



HAL
open science

Evolutionary diversification of insulin-related peptides (IRPs) in aphids and spatiotemporal distribution in *Acyrtosiphon pisum*

C. Huygens, Mélanie Ribeiro Lopes, Karen Gaget, G. Duport, S. Peignier, S. de Groef, Nicolas Parisot, Federica Calevro, P. Callaerts

► **To cite this version:**

C. Huygens, Mélanie Ribeiro Lopes, Karen Gaget, G. Duport, S. Peignier, et al.. Evolutionary diversification of insulin-related peptides (IRPs) in aphids and spatiotemporal distribution in *Acyrtosiphon pisum*. *Insect Biochemistry and Molecular Biology*, In press, 141, pp.103670. 10.1016/j.ibmb.2021.103670 . hal-03507324

HAL Id: hal-03507324

<https://hal.science/hal-03507324v1>

Submitted on 22 Jul 2024

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 - NonCommercial 4.0 International License

1 **Evolutionary diversification of insulin-related peptides (IRPs) in aphids and spatiotemporal**
2 **distribution in *Acyrtosiphon pisum***

3

4 Huygens C.^{1,2*}, Ribeiro Lopes M.^{2*}, Gaget K.^{2*}, Duport G.², Peignier S.², De Groef S.¹, Parisot N.²,
5 Calevro F.², Callaerts P.¹

6

7 Affiliation:

8 1. Laboratory of Behavioral and Developmental Genetics, Department of Human Genetics,
9 KULeuven, University of Leuven, B-3000 Leuven, Belgium

10 2. Univ Lyon, INSA Lyon, INRAE, BF2I, UMR 203, 69621 Villeurbanne, France

11

12 * Joint first authors

13

14 Authors for correspondence:

15

16 Federica Calevro: federica.calevro@insa-lyon.fr

17 Patrick Callaerts: patrick.callaerts@kuleuven.be

18 **Abstract**

19 Members of the insulin superfamily activate the evolutionarily highly conserved insulin/insulin-like
20 growth factor signaling pathway, involved in regulation of growth, energy homeostasis, and
21 longevity. In the current study we focus on aphids to gain more insight into the evolution of the IRPs
22 and how they may contribute to regulation of the insulin-signaling pathway. Using the latest
23 annotation of the pea aphid (*Acyrtosiphon pisum*) genome, and combining sequence alignments and
24 phylogenetic analyses, we identified seven putative IRP encoding-genes, with IRP1-IRP4 resembling
25 the classical insulin and insulin-like protein structures, and IRP5 and IRP6 bearing insulin-like growth
26 factor (IGF) features. We also identified IRP11 as a new and structurally divergent IRP present in at
27 least eight aphid genomes. Globally the ten aphid genomes analyzed in this work contain four to 15
28 IRPs, and only three IRPs were found in the genome of the grape phylloxera, a hemipteran insect
29 representing an earlier evolutionary branch of the aphid group. Expression analyses revealed spatial
30 and temporal variation in the expression patterns of the different *A. pisum* IRPs. IRP1 and IRP4 are
31 expressed throughout all developmental stages and morphs in neuroendocrine cells of the brain, while
32 IRP5 and IRP6 are expressed in the fat body. IRP2 is expressed in specific cells of the gut in aphids in
33 non-crowded conditions and in the head of aphids under crowded conditions, IRP3 in salivary glands,
34 and both IRP2 and IRP3 in the male morph. IRP11 expression is enriched in the carcass. This
35 complex spatiotemporal expression pattern suggests functional diversification of the IRPs.

36 **Keywords:** Hemiptera, aphids, *Acyrtosiphon pisum*, insulin, IRPs

37 **1. Introduction**

38 The insulin-signaling pathway is an evolutionarily highly conserved signal transduction pathway
39 present in all metazoans, with a central role in regulating metabolic homeostasis, growth,
40 reproduction, development, lifespan and aging (Lodish et al., 2016; Raven et al., 2015). The
41 prototypical ligand, insulin, is a peptide hormone, synthesized as a pre-proinsulin polypeptide that
42 undergoes several proteolytic modifications to proinsulin and subsequently insulin (Hancock, 2010).
43 It consists of A and B chains that are connected by two disulfide bridges between cysteine residues
44 and a third disulfide bridge between two internal cysteine residues of the A chain (Hancock, 2010).
45 Insulin is only one of the members of the insulin superfamily, which also includes insulin-related
46 peptides (IRPs), relaxins and insulin-like growth factors (IGFs). In humans, this family comprises ten
47 members, *i.e.* one insulin, four IRPs, three relaxins, and two IGFs (Hancock, 2010). The number of
48 peptides in the insulin superfamily can vary widely between species from for example three in *Hydra*
49 *magnipapillata* (Bridge et al., 2010) to 40 in *Caenorhabditis elegans* (Zheng et al., 2018).

50 In insects, the insulin signaling pathway has been extensively studied in the genetic model organism
51 *Drosophila melanogaster* where it acts in metabolic homeostasis, and in the regulation of growth,
52 reproduction, development, and lifespan (Altstein and Nässel, 2010; Brogiolo et al., 2001; Nässel and
53 Broeck, 2016). The *Drosophila* genome encodes eight so-called insulin-like peptides (DILP1-DILP8)
54 (Nässel and Vanden Broeck, 2016). While DILP1-DILP5 are *bona fide* IRPs, DILP6 is IGF-like
55 structurally and functionally (Okamoto et al., 2009; Slaidina et al., 2009), and DILP7 and DILP8 have
56 been proposed to be relaxin-like (Colombani et al., 2012; Garelli et al., 2012; Miguel-Aliaga et al.,
57 2008; Yang et al., 2008). DILP1-DILP7 bind the insulin receptor (InR), a receptor tyrosine kinase, to
58 activate the evolutionarily conserved insulin signaling pathway. DILP8 on the other hand acts via the
59 relaxin-like Lgr3 receptor (Colombani et al., 2015). Work from many groups has revealed that there is
60 considerable functional diversification among DILPs (Bai et al., 2012; Brogiolo et al., 2001; Kannan
61 and Fridell, 2013; Veenstra et al., 2008). However, this remains incompletely understood. Part of the
62 specificity probably arises from their spatiotemporal expression pattern. For instance, DILP1-3, and
63 DILP5 are produced in a specific set of neurosecretory cells of the brain, also known as insulin-
64 producing cells (IPCs) (Brogiolo et al., 2001). Furthermore, DILP3 expression is also observed in the
65 intestinal muscle cells (Veenstra et al., 2008). In contrast, expression of DILP6 is only detected in the
66 fat body (Bai et al., 2012). In addition to the spatial expression pattern, the DILPs also show a distinct
67 temporal expression profile. For example, DILP2, DILP4, and DILP7 are already detected in the late
68 embryogenic stage, while DILP3, DILP5, and DILP6 only occur from larval stage on, and DILP1 is
69 only present during the pupal stage (Brogiolo et al., 2001; Kannan and Fridell, 2013).

70 An important question regarding insulin signaling concerns the mechanistic basis of the functional
71 diversification that is observed in *Drosophila* and other insects, even though in the latter, available
72 information is rather limited (Nässel and Vanden Broeck, 2016; Wu and Brown, 2006). One source of
73 functional diversity is the number of insulin-related peptides, which differs significantly between
74 insect species. In the orthopterans *Locusta migratoria* and *Schistocerca gregaria*, only one IRP has
75 been identified (Badisco et al., 2008; Lageux et al., 1990). By contrast, the lepidopteran *Bombyx mori*
76 has 38 IRPs, the highest known number of genes encoding for IRPs in insects (Kondo et al., 1996)
77 and, as mentioned above, *Drosophila* has an intermediate number. Although the structural
78 conservation of IRPs is studied in multiple insect species comprising Diptera, Orthoptera,
79 Lepidoptera, Hymenoptera and Hemiptera, functional characterization remains limited with the
80 exception of *D. melanogaster* (Brogiolo et al., 2001; Grönke et al., 2010). A conserved function of
81 insulin signaling is the regulation of carbohydrate metabolism and growth. In *B. mori* this is regulated
82 by Bombyxin, the first discovered IRP (Nagasawa et al., 1984). Bombyxin is found in other
83 lepidopteran insect species as well, such as *Samia cynthia* and *Precis coenia*, where it also acts as a
84 growth regulator (Nagata et al., 1999; Nijhout and Grunert, 2002). Similar functions were also
85 identified in the honey bee, *Apis mellifera* (Wheeler et al., 2006), and mosquito, *Aedes aegypti*

86 (Brown et al., 2008). In addition, insulin signaling in insects has been implicated in phenotypic
87 plasticity (Emlen et al., 2012; Green and Extavour, 2014; Guo et al., 2016), diapause (Sim and
88 Denlinger, 2008; Williams et al., 2006), circadian rhythmicity (Barber et al., 2016; Barberà et al.,
89 2019; Cong et al., 2015; Vafopoulou and Steel, 2014), and behavior (Wu et al., 2005).

90 To start to address how IRPs evolve and contribute to functional diversification in processes regulated
91 by the insulin signaling pathway, we focus in the current study on aphids. These sap-sucking insects
92 are a very speciose group of which many are pest insects (Calevro et al., 2019). Aphids belong to the
93 order Hemiptera together with cicadas, planthoppers, leafhoppers, shield bugs, and whiteflies (Cryan
94 and Urban, 2012). Aphids thrive on phloem sap, a nutritionally unbalanced diet, by virtue of the
95 evolutionarily ancient symbiotic relationship with the obligatory bacterial endosymbiont, the γ -
96 proteobacterium *Buchnera aphidicola* (Akman Gündüz and Douglas, 2012; Baumann et al., 1995;
97 Shigenobu et al., 2000). These insects represent an excellent paradigm for studying evolutionary
98 changes in the insulin signaling pathway for three broad reasons. First, very little is known about the
99 role of the insulin-signaling pathway in aphid physiology, metabolism, and reproduction. Second,
100 aphids display remarkable phenotypic plasticity with winged and non-winged morphs as well as
101 asexual and sexual reproduction, processes that could be controlled by different IRPs (Barberà et al.,
102 2019; Grantham et al., 2020; Guo et al., 2016). Third, the availability of a number of aphid genome
103 sequences makes it possible to perform evolutionary analyses (Ribeiro Lopes et al., 2020) and study
104 diversification of insulin-related peptides. Given that many insects, including aphid species, have two
105 InR encoding genes (Ding et al., 2017; International Aphid Genomics Consortium, 2010; Smykal et
106 al., 2020), we hypothesize that much of the expected functional diversification comes from the
107 ligands, the insulin-related peptides.

108 In the pea aphid, *Acyrtosiphon pisum*, ten IRPs have previously been annotated by Huybrechts et al.,
109 (2010) based on the first available version of the genome, released in December 2007 (International
110 Aphid Genomics Consortium, 2010). Overall, very little information is available about their evolution
111 and their functional role, apart from the fact that they have been implicated in embryonic development
112 (IRP5) and in photoperiodism (IRP1 and IRP4) (Barberà et al., 2019; Guo et al., 2016). Here we
113 reannotated *A. pisum* IRPs based on a newly available version of the genome (Li et al., 2019), and
114 studied their expression by means of RNA-seq data, qRT-PCR, and immunohistochemistry using
115 newly generated specific antibodies. We also annotated IRPs in nine additional aphid species, and one
116 aphid-related species and described their evolutionary relationship and diversification.

117 **2. Materials and methods**

118 **2.1 Aphid IRP identification**

119 Putative *A. pisum* IRP sequences from Huybrechts et al. (2010) (IRP1-IRP10) and known *D.*
120 *melanogaster* IRP sequences (DILP1-DILP7) retrieved from FlyBase (<http://flybase.org/>) were used
121 as query to perform BLASTP searches against the latest *A. pisum* annotation (Annotation Release
122 103, available at <ftp://ftp.ncbi.nlm.nih.gov/genomes>, last accessed November 16, 2020). Proteins
123 identified by BLAST were scanned against the InterPro database using the InterProScan software
124 v77.0 (<https://www.ebi.ac.uk/interpro/>) (Jones et al., 2014) and identified as putative IRPs if they
125 possessed at least one of the following protein signatures: IPR036438, insulin-like superfamily;
126 IPR022352, Insulin family; IPR016179: insulin-like domain.

127 The putative *A. pisum* IRPs, identified both in this study and by Huybrechts et al. (2010), and known
128 *D. melanogaster* DILP sequences were then used as query to identify the putative IRPs of an
129 additional nine aphid species whose genomes were recently sequenced. Those aphids are part of either
130 the Chaitophorinae subfamily (*Sipha flava*) or of one of the Aphidinae subfamily tribes (Aphidini
131 tribe: *Aphis glycines*, *Aphis gossypii*, *Melanaphis sacchari*, *Rhopalosiphum maidis* and
132 *Rhopalosiphum padi*; Macrosiphini tribe: *Diuraphis noxia*, *Myzus cerasi* and *Myzus persicae*).
133 Genomic information about the grape phylloxera (*Daktulosphaira vitifoliae*), a historical pest of
134 grapevine belonging to a sister-family of Aphididae (*i.e.*, Phylloxeridae family), was also included in
135 this analysis (Rispe et al., 2020). The latest annotation for each of these genomes was obtained either
136 from the NCBI or the AphidBase databases (<https://bipaa.genouest.org/is/>) and results were scanned
137 against the InterPro database as described above.

138 Proteins were checked for the presence of putative signal peptide and monobasic/dibasic amino acid
139 cleavage sites with SignalP v5.0 (<http://www.cbs.dtu.dk/services/SignalP/>) (Almagro Armenteros et
140 al., 2019) and ProP v1.0 (<http://www.cbs.dtu.dk/services/ProP/>) (Duckert et al., 2004), respectively.

141 2.2 Phylogenetic reconstruction

142 Homology relationships between aphid IRPs were further inferred by phylogenetic reconstruction.
143 Candidate sequences were aligned using the MAFFT multiple alignment program v7.0
144 (<https://mafft.cbrc.jp/alignment/software/>) with the mafft-einsi alignment method. Graphical
145 representations of the alignment results were performed with ESPript v3.0 (Robert and Gouet, 2014).
146 We then selected the C20 model as the best substitution model with ModelFinder (Kalyaanamoorthy
147 et al., 2015), using the Bayesian Information Criterion metric, and built a phylogenetic tree using the
148 maximum likelihood estimation method implemented in IQ-TREE v1.6.2 (Nguyen et al., 2015). The
149 reliability of each branch was evaluated using the bootstrap method with 1000 repetitions, and weakly
150 supported branches (<80%) were collapsed using the TreeCollapseCL 4 software
151 (<http://emmahodcroft.com/TreeCollapseCL.html>). Graphical representation of the tree was performed
152 with FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

153 2.3 RNA-seq data analysis

154 Thirteen RNA-seq libraries, corresponding to different *A. pisum* morphs and tissues, were
155 downloaded from the NCBI public database and mapped on the pea_aphid_22Mar2018_4r6ur
156 (GCF_005508785.1) reference genome assembly using the Bioconductor R package Rsubread with
157 the default parameters. All RNA-seq libraries included in this analysis were produced from one of the
158 following three pea aphid strains: LSR1 (male, sexual and asexual female, winged, embryos,
159 bacteriocytes, digestive tract, salivary glands), CWR09/18 (nymph) or CR29-8 (head, 24h head
160 solitary, 24h head crowded). Gene expression of *A. pisum* IRPs was analyzed using the NCBI *A.*
161 *pisum* Annotation Release 103. Fragment read counts per gene were estimated with the Rsubread
162 count tool using the following parameters: countMultiMappingReads=TRUE and
163 allowMultiOverlap=TRUE. Normalized Transcripts Per Million (TPM) were calculated for each
164 library and genes with $\log_{10}(\text{TPM}) > 0.15$ were considered expressed.

165 2.4 Aphid culture

166 Immunohistochemistry and qRT-PCR analyses were performed on a long-established parthenogenetic
167 clone (LL01) of *A. pisum* Harris collected in 1986 in Lusignan (France) and obtained from the
168 Department of Biology of the University of York (UK) in 2008. LL01 aphids are monosymbiotic and
169 contain only the primary endosymbiont *Buchnera aphidicola*. This holocyclic clone was maintained
170 on young broad bean plants (*Vicia faba* L. cv. Aguadulce) at 21°C, with a photoperiod of 16 h light -
171 8 h dark, allowing maintaining aphids as strictly parthenogenetic matriline. To obtain a source of
172 synchronized aphids, winged adults were left on seedlings, allowing them to produce nymphs, and
173 were removed after 24 h. The remaining N1 nymphs were left to grow and nine days later (A9) aphids
174 were collected for dissection, as previously described by Simonet et al. (2016).

175 2.5 Tissue dissection

176 Bacteriocytes, brain, fat body, embryonic chains, gut and carcass (without head) were carefully
177 dissected in ice-cold isosmotic buffer (0.025 M KCl, 0.01 M MgCl₂, 0.25 M Sucrose, and 0.035 M
178 Tris-HCl, pH 7.5) under 25X-40X magnification with a MDG-17 stereomicroscope (Leica, Wild
179 Heerbrugg AG, Switzerland) using a Pasteur glass pipette attached to a vacuum pump for
180 bacteriocytes and fat body, procedure fully described in Ribeiro Lopes et al. (2020), or fine forceps
181 (Dumont no.5) for the other tissues, fully described in Sapountzis et al. (2014).

182 2.6 Genomic DNA isolation and PCR experiments

183 Four synchronized A9 aphids were collected for genomic DNA (gDNA) extraction. This was done
184 using a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's

185 instructions. At the end of the procedure, gDNA was eluted with 100 μ L of Nuclease free water.
186 gDNA quality and concentrations were checked using gel electrophoresis and a NanoDrop® ND-
187 1000 Spectrophotometer (Nanodrop technologies, Wilmington, DE, USA). Putative IRP coding-genes
188 were amplified using specific primers (Supplementary Table S1), which were designed with the
189 primer-Blast software (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). PCR reactions were carried
190 out starting from 1 μ l of a diluted gDNA solution (50 ng/ μ L) using the Taq'Ozyme DNA Polymerase
191 (Ozyme, Montigny-le-Bretonneux, France), with a total reaction volume of 20 μ L. The reaction
192 started with activation of Taq DNA polymerase at 95°C for 5 min, followed by 34 three-step
193 amplification cycles consisting of 30 s denaturation at 95°C, 45 s annealing at primer-specific
194 temperatures (see Supplementary Table S1), and 45 s elongation at 72°C. The PCR reaction was
195 concluded by a final extension for 10 min at 72°C and the program was paused at 4°C. Next, 10 μ L of
196 each amplification product was analyzed on a 1.5% agarose gel stained with Gel Red (Interchim,
197 Montluçon, France). PCR experiments on gDNA were performed to (i) test the primer specificity and
198 effectiveness, and (ii) verify the presence of the gene in the genome. For the latter, the resulting PCR
199 products were cloned in pCR4-TOPO® and transformed to One Shot® TOP10 chemically competent
200 *Escherichia coli* cells according to the manufacturer's guidelines of the TOPO® TA Cloning® Kit for
201 Sequencing (Life Technologies, Carlsbad, CA, USA). The cells were grown overnight at 37°C on LB
202 agar (Invitrogen, Carlsbad, CA, USA) plates containing 50 μ g/mL kanamycin (Merck, Darmstadt,
203 Germany). We transferred grown colonies to 5 mL LB medium (Merck, Darmstadt, Germany), with
204 50 μ g/mL kanamycin. After growing overnight at 37°C, the PCR product containing vectors were
205 purified using NucleoSpin® Plasmid EasyPure' kit (Macherey-Nagel, Dueren, Germany) and
206 sequenced by Sanger Sequencing (LGC, Berlin, Germany).

207 2.7 RNA isolation and quality control

208 Total RNA was extracted as previously described by Simonet et al. (2018), using an RNeasy Mini Kit
209 (Qiagen, Hilden, Germany). In order to get enough RNA, a total of ten aphids per tissue were used for
210 RNA isolation from bacteriocytes, brain, fat body, embryonic chains and gut. Five aphids were used
211 to isolate total RNA from the carcass. For each tissue, three independent biological replicates were
212 processed. In parallel, total RNA was also extracted from the whole body of six aphids (in triplicates)
213 and then used for data normalization in qRT-PCR analysis. Total RNA quality and concentration were
214 checked using gel electrophoresis and a NanoDrop® ND-1000 Spectrophotometer. Samples had to
215 meet the following quality parameters: $A_{260}/A_{280} \geq 1.8$ and $A_{260}/A_{230} \geq 1.8$, in order to be used
216 for subsequent analyses. To remove any genomic DNA contamination total RNA was treated with
217 DNase I (RQ1 RNase-Free DNase, Promega, Charbonnières-les-bains, France) according to the
218 manufacturer's instructions.

219 2.8 Reverse Transcription and qRT-PCR experiments

220 First strand cDNAs were synthesized starting from 200 ng of total RNA, using the SuperScript III
221 First-Strand Synthesis System (Life Technologies, Carlsbad, CA, USA) with oligo(dT)₂₀ primers. In
222 the negative controls, reverse transcriptase was replaced by nuclease free water. The absence of
223 genomic DNA contamination was checked with PCR using primers targeting the *rpl32* control gene in
224 each sample used for the following qRT-PCR experiments (Supplementary Fig. S1). Quantitative
225 real-time PCR reactions (qRT-PCRs) were performed with a Biorad CFX96 Touch Real Time PCR
226 Detection System (Biorad, Hercules, CA, USA) using either 2.5 μ L of cDNA diluted at 1:5, or water
227 (as a negative control) and SYBR Green PCR Master mix in a PCR reaction final volume of 10 μ L,
228 according to the manufacturer's instructions. An internal standard curve was generated for each gene
229 using serial dilutions (from 2000 to 0.002 fg/ μ L) of purified PCR products amplified from a pool of
230 cDNA generated from whole aphids collected at different life stages (N1 to N4 nymphal stages, and
231 adults covering the entire lifespan). The PCR reaction to prepare the control sample for the standard
232 curve was carried out starting from 1 μ L of reverse transcription product using Taq'Ozyme DNA
233 polymerase (Ozyme, Montigny-le-Bretonneux, France), according to the manufacturer's instructions.
234 The same target specific primers were used as in section 2.6. All qRT-PCR reactions were performed
235 in technical triplicates starting from each of the three biological replicates prepared as described in the
236 section 2.7.

237 2.9 *A. pisum* IRP-specific antibody design and production

238 To date, no antibodies were available targeting aphid IRPs. To enable detailed studies of the
239 spatiotemporal distribution of IRPs in *A. pisum* and other aphid species, IRP-specific antibodies were
240 raised in rat using the peptides reported in Supplementary Table S2 as immunogens. The polyclonal
241 antibodies were raised and verified by ELISA by ThermoFisher Scientific (Rockford, IL, USA) as a
242 service. The primary rat anti-IRP antibodies were all used at 1:200 dilution. Mouse anti- β -tubulin (E7;
243 DSHB) was used at 1:10 dilution. Species-specific mouse/rat secondary antibodies conjugated to
244 Alexa Fluor® 488 and Alexa Fluor® 594 (1:200 dilution; Molecular Probes Inc., Oregon, USA) were
245 used. For each IRP two primary rat anti-IRP antibodies were available, of which one was selected
246 based on intensity of the signal in preliminary immunostainings. Furthermore, specificity of
247 secondary antibodies was validated by carrying out immunostaining with either no primary antibody
248 or native Rat IgG (Vector Laboratories, Burlingame, CA, USA) as negative control. This allowed us
249 to verify that any fluorescent signal was solely due to specific binding to primary antibodies.

250 2.10 Immunohistochemistry

251 Whole-mount immunostaining of aphid brain, or digestive tract and embryonic chains was performed
252 with a technique modified from Clements et al., (2008). Freshly collected brain or digestive tract and
253 embryonic chains were individually transferred into Nunc® MicroWell® MiniTrays (Merck,

254 Darmstadt, Germany) or 12-well cell-culture plates, respectively, and fixed in ice-cold 3.7%
255 formaldehyde for 45 min. After multiple washing steps with ice-cold isosmotic buffer, the tissues
256 were preincubated in PAXD (1× PBS containing 5% BSA, 0.3% Triton X-100 and 0.3% sodium
257 deoxycholate) for at least 15 min. Aphid tissues were then sequentially incubated with primary and
258 secondary antibodies in PAXD overnight at 4°C, with multiple washing steps with PAXD after each
259 incubation. Stained tissues were mounted on glass slides in VECTASHIELD antifade mounting
260 medium (Vector Laboratories, Burlingame, CA, USA), using spacers cut from coverslips to avoid
261 crushing the tissues. For brains, all the solution-removal steps were performed under a Leica M80
262 stereomicroscope (Leica Microsystems, Buffalo Grove, IL, United States) to check the integrity of the
263 tissue and to keep them in the smallest volume possible. Confocal images were collected as single
264 optical sections or Z-series (serial optical sections at 0.1 μm intervals) using a Fluoview FV1000
265 confocal microscope (Olympus Corporation, Tokyo, Japan) and processed with ImageJ software
266 (<https://imagej.net/>).

267 2.11 Statistical analysis

268 For qRT-PCR experiments, three genes were tested as potential candidates for data normalization:
269 *rpl7* (NP_001129370.1), actin (NP_001119672.1) and *rpl32* (NP_001119682.1). *rpl7* was retained as
270 the best normalization gene based on a preliminary analysis of the data with the BestKeeper software
271 tool (Pfaffl et al., 2004). The expression levels (log₂) of the different IRPs genes were then
272 normalized relative to those of the *rpl7* gene. The relative expression ratio R (and the associated
273 standard error) of IRP mRNAs in the different tissues was calculated using as reference the mean of
274 the IRP transcript expression levels of the aphid control group (A9 whole body, see the section 2.7)
275 using the REST software tool (Pfaffl et al., 2002). This ratio was calculated taking into account the
276 real-time PCR efficiency of each gene (E) and the crossing point difference (DCP) of a test condition,
277 as compared to the reference condition, and expressed in comparison to the normalization gene (*rpl7*)
278 using the following model (Pfaffl, 2001): $R = (E_{\text{target}})^{\Delta C P_{\text{target}}(\text{control-sample})} / (E_{\text{reference}})^{\Delta C P_{\text{reference}}(\text{control-}$
279 $\text{sample})}$. Results are displayed as mean ± SD of three independent biological experiments.

280 3. Results

281 3.1 *A. pisum* IRP annotation

282 We identified seven putative IRP encoding-genes in the genome of *A. pisum* in the newly improved *A.*
283 *pisum* genome annotation, released in 2019 (Li et al. 2019). Among those seven genes,
284 LOC100169635 and LOC100570058 each encode two isoforms of IRP5 and IRP6, respectively, with
285 the mRNAs of each isoform differing only in their 3' UTR region (Table 1). We compared the
286 sequences of those proteins with the ones identified by Huybrechts et al. (2010). The protein
287 sequences encoded by LOC100568938, LOC100574788, LOC100575361 and LOC100161832 are

288 identical to the previously identified IRP1, IRP2, IRP3 and IRP4, respectively. The two proteins
289 encoded by LOC100169635 are identical to the IRP5 identified by Huybrechts et al. (2010) but have
290 an additional 117 amino acids at the C-terminus. The two proteins encoded by LOC100570058 are
291 identical to IRP6 except for six additional amino acids at the C-terminus and substitutions in the first
292 three residues (Supplementary Fig. S2). Importantly, the gene LOC100573276 encodes a protein that
293 is very divergent from the other aphid IRPs and was identified only when using *D. melanogaster*
294 proteins as query. While InterPro classified it as part of the insulin-like superfamily, no insulin-like
295 domain was predicted contrary to the six other identified IRPs. In accordance with the nomenclature
296 used by Huybrechts et al. (2010), we named this protein IRP11. The genes encoding the previously
297 identified IRP7-IRP10 were not found in the new version of the *A. pisum* genome.

298 3.2 *A. pisum* IRP sequence analysis

299 The sequences and genomic organization of the seven *A. pisum* IRPs identified in this paper (IRP1-
300 IRP6 and IRP11) were analyzed in more detail. We confirmed that all these putative IRPs possess the
301 signal peptide, the A and B chains, as well as the six conserved cysteine residues that are found in all
302 insulin-related peptides (Fig. 1) (Badisco et al., 2008; Brogiolo et al., 2001; Kondo et al., 1996;
303 Lageux et al., 1990; Steiner, 1985). The four cysteine residues of the A chain are ordered as
304 CC(X)₃C(X)₈C, in which X can be any amino acid, while the two cysteine residues of the B chain are
305 separated by 11 amino acid residues in all IRPs with the exception of IRP11 where they are separated
306 by 12 residues.

307 IRP1-IRP4 all have a long C chain, framed by dibasic convertase cleavage sites, in between the A and
308 B chains (Fig. 1). Furthermore, these *A. pisum* IRPs align on their complete sequence (99-100%
309 coverage) and are highly similar with between 62% and 89% identity at the amino acid levels, and
310 between 75% and 89% if only IRP2, IRP3 and IRP4 are considered (Table 2). IRP1-IRP4 also share
311 similar gene structures, with two introns at identical positions, one in the 5' UTR and one at the
312 beginning of the B chain-encoding sequence (Fig.2). IRP4 also possesses an additional intron in the
313 5'UTR. IRP2-IRP4 are all localized on the X chromosome, but too far apart to suggest tandem
314 duplication (Fig. 2). IRP1 is localized on chromosome A3.

315 In contrast to IRP1-IRP4, the A and B chain of IRP5 and IRP6 are separated by a shorter C chain
316 containing a single furin-like cleavage site (Fig. 1). These two proteins also possess an elongated C-
317 terminal domain following the A chain. Apart from their structures, IRP5 and IRP6 have very
318 divergent amino acid sequences with only 29% identity. Similar levels of identity are observed when
319 comparing IRP5 and IRP6 to IRP1-IRP4 (Table 2). Furthermore, the IRP5 C-terminal domain is
320 longer than in IRP6 (Fig. 1). Importantly, the arginine residue at position 97 in the IRP5 sequence
321 constitutes a potential cleavage site, suggesting that this IRP precursor could be processed at the end
322 of the A chain. IRP5 and IRP6 also have a different gene structure. They both present one intron in

323 the 5'UTR region (or two in the case of one IRP6 transcript) and one intron at the end of both chains
324 A and B-encoding sequences. The fourth intron, however, is either found in the regions corresponding
325 to the C-terminal domain for IRP5 or the signal peptide in IRP6 (Fig. 2).

326 IRP11 is unique as it comprises both a truncated B chain, where the two conserved cysteines are
327 separated by 12 amino acids, and an unusually long C chain that is framed by dibasic amino acid
328 residues (Fig. 1). This protein is also very divergent from other *A. pisum* putative IRPs (Table 2).
329 Interestingly, the gene encoding IRP11 is found directly upstream of the IRP5 encoding gene in a tail-
330 to-tail arrangement (Fig. 2).

331 3.3 IRP identification in other aphids

332 We have used the recently completed genome sequences of an additional nine aphid species (*A.*
333 *glycines*, *A. gossypii*, *D. noxia*, *M. sacchari*, *M. cerasi*, *M. persicae*, *R. maidis*, *R. padi* and *S. flava*)
334 and one aphid sister-group (*D. vitifoliae*) to examine the conservation of the IRP gene family in those
335 insects. We used the seven predicted *A. pisum* IRP sequences as query to search for homologs in each
336 of those aphids and identified a total of 77 IRP-encoding genes, corresponding to 86 different proteins
337 (Supplementary Table S3). Importantly, including the IRP7-IRP10 sequences identified before the
338 availability of the latest *A. pisum* genome annotation or *D. melanogaster* DILPs to our query did not
339 result in the identification of additional putative IRPs. Therefore, further analysis was conducted with
340 only the seven predicted IRP sequences from this study.

341 Each of the 86 aphid IRPs contains the six cysteine residues involved in disulfide bridge formation
342 except for the proteins Agl-AG006140 (from *A. glycines*) and Rp-g21416 (from *R. padi*). They lack
343 an A or B chain, respectively, and are thus unlikely to function as a *bona fide* IRP. Functional signal
344 peptides were identified in the majority of putative IRP pre-propeptides using the SignalP prediction
345 software (Almagro Armenteros et al., 2019), which indicates that they are able to act as secreted
346 peptide hormones. There are, however, 17 notable exceptions (eight proteins in *R. padi*, two in *M.*
347 *persicae*, *A. glycines* and *R. maidis*, and one in *D. noxia*, *S. flava* and *M. sacchari*) that do not have
348 this predicted signal peptide. Importantly, the sequence preceding the B chain of these 17 proteins are
349 either longer or shorter than other predicted IRPs. For the proteins with a longer sequence preceding
350 the B chain, computational shortening (deletion of the additional N-terminal amino acids) resulted in
351 a functional signal peptide detectable by the SignalP prediction software (Almagro Armenteros et al.,
352 2019).

353 Nine IRP11 homologs were found. Five of them were classified by InterPro as part of the insulin-like
354 superfamily but had no clear insulin-like domain. For the remaining four, no insulin signatures were
355 found. Furthermore, ten supplementary proteins (one from *M. persicae* and *A. gossypii*; three from *R.*
356 *padi*; and five from *S. flava*), not homologs to IRP11, had no insulin-like domain, despite being

357 identified as part of the insulin-like superfamily. Those proteins also presented a higher degree of
358 divergence with other aphid IRPs (Supplementary Table S4).

359 If we look at each species separately, we observe that aphids have varying numbers of IRP-encoding
360 genes, from the four found in *D. noxia* or *M. cerasi* to the 15 found in *R. padi* (Supplementary Table
361 S3). Importantly, there is no apparent correlation between IRP numbers and the size of each genome,
362 the number of protein-encoding genes or the database from which the genome was extracted
363 (Supplementary Table S5). Furthermore, this variability is observed even between aphids of the same
364 order. For instance, *M. persicae* has nine IRPs, which is twice the number predicted in *M. cerasi*. The
365 minimal number of IRPs was found in an aphid sister group, *D. vitifoliae* that only possesses three
366 IRPs.

367 3.4 Phylogenetic analysis of aphid IRPs

368 A phylogenetic reconstruction was made to better understand the evolutionary history that led to this
369 remarkable diversity (Fig. 3). Based on this phylogeny and their amino acid sequence similarities,
370 aphid IRPs were grouped into six subfamilies (A to F). The A, B, C and D subfamilies regroup the
371 orthologs of IRP1-IRP4, IRP5, IRP6 and IRP11 respectively. The E and F subfamilies regroup
372 sequences that have no clear orthologs in *A. pisum*. Proteins from the D, E and F subfamilies were not
373 identified when using the DILPs as query. Interestingly, among the ten proteins that lack a clearly
374 identified insulin-related domain and are not part of the D subfamily (IRP11 orthologs), eight have no
375 clear homology relationship with the other aphid IRPs and were not sorted into one of the subfamilies.
376 IRP5 and its orthologs in subfamily B are by far the most highly conserved IRPs, with an overall
377 amino acid identity of 66% between the pre-propeptides of *A. pisum* and the most distantly related
378 species *D. vitifoliae*, and increasing to 89%, when only aphids from the Aphidinae subfamily are
379 considered (Supplementary Table S4). One unique ortholog of IRP5 has been retained in each of the
380 aphids considered in this study and each of them possesses a shortened C-peptide and a long C-
381 terminal domain following the A chain. Interestingly, both the shortened C-peptide and the C-terminal
382 domain are highly conserved in the different sequences (Supplementary Fig. S3B).

383 IRP1-IRP4, which are part of the A subfamily, are the next most conserved, with a minimum of 41%
384 overall amino acid identity and 66% and 63% identity in the A and B chain, respectively vs. *D.*
385 *vitifoliae* (when comparing DV3012863 to IRP1, Supplementary Table S4). The minimum overall
386 identity increases to 59% when only aphids from the Aphidinae subfamily are considered. The A
387 subfamily is the largest subfamily with 24 proteins and at least one representative of this subfamily
388 has been conserved in each aphid species. All these proteins present the classical structure found in
389 insulin-like proteins with a long C-peptide framed by conserved dibasic convertase cleavage site, with
390 the exceptions of the ones from *S. flava* and *D. vitifoliae* and one homolog of *M. persicae* that lack
391 one of the cleavage sites (Supplementary Fig. S3A).

392 We found IRP6 homologs only in the Aphidinae subfamily while four aphids (*D. noxia*, *M. cerasi*, *M.*
393 *persicae* and *S. flava*) and phylloxera have no representative in the C subfamily. Furthermore, like
394 IRP5s, IRP6 homologs all have a very short C-peptide (Supplementary Fig. S3C). The nine proteins
395 of the C subfamily can be divided into two groups of more closely related sequences. The first group
396 includes two proteins (one from *R. padi* and one from *A. gossypi*) that are 71% identical, and the
397 second group comprises seven proteins (three from *R. maidis*; one from *M. sacchari*, *A. gossypi*, *A.*
398 *glycines* and *A. pisum*) with identity percentages ranging from 63% to 86%.
399 The D subfamily includes all the IRP11 homologs. One homolog of IRP11 was identified in each
400 aphid with the exception of *M. cerasi* but only six were classified as part of the insulin-like
401 superfamily as Agl-AG006140, Ago-XP_027847094.1 and Ms-XP_025201716.1 have no insulin-like
402 signatures. *D. vitifoliae* also lack an IRP11 homolog. Importantly, sequences from this group gave no
403 significant results when aligned against most of the other aphid IRPs. While they all contain the six
404 conserved cysteine residues involved in intra- and inter-chain disulfide bonds, their positions differ
405 from other IRPs as the two cysteines in the B chain are separated by 12 amino acids instead of 11
406 (Supplementary Fig. S3D). The E subfamily includes proteins from the six aphids that are part of the
407 Aphidini tribe. They all possess a classic A and B chain but no dibasic cleavage site. The F subfamily
408 includes 20 proteins that have an unconventional A chain that possesses nine amino acid residues
409 between the third and fourth cysteine instead of eight.

410 3.5 mRNA expression of *A. pisum* IRPs across morphs and tissues

411 We analyzed the expression profile of *A. pisum* IRPs in several RNA-seq libraries, corresponding to
412 different *A. pisum* morphs (males, sexual females, parthenogenetic females, winged females) and
413 tissues (head, gut, salivary glands, bacteriocytes, embryos) publicly available in the NCBI SRA
414 (Sequence Read Archive) database (Table 3). This analysis highlighted very different expression
415 profiles for the different pea aphid IRPs. IRP4 and IRP5 were found in all RNA-seq libraries, albeit at
416 different levels of expression. IRP5 is by far the most abundant IRP in *A. pisum* and is the most
417 expressed IRP in all the tissues and conditions considered. Two notable exceptions are: (i) IRP4 and
418 IRP5 levels are comparable in the heads of aphids at high population density, a condition that is
419 linked with alate-wingless polyphenism and (ii) IRP4 levels are higher than IRP5 in male aphids.
420 Expression of IRP1 and IRP11 are detected in all morphs and tissues with the exception of
421 bacteriocytes. Moreover, IRP1 is highly expressed in the head. IRP2 is only expressed in males and in
422 the heads of aphids at high population density, and IRP3 only in males and salivary glands. IRP6 is
423 the only IRP other than IRP4 and IRP5 for which expression was detected in the bacteriocytes and it
424 is comparatively poorly expressed in other tissues.
425 Analysis of the RNA-seq libraries gave us a first indication of the expression pattern of the IRPs, but
426 also revealed that IRP spatiotemporal expression can vary depending on *A. pisum* strain, morph and/or

427 rearing conditions. We therefore chose to complement this data with qRT-PCR experiments,
428 performed on six different tissues (embryonic chains, bacteriocytes, carcass, brain, fat body and
429 digestive tract) dissected from synchronized 9 days-old individuals collected from a unique
430 population of aphids of the LL01 strain (Table 4). IRP1 shows significantly higher expression in the
431 brain than in the other tissues. IRP4 also shows higher expression in the brain, and to a lesser extent in
432 the carcass, compared to the other tested tissues. IRP2 is mildly enriched in the carcass and the brain,
433 while IRP5 and IRP6 are more highly expressed in fat body and carcass. Interestingly, while IRP11
434 expression is detected in all tissues analyzed, it appears to be specific to the carcass as it is the only *A.*
435 *pisum* IRP to be significantly enriched in this tissue. We were not able to amplify IRP3 using qRT-
436 PCR, even though the primers we designed were reliable and able to amplify the corresponding
437 encoding gene starting from gDNA (Supplementary Fig. S4).

438

439 3.6 IRP distribution in *A. pisum* tissues by immunohistochemistry

440 Although the expression analysis gives us quantitative information on the expression in specific
441 tissues, this technique does not allow detection in specific cells or cell types within those tissues. So,
442 to further complement the expression profiles, we localized the precise tissue and cell distribution of
443 the *A. pisum* IRPs by means of immunohistochemistry for six out of seven IRP using newly
444 generated, aphid-specific antibodies. IRP distribution was analyzed in different organs throughout the
445 aphid body, more specifically the embryonic chains, digestive tract and central nervous system
446 comprising brain, suboesophageal ganglion, and thoracic ganglionic mass. Both IRP1 and IRP4
447 immunoreactivity was observed in the brain. The positive signal was localized in two groups of four
448 neuronal cells symmetrically located in both brain hemispheres of the protocerebrum (Fig. 4). IRP2,
449 on the other hand, showed mildly positive signal in the digestive tract in circular structures (Fig. 5).
450 Aside from unspecific staining in the embryonic bacteriocytes present for all IRPs, no other
451 immunoreactivity was observed in any of the investigated organs for IRP3, IRP5 and IRP6.

452 4. Discussion

453 The insulin-signaling pathway is a central component of carbohydrate metabolism in all eukaryotes
454 including insects. The number of insulin and insulin-like peptides can vary tremendously. The
455 classification of IRPs is based on similarities in the amino acid sequence organization with those of
456 mammalian insulins, with focus on the cysteine positioning within the A and B chain, and the
457 presence of a conserved pre-proinsulin signature (Kondo et al., 1996; Steiner, 1985). Based on this
458 sequence organization and signatures, we identified seven IRP encoding genes in the newest
459 improved *A. pisum* genome annotation. IRP1-IRP6 and IRP11 possess the classical IRP organization,
460 comprising a signal peptide, A and B chains, and six conserved cysteine residues which forms one
461 intrachain cysteine bridge on the A chain and two cysteine bridges connecting the A and B chain

462 (Badisco et al., 2008; Brogiolo et al., 2001; Kondo et al., 1996; Lageux et al., 1990; Steiner, 1985).
463 All the identified *A. pisum* IRPs contain an insulin-like domain (based on InterPro analysis) except for
464 IRP11, which is highly divergent. The previously described IRP7-10 are not confirmed in the newly
465 annotated genome sequence.

466 The protein structure of IRP1-IRP4 in *A. pisum* is similar to the one commonly found in insulin and
467 insulin-like peptides produced by endocrine cells or peptidergic neurons, as was previously shown by
468 Huybrechts et al. (2010). This observation, combined with the sequence similarities and intron
469 positioning of these four IRPs suggests that IRP1, IRP2, IRP3 and IRP4 likely arose by duplication
470 events. A similar hypothesis was proposed by Barberà et al. (2019), who investigated the evolutionary
471 relationship of the ten putative *A. pisum* IRPs identified by Huybrechts et al. (2010) with selected
472 sequences from insects from different orders (including information available for *M. persicae* as the
473 only other aphid). By extending our analysis to ten aphid species and using the maximum likelihood
474 estimation method, a method that gives more robust and reliable phylogenies compared to the
475 distance method used by Barberà et al. (2019), we were able to specify the evolutionary history that
476 led to this amplification and showed that the presence of IRP1-IRP4 in *A. pisum* is likely the result of
477 two duplications after the divergence between the Aphidinae and Chaitophorinae subfamilies of
478 aphids. The phylogeny highlights a complex pattern of duplication and loss, suggesting that at least
479 one duplication probably preceded the divergence of the Aphidinae subfamily and gave rise to a
480 subgroup comprising IRP1 and its direct orthologs and another subgroup comprising IRP2-IRP4 and
481 their orthologs. It is likely that another duplication occurred in the Macrosiphini tribe and that one of
482 the duplicated genes was later lost in *M. cerasi* and *D. noxia*. According to our phylogenetic analysis,
483 at least one IRP with similar structure (subfamily A) was present in all examined aphid species and
484 the aphid sister family of Phylloxeridae. This is not surprising as this structure is broadly conserved
485 within the different insect orders (mainly Diptera and Lepidoptera) in which IRPs have been studied.
486 In *D. melanogaster*, these conserved IRPs comprise DILP1-3 and -5 (Brogiolo et al., 2001). In *B.*
487 *mori*, ten such IRP genes were found (Kondo et al., 1996).

488 RNA expression analysis of *A. pisum* IRPs revealed that IRP1 and IRP4 are broadly expressed in the
489 different morphs with high expression in the head. By means of immunohistochemistry, we showed
490 the presence of IRP1 and IRP4 in four cells in each brain hemisphere corresponding to the pars
491 intercerebralis. This observation is consistent with the results of Barberà et al. (2019) who previously
492 showed expression of those IRPs using *in situ* hybridization and identified these cells as likely group I
493 neurosecretory cells (NSCI). This site of expression is a conserved feature for classical insulin and
494 insulin-like peptides, seen in many invertebrates. In the dipteran model organism, *D. melanogaster*,
495 DILP1-3, and DILP5 are produced in the IPCs also located in the pars intercerebralis, where DILP1 is
496 only expressed in the pupal stage (Brogiolo et al., 2001; Nässel and Vanden Broeck, 2016). Also, in

497 other dipterans such as *A. aegypti* and *Anopheles gambiae* the IRP expression is located in
498 neurosecretory cells of the brain (Cao and Brown, 2001; Krieger et al., 2004). In Lepidoptera, studies
499 of IRP gene expression are mainly focused on the larval stages. Nonetheless, in this insect order, the
500 expression is also localized in median neurosecretory cells (Kondo et al., 1996; Wu and Brown,
501 2006). In Orthoptera, e.g. *L. migratoria* and *S. gregaria* and the blattodean *Periplaneta americana*, or
502 more distant phylogenetic invertebrate groups, like nematodes, the expression is likewise observed in
503 the central nervous system (Badisco et al., 2008; Raabe, 1986; Smit et al., 1998). Based on RNA-seq
504 and qRT-PCR data, IRP1 and IRP4 are also expressed in other tissues than the brain, such as the
505 digestive tract. We observed no positive signal in the gut using specific antibodies. According to both
506 RNA-seq and our qRT-PCR results, IRP1 and IRP4 are highly enriched in the brain, but they are only
507 poorly expressed in the gut (2800 and 1300 times lower than in the brain, based on qRT-PCR data)
508 and their expression might be so low and diffuse that this is insufficient to give rise to a positive
509 signal using immunohistochemistry. Our RNA-seq meta-analysis provides evidence for high
510 expression of IRP1 and IRP4 in the salivary glands. We cannot exclude that the corresponding RNA-
511 seq libraries are contaminated with brain as they are contained in the aphid head and in close
512 proximity. However, we also have to consider that IRPs might be differentially expressed in salivary
513 glands of different aphid strains or depending on the rearing conditions.

514 The RNA-seq expression analysis of IRP2 and IRP3, on the other hand, shows low levels of
515 expression in the male form, and in the head of crowded populations or salivary gland, respectively.
516 Based on our immunohistochemical data, IRP2 is present in a part of the intestinal tract of adult
517 wingless parthenogenetic females in non-crowded conditions. This is reminiscent of *D. melanogaster*
518 DILP3 which is not only expressed in the IPCs but also in the intestinal muscle cells (Ikeya et al.,
519 2002; Nässel and Vanden Broeck, 2016; Veenstra et al., 2008). It is the only *Drosophila* gut hormone
520 that is not expressed by the midgut endocrine cells, which leads us to propose that in *A. pisum* the
521 observed circular structures correspond to circular muscles of the digestive tract (Veenstra et al.,
522 2008). The fact that these structures constitute only a fraction of the cells of the gut could explain that
523 expression of IRP2 was not detected by qRT-PCR and RNA-seq, which were performed on the whole
524 tissue.

525 The selective expression in specific morphs or developmental stages can be explained by a spatial
526 expression pattern that reflects separate functions of the peptide. In *D. melanogaster* for example,
527 DILP2, DILP4, and DILP7 are already detected in late embryonic stages, while DILP3, DILP5, and
528 DILP6 only occur from larval stage on, and DILP1 is only expressed during the pupal stage (Brogiolo
529 et al., 2001; Kannan and Fridell, 2013; Nässel and Broeck, 2016). The different DILPs jointly ensure
530 normal growth and development (Ikeya et al., 2002; Kannan and Fridell, 2013). Therefore, we expect
531 that also in aphids, distinct spatiotemporal expression patterns are involved in regulation of growth,

532 development and even the adaptation to their specific lifestyle (Barberà et al., 2019). One such
533 example for IRP4 is already provided by Barberà et al., (2019) who showed increased IRP4
534 expression in aphids submitted to a switch in photoperiodism, a condition that leads to the appearance
535 of males in aphid populations.

536 The IRP5 and IRP6 protein structure, in contrast, shows features similar to DILP6 and other IGF
537 peptides. Furthermore, the elongated A chain is reminiscent of human IGF. IGFs are characterized by
538 a single peptide chain with a retained C chain, linked with internal cysteine bridges (Brogiolo et al.,
539 2001; Nässel and Vanden Broeck, 2016). According to our phylogenetic analysis, an IRP5 homolog is
540 found in all examined aphid species and the sister-group phylloxera (subfamily B) and both the
541 shortened C-peptide and the C-terminal domain characteristic of this subfamily are highly conserved
542 in the different sequences. This finding may indicate that although it contains a furin-like cleavage
543 site, the C-peptide is not removed and is part of a single chain bioactive peptide, as has been recently
544 suggested for the IRP6-like BIGFLP protein in the silkworm *B. mori* and DILP6 in *D. melanogaster*,
545 which act as insulin-like growth factors (IGFs) (Okamoto et al., 2009a; Okamoto et al., 2009b). The
546 same can be hypothesized for the C-terminal domain even though *S. flava* and *D. vitifoliae* present a
547 slightly more divergent sequence that is notably devoid of a predicted cleavage site between the A
548 chain and C-terminal domain. IRP6, however, is only present in six of the examined aphid species
549 (subfamily C). Based on the short C-chain and their respective position in the tree, it is possible that
550 IRP6 is the result of an IRP5 duplication that probably took place after the divergence of Aphidinae
551 and Chaitophorinae. Furthermore, the phylogenetic division of subfamily C into two subgroups
552 suggests that a second duplication took place, possibly in the Aphidini tribe, and that one of the
553 duplicated genes was later lost in all species except *A. gossypi*. Concerning their expression patterns,
554 IRP5 is highly expressed throughout all different *A. pisum* morphs and developmental stages, as well
555 as in different organs, while IRP6 displays lower and more restricted expression levels. However, no
556 immunoreactivity was found in any of the tested tissues. One possible explanation could be that IRP5
557 and IRP6 are expressed in the diffusely spread fat body rather than in specific groups of cells or
558 tissues. This notion is supported by the qRT-PCR data that show up-regulated expression in fat body.
559 Further support for this hypothesis comes from the *Drosophila* IGF-like peptide DILP6. DILP6 is
560 produced by the fat body and functions as a regulator of carbohydrate and lipid storage as well as an
561 oxidative stress regulator (Bai et al., 2012). The fat body in insects is the functional equivalent of the
562 vertebrate liver and adipose tissue (Okamoto et al., 2009).

563 Importantly, we identified another protein with insulin-like characteristics similar to those of the *D.*
564 *melanogaster* DILPs (Brogiolo et al., 2001), IRP11. This peptide is structurally very divergent from
565 other *A. pisum* IRPs and the truncation of its B chain raises questions about the functionality of
566 IRP11. However, our comparative analysis allowed us to identify IRP11 homologs in eight of the

567 other aphid species included in this study (subfamily D) but only six, including *A. pisum* IRP11, were
568 classified as part of the insulin-like superfamily. The remaining three proteins lack an insulin
569 signature, probably due to the fact that they either have no A chain (in the case of Agl-AG006140) or
570 have an additional residue in their A chain, with a resulting CC(X)₄C(X)₈C motif. According to RNA-
571 seq and qRT-PCR data, IRP11 is enriched in *A. pisum* carcass suggesting a role in aphid physiology
572 different from all other IRPs. The IRP11 encoding-gene was previously listed as a putative gonadulin,
573 a subclass of insulin-like peptides that share similarities with *Drosophila* ILP8 (Veenstra, 2020). The
574 data presented here confirms that IRP11 belongs to this group since gonadulin ortholog sequences (i)
575 are very divergent not only from other IRPs but also among each other, even in closely related
576 species, (ii) share a common motif, containing 12 amino acid residues between the two cysteines of
577 the B-chain, (iii) generally show lower expression levels in males than females, as is the case in the
578 pea aphid according to the RNA-seq data analyzed here and (iv) IGF-like and gonadulin-like proteins
579 are encoded by genes that are next to each other in the genomes of arthropods, a structural
580 organization that we found for the IRP5 and IRP11 encoding-genes in the pea aphid genome. This, in
581 addition to the fact that IRP11 appears more closely related to IRP5 and IRP6 according to our
582 phylogenetic analysis, suggests that IRP11 and IRP5-like proteins might have evolved from
583 duplication of an ancestral gene and have since undergone multiple substitutions. Importantly, while
584 gonadulin-like proteins are often highly expressed in insect gonads, the pea aphid IRP11 is
585 specifically enriched in the pea aphid carcass, which suggests that this protein could have developed
586 new aphid-specific functions, related to their characteristic features, e.g. polyphenisms and obligatory
587 symbiosis with *B. aphidicola*. This, in addition to the fact that no IRP11 homolog was identified in the
588 aphid sister-group phylloxera, suggests that these IRPs could be the result of an aphid-specific
589 duplication. A possibility, supported by the gene location of IRP11 compared to IRP5, is that IRP11 is
590 the result of an ancestral duplication of IRP5 that has since undergone multiple substitutions. IRP11
591 could have developed new aphid-specific functions, related to their characteristic features, e.g.
592 polyphenisms, obligatory symbiosis with *B. aphidicola*.

593 The phylogenetic analysis revealed multiple aphid IRPs with no clear homolog in *A. pisum*, which
594 have been classified in subfamily E and F. None of these sequences were identified using the
595 *Drosophila* DILPs as query. This is most likely due to their higher degree of divergence compared to
596 other more conserved IRPs. It is possible that proteins that compose these groups represent aphid-
597 specific orthologs. Furthermore, eight of the identified IRPs have no clear homology relationship with
598 any of the other aphid IRPs and were not sorted into one of the subfamilies. This could be the result of
599 duplication events and subsequent functional diversification or pseudogenization.

600 The varying number of IRPs within each aphid species indicate that multiple gene duplications and/or
601 deletions must have occurred during the evolutionary history of this protein family in aphids. Overall,

602 the phylogenetic analysis shows us that, while some IRPs (subfamily A and B) have been stably
603 retained in the 10 aphid genomes, others have undergone several rounds of gene duplication/loss
604 (subfamily C and D) in a few species. The grape phylloxera only possesses three IRPs, with two
605 belonging to subfamily A and one to subfamily B. This suggests that the absence of homologs of
606 subfamily C, D, E, and F could be either the result of aphid specific duplication or losses throughout
607 evolution and diversification. Globally, the different subfamilies show distinct degrees of amino acid
608 conservation *e.g.*, the highly conserved IRP5 sequences compared to the more diverse IRP1-IRP4
609 sequences, suggesting different degrees of either functional constraint or positive selection. The
610 higher amino acid sequence conservation of IRP5 suggests that this IRP might have essential
611 functions that are different from the other IRPs and can therefore not be compensated by the others.
612 Moreover, the expression (mRNA) and immunohistochemistry analyses revealed expression of IRP1
613 and IRP4 in neurosecretory cells in the pars intercerebralis of the brain, IRP2 in circular muscles of
614 the digestive tract and IRP5 and IRP6 in the fat body. Given the similarities in tissues distribution
615 with *Drosophila* (DILP1, -2, 3, 5; DILP3; and DILP6, respectively) we propose that these expression
616 patterns are indicative of a prominent role for these proteins in aphid physiology most likely
617 comprising functions in *e.g.* carbohydrate metabolism, during development, and in phenotypic
618 plasticity (Brogiolo et al., 2001; Emlen et al., 2012; Green and Extavour, 2014; Grönke et al., 2010;
619 Guo et al., 2016).

620 **Acknowledgments:** The authors acknowledge Mattias Winant for help with confocal microscopy
621 imaging.

622 **Funding:** C.H. is supported by an FWO Strategic Basic research PhD fellowship, BF2i was supported
623 by INSA Lyon and INRAE, P.C. was supported by FWO grants G065408.N10 and G078914N and by
624 KULeuven grant C14/17/099.

625 **References**

- 626 Akman Gündüz, E., Douglas, A.E., 2012. Symbiotic bacteria enable insect to use a nutritionally
627 inadequate diet. *Proc. R. Soc. B Biol. Sci.* 276, 987–991. <https://doi.org/10.1098/rspb.2008.1476>
- 628 Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S.,
629 von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep
630 neural networks. *Nat. Biotechnol.* 37, 420–423. <https://doi.org/10.1038/s41587-019-0036-z>
- 631 Altstein, M., Nässel, D.R., 2010. Neuropeptide Signaling in Insects. *Adv. Exp. Med. Biol.* 692, 155–
632 165. https://doi.org/10.1007/978-1-4419-6902-6_8
- 633 Badisco, L., Claeys, I., Van Hiel, M., Clynen, E., Huybrechts, J., Vandersmissen, T., Van Soest, S.,
634 Vanden Bosch, L., Simonet, G., Vanden Broeck, J., 2008. Purification and characterization of an
635 insulin-related peptide in the desert locust, *Schistocerca gregaria*: immunolocalization, cDNA
636 cloning, transcript profiling and interaction with neuroparsin. *J. Mol. Endocrinol.* 40, 137–150.

637 <https://doi.org/10.1677/jme-07-0161>

638 Bai, H., Kang, P., Tatar, M., 2012. *Drosophila* insulin-like peptide-6 (dilp6) expression from fat body
639 extends lifespan and represses secretion of *Drosophila* insulin-like peptide-2 from the brain.
640 *Aging Cell* 11, 978–985. <https://doi.org/10.1111/accel.12000>

641 Barber, A.F., Erion, R., Holmes, T.C., Sehgal, A., 2016. Circadian and feeding cues integrate to drive
642 rhythms of physiology in *Drosophila* insulin-producing cells. *Genes Dev.* 30, 2596–2606.
643 <https://doi.org/10.1101/gad.288258.116>

644 Barberà, M., Cañas-Cañas, R., Martínez-Torres, D., 2019. Insulin-like peptides involved in
645 photoperiodism in the aphid *Acyrtosiphon pisum*. *Insect Biochem. Mol. Biol.* 112, 103185.
646 <https://doi.org/10.1016/j.ibmb.2019.103185>

647 Baumann, P., Baumann, L., Lai, C.Y., Rouhbakhsh, D., Moran, N.A., Clark, M.A., 1995. Genetics,
648 physiology, and evolutionary relationships of the genus *Buchnera*: Intracellular symbionts of
649 aphids. *Annu. Rev. Microbiol.* 49, 55–94. <https://doi.org/10.1146/annurev.mi.49.100195.000415>

650 Bridge, D., Theofiles, A.G., Holler, R.L., Marcinkevicius, E., Steele, R.E., Martínez, D.E., 2010.
651 FoxO and stress responses in the cnidarian *Hydra vulgaris*. *PLoS One* 5, e11686.
652 <https://doi.org/10.1371/journal.pone.0011686>

653 Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., Hafen, E., 2001. An evolutionarily
654 conserved function of the drosophila insulin receptor and insulin-like peptides in growth control.
655 *Curr. Biol.* 11, 213–221. [https://doi.org/10.1016/S0960-9822\(01\)00068-9](https://doi.org/10.1016/S0960-9822(01)00068-9)

656 Brown, M.R., Clark, K.D., Gulia, M., Zhao, Z., Garczynski, S.F., Crim, J.W., Suderman, R.J., Strand,
657 M.R., 2008. An insulin-like peptide regulates egg maturation and metabolism in the mosquito
658 *Aedes aegypti*. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5716–5721.
659 <https://doi.org/10.1073/pnas.0800478105>

660 Calevro, F., Tagu, D., Callaerts, P., 2019. *Acyrtosiphon pisum*. *Trends Genet.* 35, 781–782.
661 <https://doi.org/10.1016/j.tig.2019.07.003>

662 Cao, C., Brown, M.R., 2001. Localization of an insulin-like peptide in brains of two flies. *Cell Tissue*
663 *Res.* 304, 317–321. <https://doi.org/10.1007/s004410100367>

664 Clements, J., Hens, K., Francis, C., Schellens, A., Callaerts, P., 2008. Conserved role for the
665 *Drosophila* Pax6 homolog Eyeless in differentiation and function of insulin-producing neurons.
666 *Proc. Natl. Acad. Sci. U. S. A.* 105, 16183–16188. <https://doi.org/10.1073/pnas.0708330105>

667 Colombani, J., Andersen, D.S., Boulan, L., Boone, E., Romero, N., Virolle, V., Texada, M., Léopold,
668 P., 2015. *Drosophila* Lgr3 couples organ growth with maturation and ensures developmental
669 stability. *Curr. Biol.* 25, 2723–2729. <https://doi.org/10.1016/j.cub.2015.09.020>

670 Colombani, J., Andersen, D.S., Léopold, P., 2012. Secreted peptide dilp8 coordinates *Drosophila*
671 tissue growth with developmental timing. *Science* 336, 582–585.
672 <https://doi.org/10.1126/science.1216689>

673 Cong, X., Wang, H., Liu, Z., He, C., An, C., Zhao, Z., 2015. Regulation of sleep by insulin-like

674 peptide system in *Drosophila melanogaster*. *Sleep* 38, 1075-1083A.
675 <https://doi.org/10.5665/sleep.4816>

676 Cryan, J.R., Urban, J.M., 2012. Higher-level phylogeny of the insect order Hemiptera: Is
677 Auchenorrhyncha really paraphyletic? *Syst. Entomol.* 37, 7–21. <https://doi.org/10.1111/j.1365-3113.2011.00611.x>

678

679 Ding, B.Y., Shang, F., Zhang, Q., Xiong, Y., Yang, Q., Niu, J.Z., Smagghe, G., Wang, J.J., 2017.
680 Silencing of two insulin receptor genes disrupts nymph-adult transition of alate brown citrus
681 aphid. *Int. J. Mol. Sci.* 18, 10–12. <https://doi.org/10.3390/ijms18020357>

682 Duckert, P., Brunak, S., Blom, N., 2004. Prediction of proprotein convertase cleavage sites. *Protein*
683 *Eng. Des. Sel.* 17, 107–112. <https://doi.org/10.1093/protein/gzh013>

684 Emlen, D.J., Warren, I.A., Johns, A., Dworkin, I., Lavine, L.C., 2012. A mechanism of extreme
685 growth and reliable signaling in sexually selected ornaments and weapons. *Science* 337, 860–
686 864. <https://doi.org/10.1126/science.1224286>

687 Garelli, A., Gontijo, A.M., Miguela, V., Caparros, E., Dominguez, M., 2012. Imaginal discs secrete
688 insulin-like peptide 8 to mediate plasticity of growth and maturation. *Science* 336, 579–582.
689 <https://doi.org/10.1126/science.1216735>

690 Grantham, M.E., Shingleton, A.W., Dudley, E., Brisson, J.A., 2020. Expression profiling of winged-
691 and wingless-destined pea aphid embryos implicates insulin/insulin growth factor signaling in
692 morph differences. *Evol. Dev.* 22, 257–268. <https://doi.org/10.1111/ede.12326>

693 Green, D.A.2nd, Extavour, C.G., 2014. Insulin signalling underlies both plasticity and divergence of a
694 reproductive trait in *Drosophila*. *Proc. R. Soc. B Biol. Sci.* 281, 20132673.
695 <https://doi.org/10.1098/rspb.2013.2673>

696 Grönke, S., Clarke, D.F., Broughton, S., Andrews, T.D., Partridge, L., 2010. Molecular evolution and
697 functional characterization of *Drosophila* insulin-like peptides. *PLoS Genet.* 6, e1000857.
698 <https://doi.org/10.1371/journal.pgen.1000857>

699 Guo, S.S., Zhang, M., Liu, T.X., 2016. Insulin-related peptide 5 is involved in regulating embryo
700 development and biochemical composition in pea aphid with wing polyphenism. *Front. Physiol.*
701 7, 31. <https://doi.org/10.3389/fphys.2016.00031>

702 Hancock, J.T., 2010. Cell signalling, 3rd ed. Oxford University Press.

703 Huybrechts, J., Bonhomme, J., Minoli, S., Prunier-Leterme, N., Dombrowsky, A., Abdel-Latif, M.,
704 Robichon, A., Veenstra, J.A., Tagu, D., 2010. Neuropeptide and neurohormone precursors in the
705 pea aphid, *Acyrtosiphon pisum*. *Insect Mol. Biol.* 19, 87–95. <https://doi.org/10.1111/j.1365-2583.2009.00951.x>

706

707 Ikeya, T., Galic, M., Belawat, P., Nairz, K., Hafen, E., 2002. Nutrient-dependent expression of
708 insulin-like peptides from neuroendocrine cells in the CNS contributes to growth regulation in
709 *Drosophila*. *Curr. Biol.* 12, 1293–1300. [https://doi.org/10.1016/S0960-9822\(02\)01043-6](https://doi.org/10.1016/S0960-9822(02)01043-6)

710 International Aphid Genomics Consortium, T., 2010. Genome sequence of the pea aphid

711 *Acyrtosiphon pisum*. PLoS Biol. 8, e1000313. <https://doi.org/10.1371/journal.pbio.1000313>

712 Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J.,
713 Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong,
714 S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: Genome-scale protein function classification.
715 Bioinformatics 30, 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>

716 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2015.
717 ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 32, 587–
718 589. <https://doi.org/10.1093/molbev/msu300>

719 Kannan, K., Fridell, Y.-W.C., 2013. Functional implications of *Drosophila* insulin-like peptides in
720 metabolism, aging, and dietary restriction. Front. Physiol. 4, 288.
721 <https://doi.org/10.3389/fphys.2013.00288>

722 Kollmann, M., Minoli, S., Bonhomme, J., Homberg, U., Schachtner, J., Tagu, D., Anton, S., 2011.
723 Revisiting the anatomy of the central nervous system of a hemimetabolous model insect species:
724 The pea aphid *Acyrtosiphon pisum*. Cell Tissue Res. 343, 343–355.
725 <https://doi.org/10.1007/s00441-010-1099-9>

726 Kondo, H., Ino, M., Suzuki, A., Ishizaki, H., Iwami, M., 1996. Multiple gene copies for bombyxin, an
727 insulin-related peptide of the silkworm *Bombyx mori*: Structural signs for gene rearrangement
728 and duplication responsible for generation of multiple molecular forms of bombyxin. J. Mol.
729 Biol. 259, 926–937. <https://doi.org/10.1006/jmbi.1996.0370>

730 Krieger, M.J.B., Jahan, N., Riehle, M.A., Cao, C., Brown, M.R., 2004. Molecular characterization of
731 insulin-like peptide genes and their expression in the African malaria mosquito, *Anopheles*
732 *gambiae*. Insect Mol. Biol. 13, 305–315. <https://doi.org/10.1111/j.0962-1075.2004.00489.x>

733 Lageux, M., Lwoff, L., Meister, M., Goltzené, F., Hofmann, J.A., 1990. cDNAs from neurosecretory
734 cells of brains of *Locusta migratoria* (Insecta, Orthoptera) encoding a novel member of the
735 superfamily of insulins. Eur. J. Biochem. 187, 249–254. <https://doi.org/10.1111/j.1432-1033.1990.tb15302.x>

736

737 Li, Y., Park, H., Smith, T.E., Moran, N.A., Singh, N., 2019. Gene family evolution in the pea aphid
738 based on chromosome-level genome assembly. Mol. Biol. Evol. 36, 2143–2156.
739 <https://doi.org/10.1093/molbev/msz138>

740 Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Bretscher, A., Ploegh, H., Amon, A., Martin, K.C.,
741 2016. Molecular cell biology, 8th ed. W.H. Freeman.

742 Miguel-Aliaga, I., Thor, S., Gould, A.P., 2008. Postmitotic specification of *Drosophila* insulinergic
743 neurons from pioneer neurons. PLoS Biol. 6, e58. <https://doi.org/10.1371/journal.pbio.0060058>

744 Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, Akinori, Ishizaki, H., Mizoguchi, A.,
745 Fujiwara, Y., Suzuki, Atsushi, 1984. Amino-terminal amino acid sequence of the silkworm
746 prothoracicotropic hormone: Homology with insulin. Science 226, 1344–1345.
747 <https://doi.org/10.1126/science.226.4680.1344>

748 Nagata, K., Maruyama, K., Kojima, K., Yamamoto, M., Tanaka, M., Kataoka, H., Nagasawa, H.,
749 Isogai, A., Ishizaki, H., Suzuki, A., 1999. Prothoracicotropic activity of SBRPs, the insulin-like
750 peptides of the saturniid silkworm *Samia cynthia ricini*. *Biochem. Biophys. Res. Commun.* 266,
751 575–578. <https://doi.org/10.1006/bbrc.1999.1865>

752 Nässel, D.R., Broeck, J. Vanden, 2016. Insulin/IGF signaling in *Drosophila* and other insects: Factors
753 that regulate production, release and post-release action of the insulin-like peptides. *Cell. Mol.*
754 *Life Sci.* 73, 271–290. <https://doi.org/10.1007/s00018-015-2063-3>

755 Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective
756 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–
757 274. <https://doi.org/10.1093/molbev/msu300>

758 Nijhout, H.F., Grunert, L.W., 2002. Bombyxin is a growth factor for wing imaginal disks in
759 lepidoptera. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15446–15450.
760 <https://doi.org/10.1073/pnas.242548399>

761 Okamoto, Naoki ; Yamanaka, Naoki; Yoshimasa, Yagi; Nishida, Yasuyoshi; Kataoka, Hiroshi;
762 O'Connor, Michael B.; Mizoguchi, A., 2009a. A Fat Body-Derived IGF-like Peptide Regulates
763 Postfeeding Growth in *Drosophila*. *Dev Cell* 17, 885–891.
764 <https://doi.org/10.1016/j.physbeh.2017.03.040>

765 Okamoto, N., Yamanaka, N., Satake, H., Saegusa, H., Kataoka, H., Mizoguchi, A., 2009b. An
766 ecdysteroid-inducible insulin-like growth factor-like peptide regulates adult development of the
767 silkworm *Bombyx mori*. *FEBS J.* 276, 1221–1232. [https://doi.org/10.1111/j.1742-](https://doi.org/10.1111/j.1742-4658.2008.06859.x)
768 [4658.2008.06859.x](https://doi.org/10.1111/j.1742-4658.2008.06859.x)

769 Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR.
770 *Nucleic Acids Res.* 29, e45. <https://doi.org/10.1093/nar/29.9.e45>

771 Pfaffl, M.W., Horgan, G.W., Dempfle, L., 2002. Relative expression software tool (REST) for group-
772 wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic*
773 *Acids Res.* 30, e36. <https://doi.org/10.1093/nar/30.9.e36>

774 Pfaffl, M.W., Tichopad, A., Prgomet, C., Neuvians, T., 2004. Determination of most stable
775 housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper.
776 *Biotechnol. Lett.* 26, 509–515.

777 Raabe, M., 1986. Comparative immunocytochemical study of release sites of insulin, glucagon and
778 AKH-like products in *Locusta migratoria*, *Periplaneta americana*, and *Carausius morosus*. *Cell*
779 *Tissue Res.* 245, 267–271. <https://doi.org/10.1007/BF00213931>

780 Raven, P.H., Johnson, G.B., Mason, K.A., Losos, J.B., Singer, S.R., 2015. *Biology*, 11th ed.
781 McGraw-Hill.

782 Ribeiro Lopes, M., Parisot, N., Gaget, K., Huygens, C., Peignier, S., Duport, G., Orlans, J., Charles,
783 H., Baatsen, P., Jouselin, E., Silva, P. Da, Hens, K., Callaerts, P., Calevro, F., 2020.
784 Evolutionary novelty in the apoptotic pathway of aphids. *Proc. Natl. Acad. Sci. U. S. A.* 117,

785 32545–32556. <https://doi.org/10.1073/pnas.2013847117>

786 Rispe, C., Legeai, F., Nabity, P.D., Fernández, R., Arora, A.K., Baa-Puyoulet, P., Banfill, C.R., Bao,
787 L., Barberà, M., Bouallègue, M., Bretaudeau, A., Brisson, J.A., Calevro, F., Capy, P., Catrice,
788 O., Chertemps, T., Couture, C., Delière, L., Douglas, A.E., Dufault-Thompson, K., Escuer, P.,
789 Feng, H., Forneck, A., Gabaldón, T., Guigó, R., Hilliou, F., Hinojosa-Alvarez, S., Hsiao, Y.M.,
790 Hudaverdian, S., Jacquin-Joly, E., James, E.B., Johnston, S., Joubard, B., Le Goff, G., Le
791 Trionnaire, G., Librado, P., Liu, S., Lombaert, E., Lu, H.L., Maïbèche, M., Makni, M., Marcet-
792 Houben, M., Martínez-Torres, D., Meslin, C., Montagné, N., Moran, N.A., Papura, D., Parisot,
793 N., Rahbé, Y., Lopes, M.R., Ripoll-Cladellas, A., Robin, S., Roques, C., Roux, P., Rozas, J.,
794 Sánchez-Gracia, A., Sánchez-Herrero, J.F., Santesmasses, D., Scatoni, I., Serre, R.F., Tang, M.,
795 Tian, W., Umina, P.A., Van Munster, M., Vincent-Monégat, C., Wemmer, J., Wilson, A.C.C.,
796 Zhang, Y., Zhao, C., Zhao, J., Zhao, S., Zhou, X., Delmotte, F., Tagu, D., 2020. The genome
797 sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion
798 routes of an iconic pest. *BMC Biol.* 18, 90. <https://doi.org/10.1186/s12915-020-00820-5>

799 Robert, X., Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript
800 server. *Nucleic Acids Res.* 42, 320–324. <https://doi.org/10.1093/nar/gku316>

801 Sapountzis, P., Duport, G., Balmand, S., Gaget, K., Jaubert-Possamai, S., Febvay, G., Charles, H.,
802 Rahbé, Y., Colella, S., Calevro, F., 2014. New insight into the RNA interference response
803 against cathepsin-L gene in the pea aphid, *Acyrtosiphon pisum*: Molting or gut phenotypes
804 specifically induced by injection or feeding treatments. *Insect Biochem. Mol. Biol.* 51, 20–32.
805 <https://doi.org/10.1016/j.ibmb.2014.05.005>

806 Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., Ishikawa, H., 2000. Genome sequence of the
807 endocellular bacterial symbiont of aphids *Buchnera* sp. *APS. Nature* 407, 81–86.
808 <https://doi.org/10.1038/35024074>

809 Sim, C., Denlinger, D.L., 2008. Insulin signaling and FOXO regulate the overwintering diapause of
810 the mosquito *Culex pipiens*. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6777–6781.
811 <https://doi.org/10.1073/pnas.0802067105>

812 Simonet, P., Duport, G., Gaget, K., Weiss-Gayet, M., Colella, S., Febvay, G., Charles, H., Viñuelas,
813 J., Heddi, A., Calevro, F., 2016. Direct flow cytometry measurements reveal a fine-tuning of
814 symbiotic cell dynamics according to the host developmental needs in aphid symbiosis. *Sci.*
815 *Rep.* 6, 19967. <https://doi.org/10.1038/srep19967>

816 Simonet, P., Gaget, K., Balmand, S., Ribeiro Lopes, M., Parisot, N., Buhler, K., Duport, G., Vulsteke,
817 V., Febvay, G., Heddi, A., Charles, H., Callaerts, P., Calevro, F., 2018. Bacteriocyte cell death
818 in the pea aphid/ *Buchnera* symbiotic system. *Proc. Natl. Acad. Sci. U. S. A.* 115, E1819–E1828.
819 <https://doi.org/10.1073/pnas.1720237115>

820 Slaidina, M., Delanoue, R., Gronke, S., Partridge, L., Léopold, P., 2009. A *Drosophila* Insulin-like
821 Peptide promotes growth during nonfeeding states. *Dev. Cell* 17, 874–884.

822 <https://doi.org/10.1016/j.devcel.2009.10.009>

823 Smit, A.B., Van Kesteren, R.E., Li, K.W., Van Minnen, J., Spijker, S., Van Heerikhuizen, H.,
824 Geraerts, W.P.M., 1998. Towards understanding the role of insulin in the brain: Lessons from
825 insulin-related signaling systems in the invertebrate brain. *Prog. Neurobiol.* 54, 35–54.
826 [https://doi.org/10.1016/S0301-0082\(97\)00063-4](https://doi.org/10.1016/S0301-0082(97)00063-4)

827 Smykal, V., Pivarci, M., Provaznik, J., Bazalova, O., Jedlicka, P., Luksan, O., Horak, A., Vaneckova,
828 H., Benes, V., Fiala, I., Hanus, R., Dolezel, D., 2020. Complex evolution of insect insulin
829 receptors and homologous decoy receptors, and functional significance of their multiplicity.
830 *Mol. Biol. Evol.* 37, 1775–1789. <https://doi.org/10.1093/molbev/msaa048>

831 Steiner, D., 1985. Structure and Evolution of the Insulin gene. *Annu. Rev. Genet.* 19, 463–484.
832 <https://doi.org/10.1146/annurev.genet.19.1.463>

833 Vafopoulou, X., Steel, C.G.H., 2014. Synergistic induction of the clock protein PERIOD by insulin-
834 like peptide and prothoracicotropic hormone in *Rhodnius prolixus* (Hemiptera): Implications for
835 convergence of hormone signaling pathways. *Front. Physiol.* 5, 41.
836 <https://doi.org/10.3389/fphys.2014.00041>

837 Veenstra, J.A., Agricola, H.J., Sellami, A., 2008. Regulatory peptides in fruit fly midgut. *Cell Tissue*
838 *Res.* 334, 499–516. <https://doi.org/10.1007/s00441-008-0708-3>

839 Veenstra, J.A., 2020. Arthropod IGF, relaxin and gonadulin, putative orthologs of *Drosophila* insulin-
840 like peptides 6, 7 and 8, likely originated from an ancient gene triplication. *PeerJ*, 8:e9534.
841 <http://doi:10.7717/peerj.9534>.

842 Wheeler, D.E., Buck, N., Evans, J.D., 2006. Expression of insulin pathway genes during the period of
843 caste determination in the honey bee, *Apis mellifera*. *Insect Mol. Biol.* 15, 597–602.
844 <https://doi.org/10.1111/j.1365-2583.2006.00681.x>

845 Williams, K.D., Busto, M., Suster, M.L., So, A.K.C., Ben-Shahar, Y., Leever, S.J., Sokolowski,
846 M.B., 2006. Natural variation in *Drosophila melanogaster* diapause due to the insulin-regulated
847 PI3-kinase. *Proc. Natl. Acad. Sci. U. S. A.* 103, 15911–15915.
848 <https://doi.org/10.1073/pnas.0604592103>

849 Wu, Q., Brown, M.R., 2006. Signaling and function of Insulin-Like Peptides in insects. *Annu. Rev.*
850 *Entomol.* 51, 1–24. <https://doi.org/10.1146/annurev.ento.51.110104.151011>

851 Wu, Q., Zhang, Y., Xu, J., Shen, P., 2005. Regulation of hunger-driven behaviors by neural ribosomal
852 S6 kinase in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13289–13294.
853 <https://doi.org/10.1073/pnas.0501914102>

854 Yang, C.H., Belawat, P., Hafen, E., Jan, L.Y., Jan, Y.N., 2008. *Drosophila* egg-laying site selection as
855 a system to study simple decision-making processes. *Science* 319, 1679–1683.
856 <https://doi.org/10.1126/science.1151842>

857 Zheng, S., Chiu, H., Boudreau, J., Papanicolaou, T., Bendena, W., Chin-Sang, I., 2018. A functional
858 study of all 40 *Caenorhabditis elegans* insulin-like peptides. *J. Biol. Chem.* 293, 16912–16922.

860

861 **Figures and Tables**

862 **Figure 1. Schematic representation of the predicted structure of *Acyrtosiphon pisum* insulin**
863 **related peptides (IRPs) and comparison with known insulin- and IGF-like peptides.**

864 Schematic representation and alignments of *A. pisum* IRP1-IRP4 (A), IRP5-6 (B) and IRP11
865 (C). Domains are denoted by letters. The spaces between domains represent predicted
866 proteolytic cleavage during maturation of the propeptide. Disulfide bonds between conserved
867 cysteines are indicated. Amino acid sequences of the predicted IRPs from *A. pisum* are
868 aligned. Amino acid residues are highlighted in red when fully conserved and written in red
869 when partially conserved. Color bars below the alignment indicate the predicted domains in
870 the precursor peptides: green, signal peptide; orange, B chain; yellow, C chain; blue, A chain.
871 Asterisks denote cysteine residues, and paired triangles denote canonical prohormone
872 convertase or furin cleavage sites (dibasic amino acids).

873 **Figure 2. Genomic organization of *Acyrtosiphon pisum* insulin related peptides (IRPs) mapping**
874 **on the A1, A3 and X chromosomes.** Genomic DNA (GenBank accession number

875 GCF_005508785.1) is represented by a line (top) with distances in kb. White and grey boxes
876 indicate predicted introns and predicted exons respectively. In the case of IRP5 and IRP6, the
877 two predicted transcripts are represented. Positions of the primers used for the (qRT-)PCR
878 experiments are indicated as red arrows above each transcript.

879 **Figure 3. Maximum likelihood tree showing relationships among IRP amino acid sequences**
880 **from different aphid species.** Unrooted tree of IRP sequences from 10 aphid species and the

881 grape phylloxera, *Daktulosphaira vitifoliae*, an evolutionary ancient aphid-related species.
882 Sequence names are indicated as a prefix formed by the abbreviated species name
883 (*Acyrtosiphon pisum*, Ap; *Aphis glycines*, Agl; *Aphis gossypii*, Ago; *Daktulosphaira*
884 *vitifoliae*, Dv; *Diuraphis noxia*, Dn; *Melanaphis sacchari*, Ms; *Myzus cerasi*, Mc; *Myzus*
885 *persicae*, Mp; *Rhopalosiphum maidis*, Rm; *Rhopalosiphum padi*, Rp; *Sipha flava*, Sf),
886 followed by the IRP common name in the case of the pea aphid (sequences in red) or protein
887 accession number for other species. The putative IRP homolog groups identified in this study,
888 referred to in the text as IRP subfamilies, are highlighted in different colors with a letter
889 corresponding to each subfamily. All the displayed nodes have a bootstrap value above 80%.

890 **Figure 4. Localization of IRP1 and IRP4 in adult pea aphid brain by immunohistochemistry.**

891 Images correspond to confocal Z-stack with tubulin in green and target IRP in magenta. For
892 each antibody 15 aphid brains were analyzed. (A) Schematic representation of the central
893 nervous system of the pea aphid adapted from Kollmann et al. (2011). (B) and (C) expression
894 of IRP1 and IRP4 respectively in eight neurons of the pars intercerebralis with OL, optic lobe;

895 PC, protocerebrum; PI, pars intercerebralis; SEG, subesophageal ganglion; TGM, thoracic
896 ganglionic mass. A closer view of the cells is shown on the right side of the figure.

897 **Figure 5. Localization of IRP2 in adult pea aphid digestive tract by immunohistochemistry.**

898 Images correspond to confocal Z-stack with tubulin in green and IRP2 in magenta. In total 6
899 digestive tracts were analyzed. (A) Tubulin staining of the intestinal tract to indicate the
900 boundaries and structure of the intestinal tract. (B) IRP2 is expressed at low levels in circular
901 structures of the intestinal tract.

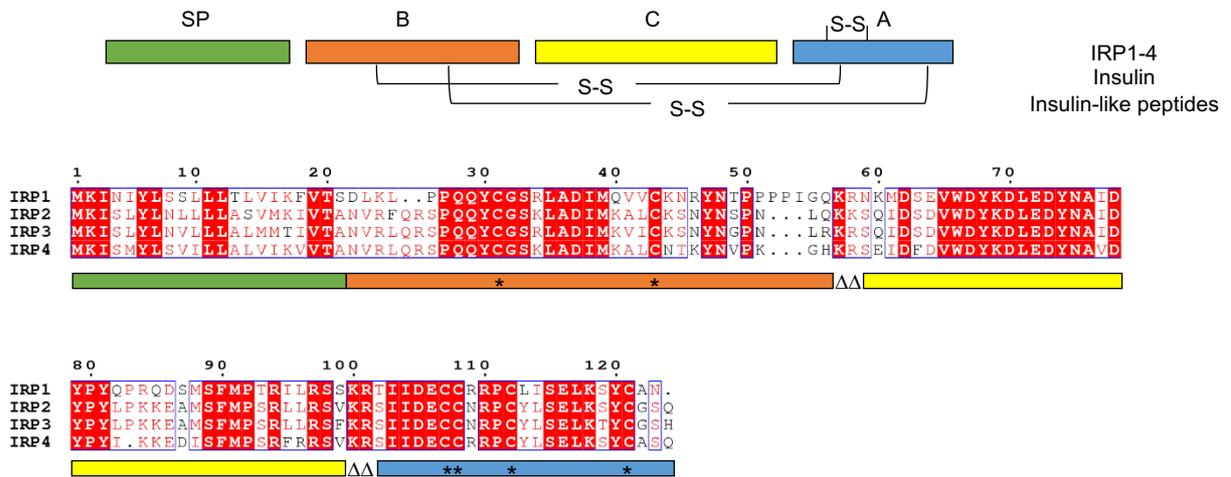
902 **Table 1. Putative IRP-encoding genes in the *A. pisum* genome**

903 **Table 2. Similarity between *A. pisum* IRP protein sequences.**

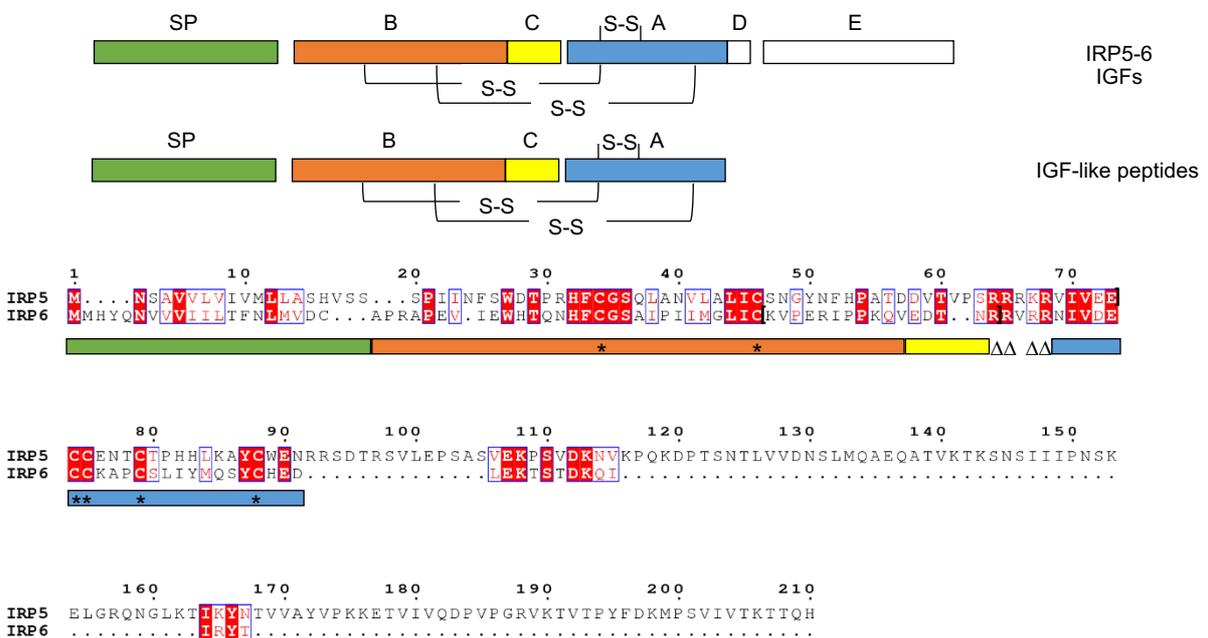
904 **Table 3. Expression of *A. pisum* IRPs in publicly available RNA-seq libraries corresponding to
905 different tissues and morphs.**

906 **Table 4. qRT-PCR analyses of *A. pisum* IRPs expression within various tissues relative to whole
907 body.**

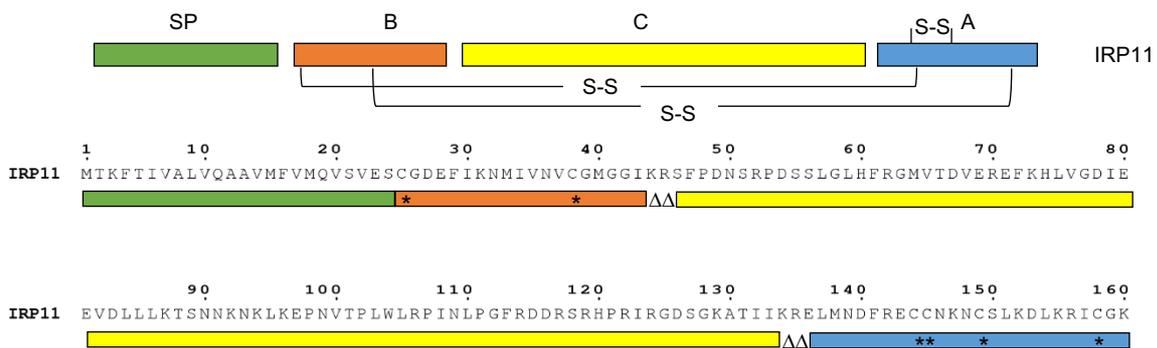
A)

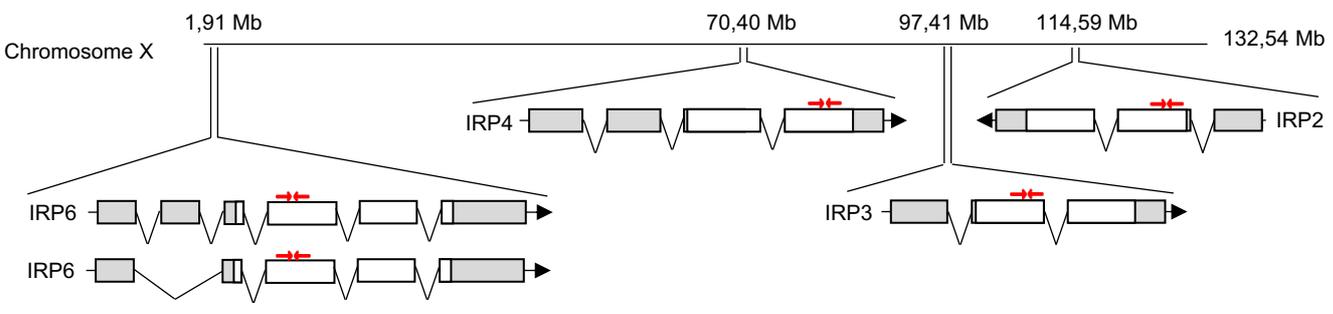
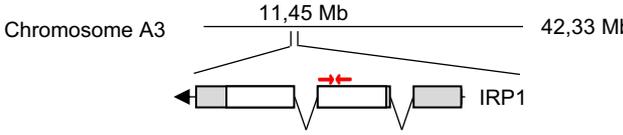
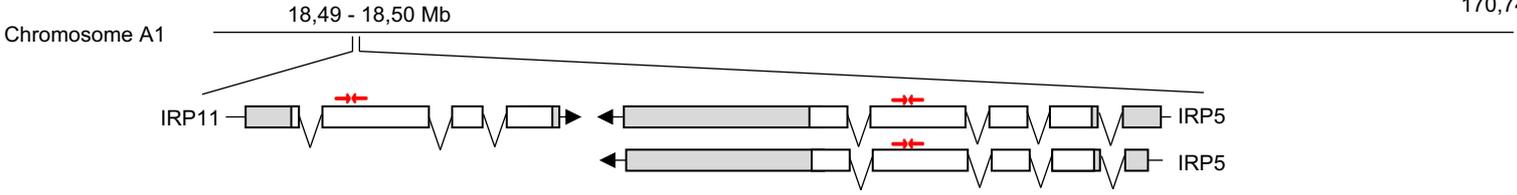


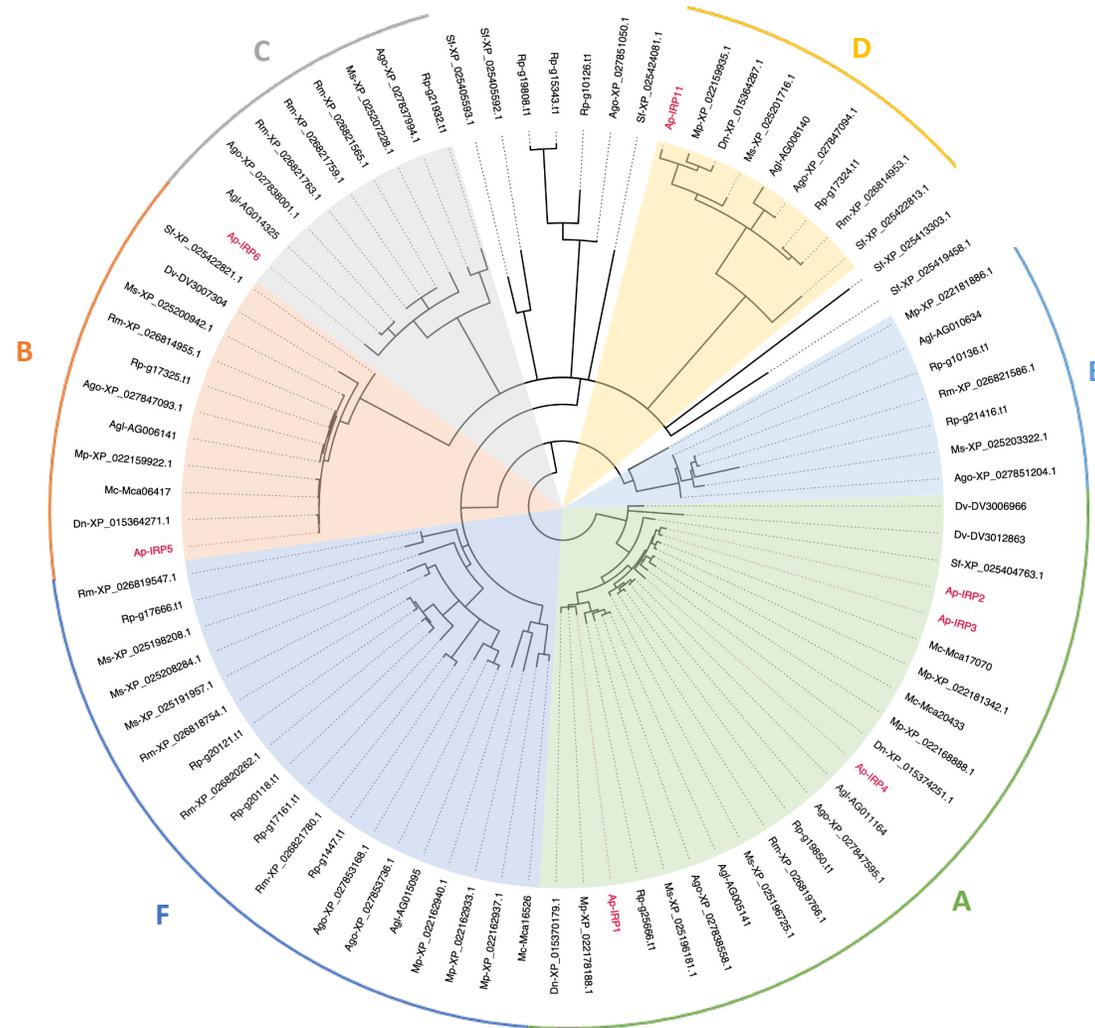
B)



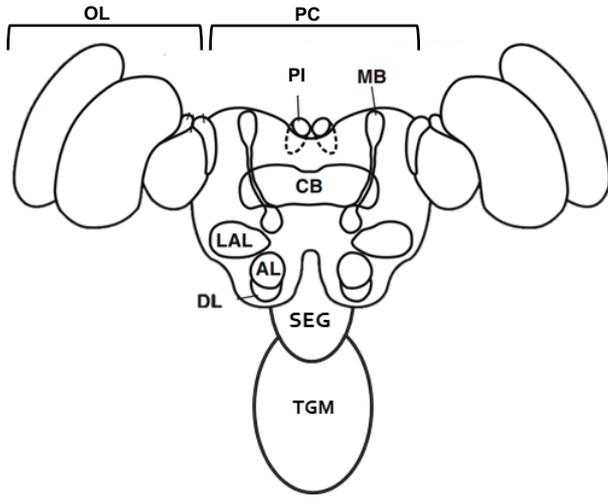
C)



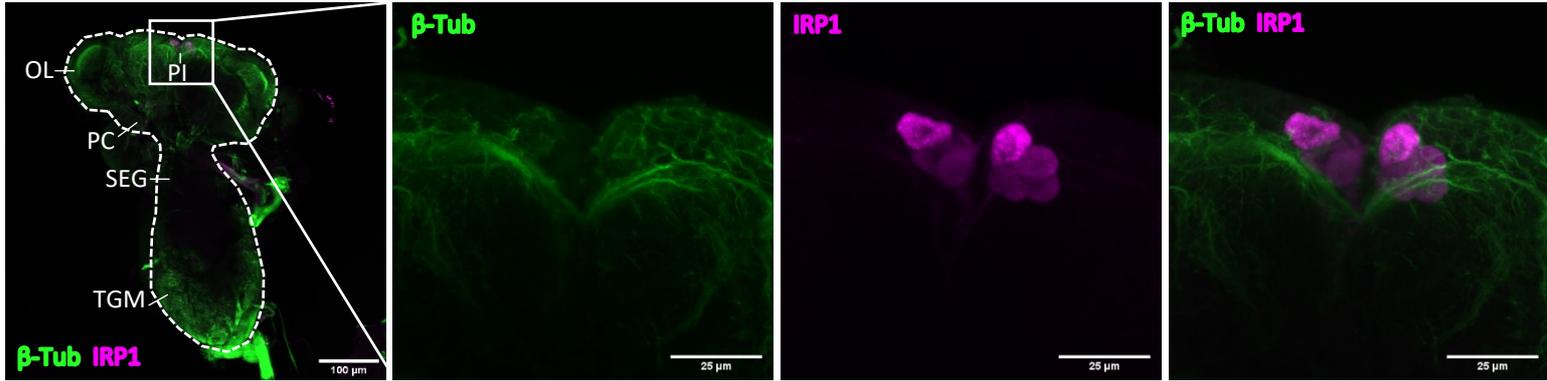
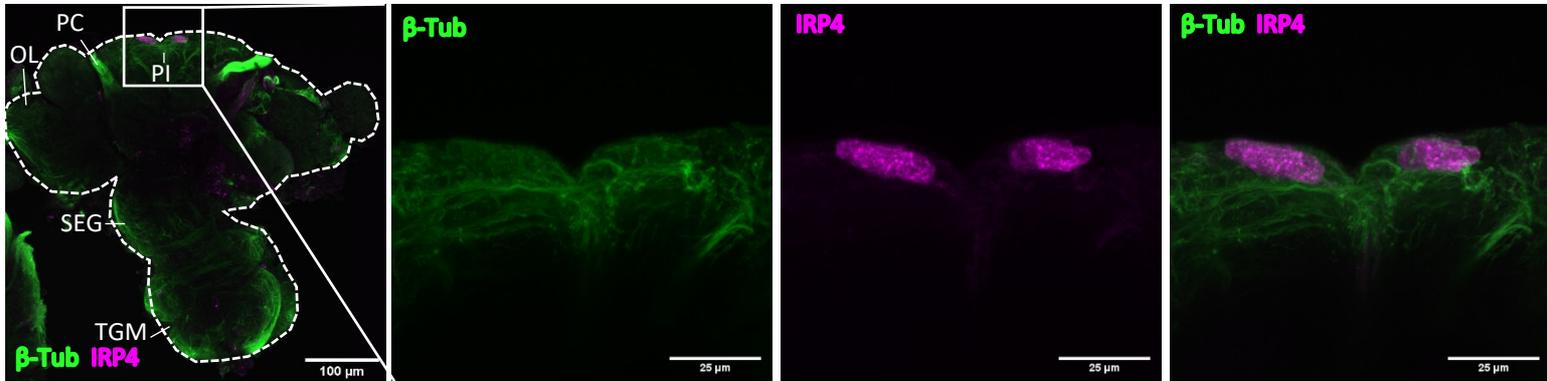


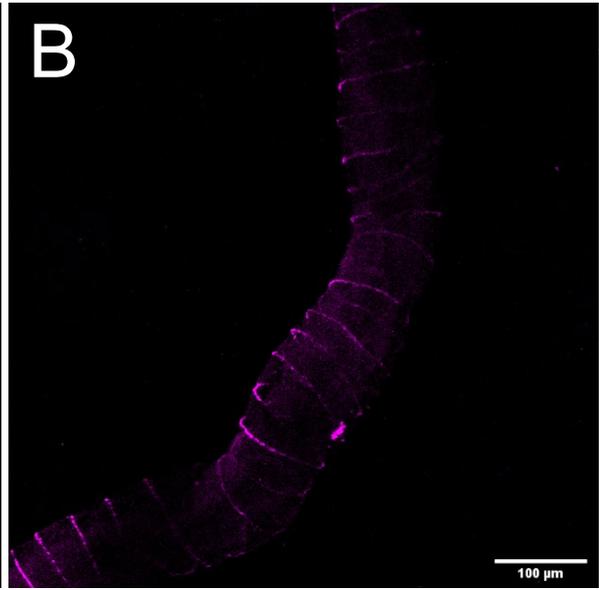
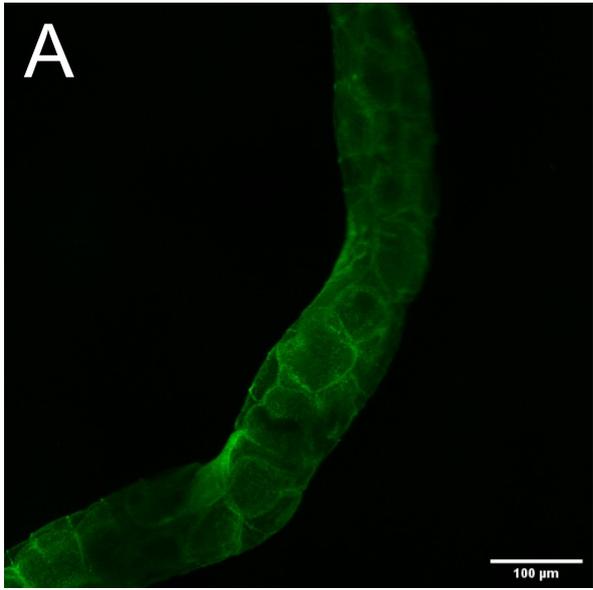


0.6

A

OL - Optic Lobe
 PC - Protocerebrum
 MB - Mushroom Body
 PI - Pars Intercerebralis
 CB - Central Body
 LAL - Lateral Accessory Lobe
 AL - Antennal Lobe
 DL - Dorsal Lobe
 SEG - Subesophageal Ganglion
 TGM - Thoracic Ganglionic Mass

B**C**



Protein name	Locus identifier	Protein identifier Annotation Release 103	Length (AA)	Chromosome	Exons
IRP1	LOC100568938	XP_003247548.1	123	A3	3
IRP2	LOC100574788	XP_003244126.1	123	X	3
IRP3	LOC100575361	XP_003240930.1	123	X	3
IRP4	LOC100161832	XP_001949438.1	122	X	4
IRP5	LOC100169635	XP_029342630.1	210	A1	5
		XP_001949253.1	210	A1	5
IRP6	LOC100570058	XP_016663396.1	106	X	6
		XP_003240733.1	106	X	5
IRP11	LOC100573276	XP_003246343.1	161	A1	4

	IRP1		IRP2		IRP3		IRP4		IRP5		IRP6		IRP11	
	Id	Cov	Id	Cov										
IRP1	100	100	62	99	63	99	63	99	26	28	27	52	36	20
IRP2	62	100	100	100	89	100	78	100	26	29	29	52	41	18
IRP3	63	100	89	100	100	100	75	100	27	29	28	52	59	11
IRP4	63	100	78	100	75	100	100	100	27	29	25	78	34	18
IRP5	26	77	25	96	27	76	27	75	100	100	29	92		
IRP6	27	76	29	73	25	93	25	87	29	54	100	100		
IRP11	36	20	41	21	59	14	34	21					100	100

Note : identity (Id) and coverage (cov) percentages are indicated, as determined by systematic BLASTP analysis between (column) and subject (lines).

	Male (Library ID: SRR924119)	Sexual female (Library ID: SRR924121)	Asexual female (Library ID: SRR924118)	Winged (Library ID: SRR5045469)	Nymph (Library ID: SRR1793299)	Embryos (Library ID: SRR7454536)	Bacteriocytes (Library ID: SRR073576)	Digestive tract (Library ID: SRR7037540)	Head (Library ID: SRR353539)	24h head solitary (Library ID: SRR074231)	24h head crowded (Library ID: SRR074233)	Salivary glands (Library ID: SRR7037544)
IRP1	++	+	++	++	++	+		+	+++	+++	+++	+++
IRP2	++										+	
IRP3	++											+
IRP4	++++	+++	+++	+++	++	++	++	++	++++	++++	++++	+++
IRP5	+++	+++	+++	+++	+++	+++	+++	+++	++++	++++	++++	+++
IRP6	++	++	++	+	+		++	+			+	+
IRP11	+	++	+++	+++	++	+		++		+	+	++

Note : RNAseq libraries were downloaded from the NCBI SRA database. Expression values are ranked based on log2 of the Transcript Count Per Million (+ [0,15-0,5] ; ++ [0,5-1] ; +++ [1-2] ; ++++ [>2]). Cells marked in grey indicate that no expression of the corresponding IRP was observed in this library.

	IRP 1		IRP 2		IRP 4		IRP 5		IRP 6		IRP11	
	Relative expression (SD)	Variation										
Embryonic chains	0,999 (±0,0690)	-	0,41 (±0,0840)	-	0,678 (±0,0442)	-	0,827 (±0,1025)	-	0,641 (±0,1305)	-	1,347 (±0,15671)	-
Bacteriocytes	0,448 (±0,0477)	-	0,432 (±0,1600)	-	0,550 (±0,1204)	-	1,849 (±0,2298)	-	2,091 (±0,5452)	-	0,080 (±0,0197)	-
Carcass	0,173 (±58,830)	-	4,412 (±1,0641)	↑	11,025 (±0,9499)	↑	7,281 (±0,7758)	↑	17,028 (±3,7747)	↑	9,5780 (±0,9068)	↑
Brain	323,704 (±58,830)	↑	39,356 (±10,327)	↑	158,511 (±30,471)	↑	8,893 (±1,6748)	-	1,753 (±0,5262)	-	0,277 (±0,0723)	-
Fat Body	0,317 (±0,1268)	-	0,957 (±0,4193)	-	2,968 (±0,9214)	-	17,491 (±5,9433)	↑	17,487 (±8,3910)	↑	0,114 (±0,0363)	-
Digestive tract	0,115 (±0,0247)	↓	0,228 (±0,0795)	-	0,122 (±0,0316)	↓	0,891 (±0,1101)	-	1,060 (±0,1947)	-	0,157 (±0,0179)	↓

Note : Gene expression levels are expressed relative to the expression in 9 days-old *A. pisum* whole body. The rpl7 gene was used for data normalization. Results are reported as mean ± SD from three independent experiments and variation (-: non-significant; ↑: enriched expression; ↓: reduced expression) is given for each IRP.

Aphid IRPs

