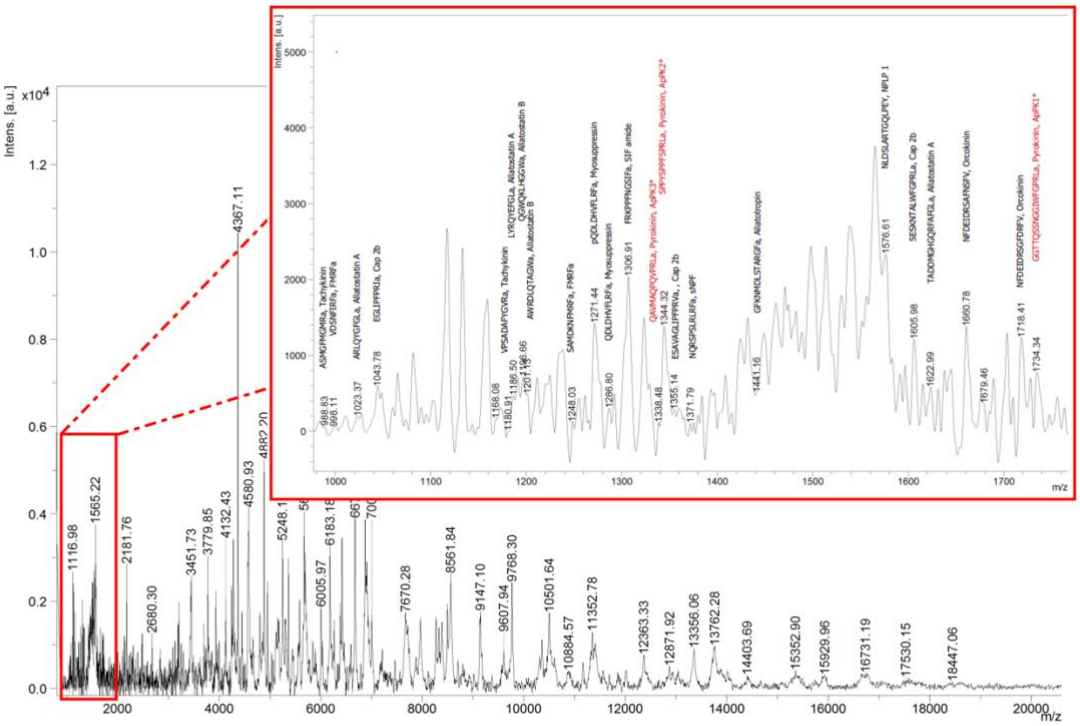



**Figure S7. Representative MALDI mass spectrum of a central nervous tissue sample of *Acyrthosiphon pisum* (crude sample).**

The main spectrum is acquired in a linear positive mode in the m/z range of 1 to 20 kDa (low molecular mass). The enlarged panel shows the ions obtained between m/z 900 and 1,750. All peptide sequences indicated on peak tops (associated to m/z values) refer to the sequences predicted by Huybrechts and colleagues (2010, Table 1, page 89) for which the theoretical average mass could be assigned to an ion detected in our conditions (linear mode). The amino acid sequences in red (enlarged panel) correspond to the three predicted sequence of the pyrokinin according to us and Huybrechts and colleagues. The three pyrokinin peptides were named ApPK1, ApPK2 and ApPK3 for GGTQSSNGGIWFGPRLa (m/z 1,734.34), SPPYSPPFSPRLa (m/z 1,344.32) and (p)QAVMAQPQVPRLa m/z 1,338.48, respectively. In the displayed sequences, ‘a’ corresponds to a C-terminal amidation and ‘pQ’ to a N-terminal pyroglutamic acid. *Related to Fig.4.*

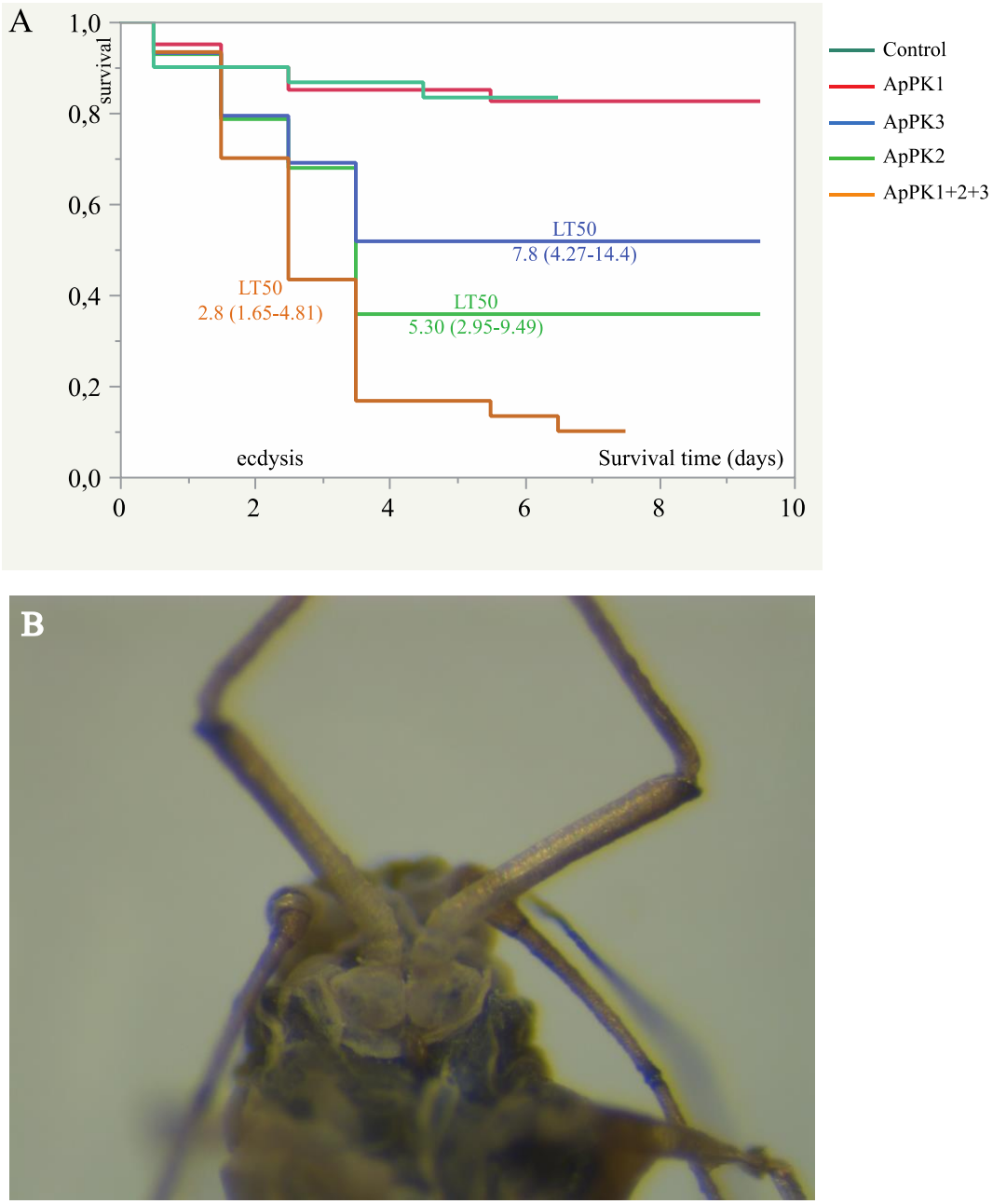



**Figure S8: Pyrokinin injection supplemental experiment**

**A**) Aphid survival curves after the injection of individual pyrokinin peptides, before ecdysis and compared to the equimolar bouquet at same individual molarity (2.5 pmol per peptide/23nL/insect); Lethal times fifty (LT50) for all active injections (3,2,1+2+3) are labelling corresponding curves, and include confidence intervals as calculated for a log-normal fit (n≈30). Approximate observed ecdysis time slot labels x-axis (differs slightly from Fig.4).

**B**) Detailed view of headless phenotype.

*Related to Fig.4.*

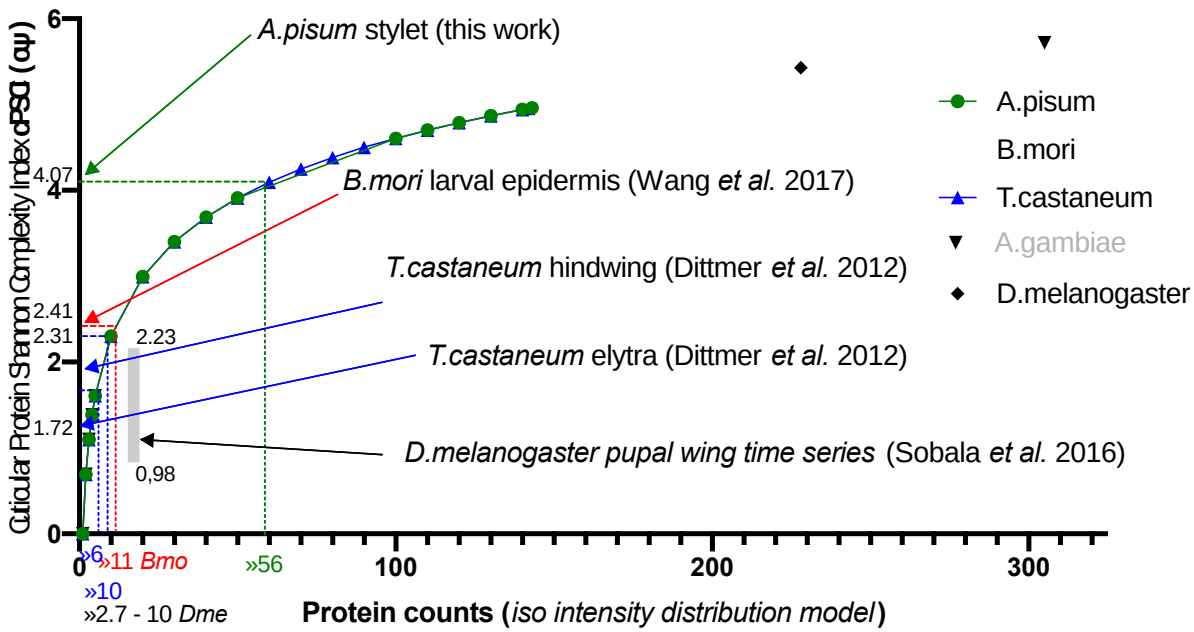



**Figure S9. Plot of Shannon Cuticular protein complexity index (cPSCI) versus equivalent protein counts under iso-distribution model for the four published data of whole-genome transcriptional analyses of cuticle producing tissues (*Bombyx* epidermis, *Tribolium* wings and *Acyrthosiphon* stylets).**

cPSCI index computed for the various cuticles are plotted on y-axis, and corresponding protein counts inferred on the x-axis, and showing the higher complexity of aphid stylets as compared to the other polymer alloys. *Related to Fig.5.*

# Transparent methods

## Contact for reagent and resource sharing

Yvan.Rahbe@inra.fr, and Marilyne.Uzest@inra.fr for antibodies

## Experimental model and subject details

### Insects and insect rearing

A long-established parthenogenetic clone (LL01) of *A. pisum* Harris was maintained on young broad bean plants (*Vicia faba* L. cv. *Aguadulce*) at 21°C, with a photoperiod of 16 h light/8 h dark. For synchronized aphid production, wingless adults were put on seedlings, allowed to produce nymphs, and removed after 4 h. These broadly synchronized aphids were sampled at different stages, but a similar 2 h-synchronization protocol was used for precise L4-cohort production when needed: L3 nymphs, easily distinguishable by caudal morphology (Digilio, 1995), were specifically isolated and screened every 2 h for L4 emergence.

## Method details

### Sampling for analysis of mRNA expression of aphid genes

Aphids were collected at 16 time points within the L4 nymphal instar: at 0, 4, 6, 9, 12, 18, 22, 24, 26, 30, 34, 36, 42, 46 and 48 h from the beginning of the L4 instar and immediately after molting to adult. Whole bodies or dissected heads were then placed in RNAlater® solution (ThermoFisher Scientific, Waltham, MA, USA) and stored at −80°C until qRT-PCR analyses were performed. For each time point, three independent biological replicates were processed.

### Fixation and clearing of aphids for CLSM

Aphids were fixed overnight in 4% paraformaldehyde (PFA)-PBS solution at 4°C. After fixation, the samples were washed 3 times for 10 min each in phosphate buffered saline (PBS). The aphids were then transferred into aqueous 35% $H_2O_2$ for one week until the cuticle turned light-yellow. After washing again 3 times for 10 min each in PBS, the samples were incubated in graded ethanol solutions (50%, 70%, 90%, 95% and 100%; 10 min each). The samples were finally transferred to methyl salicylate (M6752, Sigma-Aldrich) for clearing (Smolla et al. 2014).

### In situ confocal laser scanning microscopy

For *in situ* confocal laser scanning microscopy (488 nm argon laser source), L4 nymphal instars were placed in scanning chambers with methyl salicylate. The aphids were examined using an LSM 800 (Zeiss, Germany) confocal laser scanning microscope with a 10x dry objective (0.40 NA) and 40s and 100x oil immersion objectives (0.7 and 0.95 NA). Optical sections of whole heads were obtained at z-steps ranging from 1 to 5 µm resulting in an x-y plane at 1024x1024 pixel resolution.

### 3D reconstruction

Three-dimensional surface-rendered models were created from optical image stacks using Imaris 7.5 (Bitplane Inc.). The retort glands, salivary glands, digestive tract and partial inner stylets were manually segmented.

### Real-time quantitative RT-PCR

Total RNA was extracted from whole aphids or dissected tissues using the RNeasy Mini Kit. The RNA was treated with DNase I (Promega, Madison, WI, USA), and first-strand cDNA was synthesized using the Sensiscript RT Kit (Qiagen, Hilden, Germany) with oligo(dT) primers (ThermoFischer Scientific, Waltham, MA, USA). Real-time RT-PCR reactions were performed on a LightCycler® 480 instrument (Roche, Basel, Switzerland) using 1:2.5 diluted cDNAs and SYBR Green PCR Master Mix, according to the manufacturer's instructions. The mRNA levels of genes of interest were quantified relative to two constitutively expressed genes, EF1-α (*ACYPI004434*)

and actin (*ACYPI000064*). These two genes were chosen as the best normalization genes according to Pan *et al*. (Pan et al. 2015). The primers used in this study are listed in SI Table 4. All measurements on biological replicates were performed in technical triplicate. The relative expression was calculated taking into account the real-time PCR efficiency of each gene (E) and the crossing point difference (ΔCP) of a test condition compared to the reference condition and expressed in comparison to the normalization genes (actin and EF1-α) using the following equation (Pfaffl, 2001):

$$R = \frac{(E_{target})^{\Delta \mathrm{CP}\ target\ (\mathrm{control-sample})}}{(E_{reference})^{\Delta \mathrm{CP}\ reference\ (\mathrm{control-sample})}}$$

### Dissection of retort glands

Maxillary and mandibular glands were dissected from *A.pisum* individuals on the L4-embryo stage (L4_e, 46-48 hours from start of L4). The insects were immobilized on glass slides, using crystal adhesive tape to allow access to rostrum, stylets and labium. Dissections were performed under binocular microscope in RCL2 solution (Masir et al., 2012) to preserve the anatomic integrity of retort glands, and avoid the cross-contamination between both type of retort glands, salivary glands and gut. For every dissection, the retort organ type (*mxr* / *mdr*) was carefully analyzed by the shape of the old/pharate stylet (Fig. S3).

### Immunolabeling of ROs

The ROs and salivary glands of nymphal *A. pisum* were dissected from the L4_e stage under a dissecting binocular microscope using thin tweezers and needles. The dissected organs were immobilized on glass slides by double-faced adhesive, washed with Tris-buffered saline TS (50 mM Tris pH 7.4, 150 mM NaCl) and blocked with 2% skim milk. After three washing steps with TS buffer, the samples were incubated overnight at 4°C with the polyclonal primary antibody anti-1-07 already described (Webster et al., 2018). Then, the samples were washed three times with TS buffer and finally incubated with secondary Alexa Fluor 488-conjugated anti-rabbit IgG for 2 h at room temperature. After three washing steps in TS buffer, the samples were covered and observed with a Zeiss LSM 800 confocal microscope. For each of the antibodies used, 5 dissected individual glands were labelled (*mxr* RO, *mdr* RO and SG, as well as gut, and mock cutting slide). Antibodies used were @Ap 1.04, @Ap 1.03, @Ap 1.07, @Mp 1.01, @Mp1.11 and @pepL (Uzest Lab), but only @Ap 1.07 labels are shown in this work (Fig. 2 and S5).

### RNA extraction, cDNA library construction, and Illumina sequencing

Total RNA was isolated from 40 dissected ROs, with the maxillary (*mxr*) and mandibular (*mdr*) separated, from L4 nymphal instars collected between 46-48 h from the beginning of the L4 instar (L4_e). The dissected organs were collected in RLT buffer (Qiagen lysis buffer) supplemented with 1% β-mercaptoethanol and frozen until RNA extraction (-80°C). RNA extraction was performed using the ProfileXpert platform (Lyon, France) with the RNeasy Mini Kit (Qiagen, Hilden, Germany). The RNA quantity and quality were assessed by an Agilent 2100 Bioanalyzer. Specific preamplification of mRNA was performed with the SMART-Seq v4 Ultra Low Input RNA (Clontech) Kit from 4 ng of total RNA. The cDNA quantity and quality were assessed by an Agilent 2100 Bioanalyzer. Double indexes were added to 1 ng of each of the 6 libraries of cDNAds (3 *mxr* and 3 *mdr*) with a Nextera XT kit (Illumina, USA). The quality of the cDNA library constructed was tested using an Agilent 2100 Bioanalyzer. The sequencing was performed using an Illumina NextSeq500 system (Illumina, USA) with single reads of 76 bp. Runs yielded 35,469 Mb of data with more than 5,300 Mb per library and Qscores (%≥Q30) of more than 93.5 per library (Mean quality score >34.2). Pre-analysis quality control of delivered libraries was run

with FastQC (0.11.5) under Galaxy with all critical tests passed (per base sequence end tile quality, per sequence quality scores, per base N content, no overrepresented sequences nor adapter present).

### Mass spectrometry methods

For sample preparation, RapiGest SF surfactant was purchased from Waters (Milford, MA) and ammonium bicarbonate ($NH_4HCO_3$), Hexofluoroisopopanol (HFIP), iodoacétamide (IAA), dithiothreitol (DTT) were reagent grade from Sigma-Aldrich (St. Louis, MO). For LC-MS/MS, MilliQ water (Merck Millipore, Billerica, MA) was used and LC- MS-grade formic acid (FA) was obtained from Fluka. Acetonitrile (ACN)and tri- fluoroacetic acid (TFA) were of HPLC grade or higher and obtained from Carlo Erba (Val de Reuil, France). For protein digestion, sequencing- grade modified trypsin (Promega, Madison, WI) was used and PBS buffer was purchased from Thermo Fisher Scientific (Waltham, MA). All other sample procedures were published (Masson et al. 2018) and are summarized as follows; Briefly, all cuticular structures were washed twice with PBS buffer in an ultrasonic bath (Branson, Danbury, CT) for 15 minutes to remove the internal tissues. Cuticles were intact after treatment; glands were washed once after dissection and not sonicated. Samples were then pelleted by a quick centrifugation step, and dried by centrifugation under vacuum (Labconco, Kansas City, MO). Dried samples were incubated in pure HFIP with a sufficient volume to cover the biological matrices for 4h at 4°C under gentle shaking. HFIP was evaporated by centrifugation under vacuum, and samples were incubated overnight at 4°C in ammonium bicarbonate buffer (50 mM, pH 7.8) supplemented with 0.1% RapiGest™ surfactant. Proteins were reduced with 10 mM DTT in 50 mM NH4HCO3 for 1h in the dark, at 50°C, prior to alkylation with 30 mM IAA in 50 mM NH4HCO3 for 30 minutes at 50°C. Digestion was carried out overnight at 37°C by addition of 0.5 µg of trypsin prepared in 50 mM NH4HCO3. To stop proteolysis and cleave the RapiGest™ surfactant, samples were acidified with TFA to obtain a 0.5% concentration and incubated for 30 minutes at 37°C. Finally, samples were dried by centrifugation under vacuum and dry pellets re-suspended in 2% ACN/0.1% TFA final concentration prior to MS analysis.

Maldi-ToF analyses were performed on a AutoFlexTM III MALDI TOF/TOF mass spectrometer equipped with a SmartBeam IITM laser (Bruker Daltonik). The molecular mass fingerprints were acquired on raw organs dissected in RCL2 solution under the microscope to avoid the cross-contaminations with other tissues. Maxillary and mandibular retort gland, as well as part of the nervous system were put on a MALDI target plate (MALDI MTP 384 polished Ground Steel plate from Bruker Daltonik), and resuspended in 1µL of 4-HCCA (10 mg/ml prepared in 0,5% of trifluoroacetic acid mixed by volume, Sigma Aldrich, France, and 35% of acetonitrile). MALDI MS spectra were recorded in linear positive mode and in automatic data acquisitions using FlexControl 4.0 software (Bruker Daltonik) using 1.5kV as difference of potential. An external calibration of the mass spectrometer was performed using a standard mixture of peptides and proteins (Peptide Standard Calibration II and Protein Standard Calibration I, Bruker Daltonik) covering the dynamic range of analysis. Nano-LC-MS/MS analyses were carried out on an Ultimate 3000 nano-HPLC, coupled with a Q-Exactive Orbitrap high- resolution mass spectrometer (Thermo Fisher Scientific, MA). The Q-Exactive mass spectrometer, equipped with a nanospray ion source, was used in positive mode and data- dependent acquisition. The software Chromeleon Xpress and Xcalibur 2.2 were used to control the HPLC and the mass spectrometer, respectively. Full procedures were as described in Masson et al. (Masson et al., 2018).

### Injection of PKs

Equimolar PK mixtures, or individual peptides, were injected at 23 nL/insect and 2.5 pmol per peptide/insect through a nano-Inject II device (Drummond, Broomall PA USA). Needles were forged and beveled at less than 10

µm external terminal diameter with a Sutter P1000 µcapillary puller (http://dtamb.univ-lyon1.fr). Injections were performed through the anterior joint of the first abdominal tergite in L4_e aphids (or adults in the additional postecdysis control injection). Survival analysis following this injection experiment was scored on individual faba bean plantlets, with a mean repeat number of n=30 (28-40 depending on insect availabilities).

## Quantification and statistical analysis

### Statistical analyses

All statistical analyses, except for NGS, were carried out using the JMP Pro 14SW® software (SAS Institute, Cary NC USA) with values of $p<0.05$ considered significant. Multiple comparisons using the Tukey-Kramer HSD test were used for most non-NGS applications.

### RNA-Seq data and differential analysis of transcriptomic libraries

All whole-genome analyses were run under a Galaxy V0.2 (Giardine et al., 2005) instance installed at the INRA BIPAA platform (http://bipaa.genouest.org) or with the B2G client for MacOS (https://www.blast2go.com). Quality score analysis and filtering were conducted using the FASTQ tool. The list of RO-expressed genes was identified as FPKM>0.5, generating a list of 10587 genes. When using published salivary gland libraries as external controls (Boulain et al., 2018) , the same parameters was used to output the SG expressed gene list, containing 15435 genes. To identify the differentially expressed genes and calculate rpkm, EdgeR software was used with a minimum CPM value of 0.5 for gene selection, TMM normalization and FDR multitest correction (AskoR DE EdgeR for AskOmics, Galaxy V0.2). DESeq2 was also used at the initial steps of analysis, and served for selecting the two *mxr*/*mdr* differential genes (adjusted p-values <0.05; 0.01 threshold gave the same result)). All results are included in the SI material Table S3.

### CP complexity index

For the sake of comparison with other whole-genome transcriptomic analyses of CPs (microarrays or NGS), we calculated a "species" complexity index, where the species is an individual protein within the set of predicted protein constituents of the cuticle, defined as a polymer assemblage. We selected Shannon's entropy index for its generality, its mathematical properties and its widespread use in ecology for measuring species diversity: $-\sum_{n=1}^{N}(p_i \,.\, \ln(p_i))$, with N being the total number of CPs in a given arthropod genome (or whole-genome experiment) and $p_i$ the proportion of *species* belonging to the *i*th type in the quantitative analysis (*i.e.*, part of *i*th protein in the expression profile, such as that of Fig. 3). N=143 for our aphid work (excluding the newly identified CPs) and may be computed for any given genome (Ioannidou et al. 2014). Computing the cPSCI, or **cψ** (CP Shannon complexity index), is easy, and computing a valuable proxy of this measure is also straightforward: $e_{c\psi}$ is a measure of the *effective number of proteins* in a cuticle, *i.e.*, the number of individual proteins, under isodistribution law, that would give the same complexity observed in the experiment. Fig. S9 gives an illustration of these two variables for our work and for the three other full genome datasets available in the current literature.

### New cuticular family detection

The rationale for the detection of unidentified (low-complexity) CPs was as follows and was first run on a test set of the 200 most expressed proteins in RO (including known CPs) to adjust the thresholds: The pea aphid official gene set proteome (OGS v2.1b; http://bipaa.genouest.org/sp/acyrthosiphon_pisum) was downloaded (fasta) and submitted to InterProScan to compute all identifiable protein domains and their coordinates and to identify all proteins with a known protein domain. All proteins with no identifiable domain were then retrieved, analyzed using SignalP v4.1, trimmed to remove signal peptides, and then their amino-acid composition table was computed

(JMP v14 script). The 143 identified *A. pisum* CPs were added to the set (trimmed to remove both their signal peptides and their chitin-binding domains) to serve as a compositional template. A PCA on the amino-acid composition of the 200 RO most expressed proteins and the whole CuP test set (Figure 4A) allowed us to group cognate CPs with groups of unknown proteins to define a first set/list of (19) candidate CPs. The whole aphid proteome was then computed for mature peptide amino-acid composition and plotted onto the first five axes of the previous PCA. Then, a selection scheme for i) predicted secreted proteins (presence of signal peptide by SignaP 4.1) *and* ii) unannotated proteins (no identified domain by InterProScan) *and* iii) a compositional bias analogous to the test set selection (PCA plot with a negative axis one coordinate) allowed us to define a group of putative CPs that was checked manually against the other PCA axes to determine whether they fell within the compositional biases exhibited by cognate CPs and analyzed for their expression pattern in the RO (see results).

## Data and software availability

### Data availability

The sequence data have been submitted to the Genbank SRA archive with accession number PRJNA510301. Serial images of aphid heads/ROs (Fig. 2B, D-E) and confocal stacks of the same (Fig. 2B) are available as supplementary data at data.inra.fr (Two datasets: CLSM : https://doi.org/10.15454/VPUGUG, Histology : https://doi.org/10.15454/QUM1LQ).

### Software availability

NA

### Additional resources

NA

## Supplemental references


Digilio M.C. (1995) Rapid identification of the nymphal stages of *Acyrthosiphon pisum* (Harris) (Homoptera Aphidoidea). [Identificazione rapida degli stadi giovanili di *Acyrthosiphon pisum* (Harris) (Homoptera, Aphidoidea).]. Bollettino di Zoologia Agraria e di Bachicoltura 27(1):111-116.

Smolla M, Ruchty M, Nagel M, & Kleineidam CJ (2014) Clearing pigmented insect cuticle to investigate small insects' organs in situ using confocal laser-scanning microscopy (CLSM). Arthropod Structure & Development 43(2):175-181.

Pan H., Yang X., Siegfried B.D., and Zhou X. (2015) A Comprehensive Selection of Reference Genes for RT-qPCR Analysis in a Predatory Lady Beetle, *Hippodamia convergens* (Coleoptera: Coccinellidae). PLOS ONE 10(4).

Pfaffl M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic acids research 29(9):e45.

Masir N., et al. (2012) RCL2, a potential formalin substitute for tissue fixation in routine pathological specimens. Histopathology 60(5):804-815.

Webster C.G., et al. (2018) Identification of plant virus receptor candidates in the stylets of their aphid vectors. J Virol.

Masson V., Arafah K., Voisin S., and Bulet P. (2018) Comparative Proteomics Studies of Insect Cuticle by Tandem Mass Spectrometry: Application of a Novel Proteomics Approach to the Pea Aphid Cuticular Proteins. Proteomics 18(3-4).

Giardine B., et al. (2005) Galaxy: a platform for interactive large-scale genome analysis. Genome Res 15(10):1451-1455.

Boulain H., et al. (2018) Fast Evolution and Lineage-Specific Gene Family Expansions of Aphid Salivary Effectors Driven by Interactions with Host-Plants. Genome Biology and Evolution 10(6):1554-1572.

Ioannidou Z.S., Theodoropoulou M.C., Papandreou N.C., Willis J.H., and Hamodrakas S.J. (2014) CutProtFam-Pred: Detection and classification of putative structural cuticular proteins from sequence alone, based on profile Hidden Markov Models. Insect Biochem. Mol. Biol. 52:51-59.