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

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

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

Huygens C.[1,2*], Ribeiro Lopes M.[2*], Gaget K.[2*], Duport G.[2], Peignier S [2]., De Groef S.[1], Parisot N.[2], Calevro F.[2], Callaerts P.[1]

Affiliation:
1. Laboratory of Behavioral and Developmental Genetics, Department of Human Genetics, KULeuven, University of Leuven, B-3000 Leuven, Belgium
2. Univ Lyon, INSA Lyon, INRAE, BF2I, UMR 203, 69621 Villeurbanne, France

* Joint first authors

Authors for correspondence:

Federica Calevro: federica.calevro@insa-lyon.fr
Patrick Callaerts: patrick.callaerts@kuleuven.be

Abstract

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

Keywords: Hemiptera, aphids, *Acyrthosiphon pisum,* insulin, IRPs

1. Introduction

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

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

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

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

To start to address how IRPs evolve and contribute to functional diversification in processes regulated by the insulin signaling pathway, we focus in the current study on aphids. These sap-sucking insects are a very speciose group of which many are pest insects (Calevro et al., 2019). Aphids belong to the order Hemiptera together with cicadas, planthoppers, leafhoppers, shield bugs, and whiteflies (Cryan and Urban, 2012). Aphids thrive on phloem sap, a nutritionally unbalanced diet, by virtue of the evolutionarily ancient symbiotic relationship with the obligatory bacterial endosymbiont, the γ3-proteobacterium *Buchnera aphidicola* (Akman Gündüz and Douglas, 2012; Baumann et al., 1995; Shigenobu et al., 2000). These insects represent an excellent paradigm for studying evolutionary changes in the insulin signaling pathway for three broad reasons. First, very little is known about the role of the insulin-signaling pathway in aphid physiology, metabolism, and reproduction. Second, aphids display remarkable phenotypic plasticity with winged and non-winged morphs as well as asexual and sexual reproduction, processes that could be controlled by different IRPs (Barberà et al., 2019; Grantham et al., 2020; Guo et al., 2016). Third, the availability of a number of aphid genome sequences makes it possible to perform evolutionary analyses (Ribeiro Lopes et al., 2020) and study diversification of insulin-related peptides. Given that many insects, including aphid species, have two InR encoding genes (Ding et al., 2017; International Aphid Genomics Consortium, 2010; Smykal et al., 2020), we hypothesize that much of the expected functional diversification comes from the ligands, the insulin-related peptides.

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

2. Materials and methods

2.1 Aphid IRP identification

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

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

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

2.2 Phylogenetic reconstruction

Homology relationships between aphid IRPs were further inferred by phylogenetic reconstruction. Candidate sequences were aligned using the MAFFT multiple alignment program v7.0 (https://mafft.cbrc.jp/alignment/software/) with the mafft-einsi alignment method. Graphical representations of the alignment results were performed with ESPript v3.0 (Robert and Gouet, 2014). We then selected the C20 model as the best substitution model with ModelFinder (Kalyaanamoorthy et al., 2015), using the Bayesian Information Criterion metric, and built a phylogenetic tree using the maximum likelihood estimation method implemented in IQ-TREE v1.6.2 (Nguyen et al., 2015). The reliability of each branch was evaluated using the bootstrap method with 1000 repetitions, and weakly supported branches (<80%) were collapsed using the TreeCollapseCL 4 software (http://emmahodcroft.com/TreeCollapseCL.html). Graphical representation of the tree was performed 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 log10(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 matrilines. 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 $MgCl_2$, 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

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

2.7 RNA isolation and quality control

Total RNA was extracted as previously described by Simonet et al. (2018), using an RNeasy Mini Kit (Qiagen, Hilden, Germany). In order to get enough RNA, a total of ten aphids per tissue were used for RNA isolation from bacteriocytes, brain, fat body, embryonic chains and gut. Five aphids were used to isolate total RNA from the carcass. For each tissue, three independent biological replicates were processed. In parallel, total RNA was also extracted from the whole body of six aphids (in triplicates) and then used for data normalization in qRT-PCR analysis. Total RNA quality and concentration were checked using gel electrophoresis and a NanoDrop® ND-1000 Spectrophotometer. Samples had to meet the following quality parameters: A260/A280 $\geq$ 1.8 and A260/A230 $\geq$ 1.8, in order to be used for subsequent analyses. To remove any genomic DNA contamination total RNA was treated with DNase I (RQ1 RNase-Free DNase, Promega, Charbonnières-les-bains, France) according to the manufacturer's instructions.

2.8 Reverse Transcription and qRT-PCR experiments

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

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

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

2.10 Immunohistochemistry

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

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

2.11 Statistical analysis

For qRT-PCR experiments, three genes were tested as potential candidates for data normalization: *rpl7* (NP_001129370.1), actin (NP_001119672.1) and *rpl32* (NP_001119682.1). *rpl7* was retained as the best normalization gene based on a preliminary analysis of the data with the BestKeeper software tool (Pfaffl et al., 2004). The expression levels (log2) of the different IRPs genes were then normalized relative to those of the *rpL7* gene. The relative expression ratio R (and the associated standard error) of IRP mRNAs in the different tissues was calculated using as reference the mean of the IRP transcript expression levels of the aphid control group (A9 whole body, see the section 2.7) using the REST software tool (Pfaffl et al., 2002). This ratio was calculated taking into account the real-time PCR efficiency of each gene (E) and the crossing point difference (DCP) of a test condition, as compared to the reference condition, and expressed in comparison to the normalization gene (*rpl7*) using the following model (Pfaffl, 2001): $R=(Etarget)^{\Delta CPtarget(control-sample)}/ (Ereference)^{\Delta CPreference(control-sample)}$. Results are displayed as mean ± SD of three independent biological experiments.

3. Results

3.1 *A. pisum* IRP annotation

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

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

3.2 *A. pisum* IRP sequence analysis

The sequences and genomic organization of the seven *A. pisum* IRPs identified in this paper (IRP1-IRP6 and IRP11) were analyzed in more detail. We confirmed that all these putative IRPs possess the signal peptide, the A and B chains, as well as the six conserved cysteine residues that are found in all insulin-related peptides (Fig. 1) (Badisco et al., 2008; Brogiolo et al., 2001; Kondo et al., 1996; Lageux et al., 1990; Steiner, 1985). The four cysteine residues of the A chain are ordered as $CC(X)_3C(X)_8C$, in which X can be any amino acid, while the two cysteine residues of the B chain are separated by 11 amino acid residues in all IRPs with the exception of IRP11 where they are separated by 12 residues.

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

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

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

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

3.3 IRP identification in other aphids

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

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

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

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

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

3.4 Phylogenetic analysis of aphid IRPs

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

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

We found IRP6 homologs only in the Aphidinae subfamily while four aphids (*D. noxia, M. cerasi, M. persicae* and *S. flava)* and phylloxera have no representative in the C subfamily. Furthermore, like IRP5s, IRP6 homologs all have a very short C-peptide (Supplementary Fig. S3C). The nine proteins of the C subfamily can be divided into two groups of more closely related sequences. The first group includes two proteins (one from *R. padi* and one from *A. gossypi)* that are 71% identical, and the second group comprises seven proteins (three from *R. maidis*; one from M. *sacchari, A. gossypi, A. glycines* and *A. pisum*) with identity percentages ranging from 63% to 86%.

The D subfamily includes all the IRP11 homologs. One homolog of IRP11 was identified in each aphid with the exception of *M. cerasi* but only six were classified as part of the insulin-like superfamily as Agl-AG006140, Ago-XP_027847094.1 and Ms-XP_025201716.1 have no insulin-like signatures. *D. vitifoliae* also lack an IRP11 homolog. Importantly, sequences from this group gave no significant results when aligned against most of the other aphid IRPs. While they all contain the six conserved cysteine residues involved in intra- and inter-chain disulfide bonds, their positions differ from other IRPs as the two cysteines in the B chain are separated by 12 amino acids instead of 11 (Supplementary Fig. S3D). The E subfamily includes proteins from the six aphids that are part of the Aphidini tribe. They all possess a classic A and B chain but no dibasic cleavage site. The F subfamily includes 20 proteins that have an unconventional A chain that possesses nine amino acid residues between the third and fourth cysteine instead of eight.

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

We analyzed the expression profile of *A. pisum* IRPs in several RNA-seq libraries, corresponding to different *A. pisum* morphs (males, sexual females, parthenogenetic females, winged females) and tissues (head, gut, salivary glands, bacteriocytes, embryos) publicly available in the NCBI SRA (Sequence Read Archive) database (Table 3). This analysis highlighted very different expression profiles for the different pea aphid IRPs. IRP4 and IRP5 were found in all RNA-seq libraries, albeit at different levels of expression. IRP5 is by far the most abundant IRP in *A. pisum* and is the most expressed IRP in all the tissues and conditions considered. Two notable exceptions are: (i) IRP4 and IRP5 levels are comparable in the heads of aphids at high population density, a condition that is linked with alate-wingless polyphenism and (ii) IRP4 levels are higher than IRP5 in male aphids. Expression of IRP1 and IRP11 are detected in all morphs and tissues with the exception of bacteriocytes. Moreover, IRP1 is highly expressed in the head. IRP2 is only expressed in males and in the heads of aphids at high population density, and IRP3 only in males and salivary glands. IRP6 is the only IRP other than IRP4 and IRP5 for which expression was detected in the bacteriocytes and it is comparatively poorly expressed in other tissues.

Analysis of the RNA-seq libraries gave us a first indication of the expression pattern of the IRPs, but also revealed that IRP spatiotemporal expression can vary depending on *A. pisum* strain, morph and/or

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

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

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

4. Discussion

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

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

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

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

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

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

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

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

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

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

other aphid species included in this study (subfamily D) but only six, including *A. pisum* IRP11, were classified as part of the insulin-like superfamily. The remaining three proteins lack an insulin signature, probably due to the fact that they either have no A chain (in the case of Agl-AG006140) or have an additional residue in their A chain, with a resulting $CC(X)_4C(X)_8C$ motif. According to RNA-seq and qRT-PCR data, IRP11 is enriched in *A. pisum* carcass suggesting a role in aphid physiology different from all other IRPs. The IRP11 encoding-gene was previously listed as a putative gonadulin, a subclass of insulin-like peptides that share similarities with *Drosophila* ILP8 (Veenstra, 2020). The data presented here confirms that IRP11 belongs to this group since gonadulin ortholog sequences (i) are very divergent not only from other IRPs but also among each other, even in closely related species, (ii) share a common motif, containing 12 amino acid residues between the two cysteines of the B-chain, (iii) generally show lower expression levels in males than females, as is the case in the pea aphid according to the RNA-seq data analyzed here and (iv) IGF-like and gonadulin-like proteins are encoded by genes that are next to each other in the genomes of arthropods, a structural organization that we found for the IRP5 and IRP11 encoding-genes in the pea aphid genome. This, in addition to the fact that IRP11 appears more closely related to IRP5 and IRP6 according to our phylogenetic analysis, suggests that IRP11 and IRP5-like proteins might have evolved from duplication of an ancestral gene and have since undergone multiple substitutions. Importantly, while gonadulin-like proteins are often highly expressed in insect gonads, the pea aphid IRP11 is specifically enriched in the pea aphid carcass, which suggests that this protein could have developed new aphid-specific functions, related to their characteristic features, e.g. polyphenisms and obligatory symbiosis with *B. aphidicola*. This, in addition to the fact that no IRP11 homolog was identified in the aphid sister-group phylloxera, suggests that these IRPs could be the result of an aphid-specific duplication. A possibility, supported by the gene location of IRP11 compared to IRP5, is that IRP11 is the result of an ancestral duplication of IRP5 that has since undergone multiple substitutions. IRP11 could have developed new aphid-specific functions, related to their characteristic features, e.g. polyphenisms, obligatory symbiosis with *B. aphidicola.*

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

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

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

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

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

References

Akman Gündüz, E., Douglas, A.E., 2012. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. Proc. R. Soc. B Biol. Sci. 276, 987–991. https://doi.org/10.1098/rspb.2008.1476

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37, 420–423. https://doi.org/10.1038/s41587-019-0036-z

Altstein, M., Nässel, D.R., 2010. Neuropeptide Signaling in Insects. Adv. Exp. Med. Biol. 692, 155–165. https://doi.org/10.1007/978-1-4419-6902-6_8

Badisco, L., Claeys, I., Van Hiel, M., Clynen, E., Huybrechts, J., Vandersmissen, T., Van Soest, S., Vanden Bosch, L., Simonet, G., Vanden Broeck, J., 2008. Purification and characterization of an insulin-related peptide in the desert locust, *Schistocerca gregaria*: immunolocalization, cDNA cloning, transcript profiling and interaction with neuroparsin. J. Mol. Endocrinol. 40, 137–150.

https://doi.org/10.1677/jme-07-0161

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

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

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

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

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

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

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

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

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

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

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

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

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-
678 3113.2011.00611.x
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., Dombrovsky, A., Abdel-Latief, M.,
704 Robichon, A., Veenstra, J.A., Tagu, D., 2010. Neuropeptide and neurohormone precursors in the
705 pea aphid, *Acyrthosiphon pisum*. Insect Mol. Biol. 19, 87–95. https://doi.org/10.1111/j.1365-
706 2583.2009.00951.x
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
710 International Aphid Genomics Consortium, T., 2010. Genome sequence of the pea aphid

711 *Acyrthosiphon 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 *Acyrthosiphon 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 silkmoth *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-
736 1033.1990.tb15302.x
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 silkmoth *Bombyx mori*. FEBS J. 276, 1221–1232. https://doi.org/10.1111/j.1742-
768 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., Jousselin, 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, *Acyrthosiphon 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.

https://doi.org/10.1016/j.devcel.2009.10.009
Smit, A.B., Van Kesteren, R.E., Li, K.W., Van Minnen, J., Spijker, S., Van Heerikhuizen, H., Geraerts, W.P.M., 1998. Towards understanding the role of insulin in the brain: Lessons from insulin-related signaling systems in the invertebrate brain. Prog. Neurobiol. 54, 35–54. https://doi.org/10.1016/S0301-0082(97)00063-4
Smykal, V., Pivarci, M., Provaznik, J., Bazalova, O., Jedlicka, P., Luksan, O., Horak, A., Vaneckova, H., Benes, V., Fiala, I., Hanus, R., Dolezel, D., 2020. Complex evolution of insect insulin receptors and homologous decoy receptors, and functional significance of their multiplicity. Mol. Biol. Evol. 37, 1775–1789. https://doi.org/10.1093/molbev/msaa048
Steiner, D., 1985. Structure and Evolution of the Insulin gene. Annu. Rev. Genet. 19, 463–484. https://doi.org/10.1146/annurev.genet.19.1.463
Vafopoulou, X., Steel, C.G.H., 2014. Synergistic induction of the clock protein PERIOD by insulin-like peptide and prothoracicotropic hormone in *Rhodnius prolixus* (Hemiptera): Implications for convergence of hormone signaling pathways. Front. Physiol. 5, 41. https://doi.org/10.3389/fphys.2014.00041
Veenstra, J.A., Agricola, H.J., Sellami, A., 2008. Regulatory peptides in fruit fly midgut. Cell Tissue Res. 334, 499–516. https://doi.org/10.1007/s00441-008-0708-3
Veenstra, J.A., 2020. Arthropod IGF, relaxin and gonadulin, putative orthologs of *Drosophila* insulin-like peptides 6, 7 and 8, likely originated from an ancient gene triplication. PeerJ, 8:e9534. http:// doi: 10.7717/peerj.9534.
Wheeler, D.E., Buck, N., Evans, J.D., 2006. Expression of insulin pathway genes during the period of caste determination in the honey bee, *Apis mellifera*. Insect Mol. Biol. 15, 597–602. https://doi.org/10.1111/j.1365-2583.2006.00681.x
Williams, K.D., Busto, M., Suster, M.L., So, A.K.C., Ben-Shahar, Y., Leevers, S.J., Sokolowski, M.B., 2006. Natural variation in *Drosophila melanogaster* diapause due to the insulin-regulated PI3-kinase. Proc. Natl. Acad. Sci. U. S. A. 103, 15911–15915. https://doi.org/10.1073/pnas.0604592103
Wu, Q., Brown, M.R., 2006. Signaling and function of Insulin-Like Peptides in insects. Annu. Rev. Entomol. 51, 1–24. https://doi.org/10.1146/annurev.ento.51.110104.151011
Wu, Q., Zhang, Y., Xu, J., Shen, P., 2005. Regulation of hunger-driven behaviors by neural ribosomal S6 kinase in *Drosophila*. Proc. Natl. Acad. Sci. U. S. A.102, 13289–13294. https://doi.org/10.1073/pnas.0501914102
Yang, C.H., Belawat, P., Hafen, E., Jan, L.Y., Jan, Y.N., 2008. *Drosophila* egg-laying site selection as a system to study simple decision-making processes. Science 319, 1679–1683. https://doi.org/10.1126/science.1151842
Zheng, S., Chiu, H., Boudreau, J., Papanicolaou, T., Bendena, W., Chin-Sang, I., 2018. A functional study of all 40 *Caenorhabditis elegans* insulin-like peptides. J. Biol. Chem. 293, 16912–16922.

https://doi.org/10.1074/jbc.RA118.004542

Figures and Tables

Figure 1. Schematic representation of the predicted structure of *Acyrthosiphon pisum* insulin related peptides (IRPs) and comparison with known insulin- and IGF-like peptides. Schematic representation and alignments of *A. pisum* IRP1-IRP4 (A), IRP5-6 (B) and IRP11 (C). Domains are denoted by letters. The spaces between domains represent predicted proteolytic cleavage during maturation of the propeptide. Disulfide bonds between conserved cysteines are indicated. Amino acid sequences of the predicted IRPs from *A. pisum* are aligned. Amino acid residues are highlighted in red when fully conserved and written in red when partially conserved. Color bars below the alignment indicate the predicted domains in the precursor peptides: green, signal peptide; orange, B chain; yellow, C chain; blue, A chain. Asterisks denote cysteine residues, and paired triangles denote canonical prohormone convertase or furin cleavage sites (dibasic amino acids).

Figure 2. Genomic organization of *Acyrthosiphon pisum* insulin related peptides (IRPs) mapping on the A1, A3 and X chromosomes. Genomic DNA (GenBank accession number GCF_005508785.1) is represented by a line (top) with distances in kb. White and grey boxes indicate predicted introns and predicted exons respectively. In the case of IRP5 and IRP6, the two predicted transcripts are represented. Positions of the primers used for the (qRT-)PCR experiments are indicated as red arrows above each transcript.

Figure 3. Maximum likelihood tree showing relationships among IRP amino acid sequences from different aphid species. Unrooted tree of IRP sequences from 10 aphid species and the grape phylloxera, *Daktulosphaira vitifoliae*, an evolutionary ancient aphid-related species. Sequence names are indicated as a prefix formed by the abbreviated species name (*Acyrthosiphon pisum,* Ap; *Aphis glycines*, Agl; *Aphis gossypii*, Ago; *Daktulosphaira vitifoliae*, Dv; *Diuraphis noxia*, Dn; *Melanaphis sacchari*, Ms; *Myzus cerasi*, Mc; *Myzus persicae*, Mp; *Rhopalosiphum maidis*, Rm; *Rhopalosiphum padi*, Rp; *Sipha flava*, Sf), followed by the IRP common name in the case of the pea aphid (sequences in red) or protein accession number for other species. The putative IRP homolog groups identified in this study, referred to in the text as IRP subfamilies, are highlighted in different colors with a letter corresponding to each subfamily. All the displayed nodes have a bootstrap value above 80%.

Figure 4. Localization of IRP1 and IRP4 in adult pea aphid brain by immunohistochemistry. Images correspond to confocal Z-stack with tubulin in green and target IRP in magenta. For each antibody 15 aphid brains were analyzed. (A) Schematic representation of the central nervous system of the pea aphid adapted from Kollmann et al. (2011). (B) and (C) expression of IRP1 and IRP4 respectively in eight neurons of the pars intercerebralis with OL, optic lobe;

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

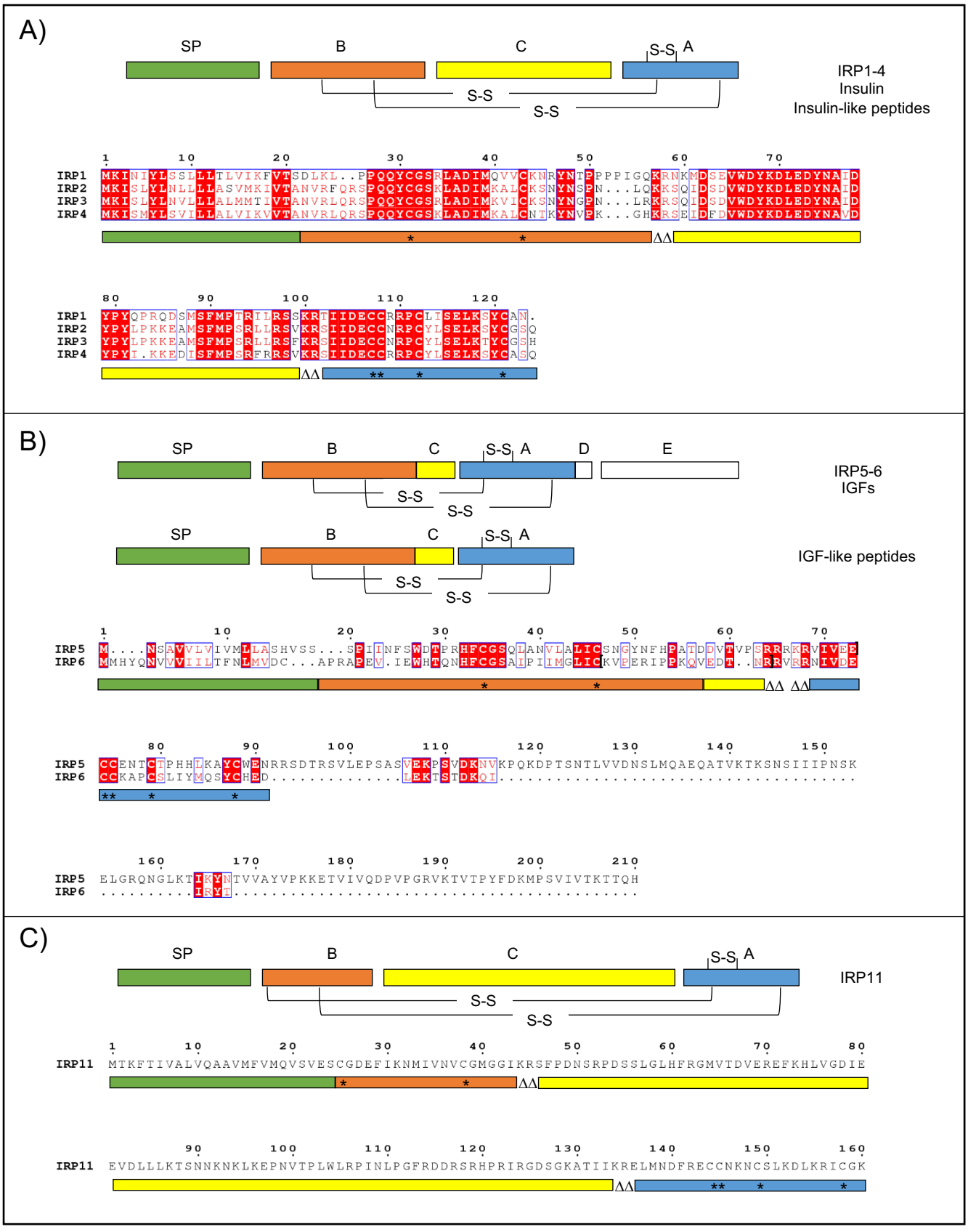
Figure 5. Localization of IRP2 in adult pea aphid digestive tract by immunohistochemistry. Images correspond to confocal Z-stack with tubulin in green and IRP2 in magenta. In total 6 digestive tracts were analyzed. (A) Tubulin staining of the intestinal tract to indicate the boundaries and structure of the intestinal tract. (B) IRP2 is expressed at low levels in circular structures of the intestinal tract.

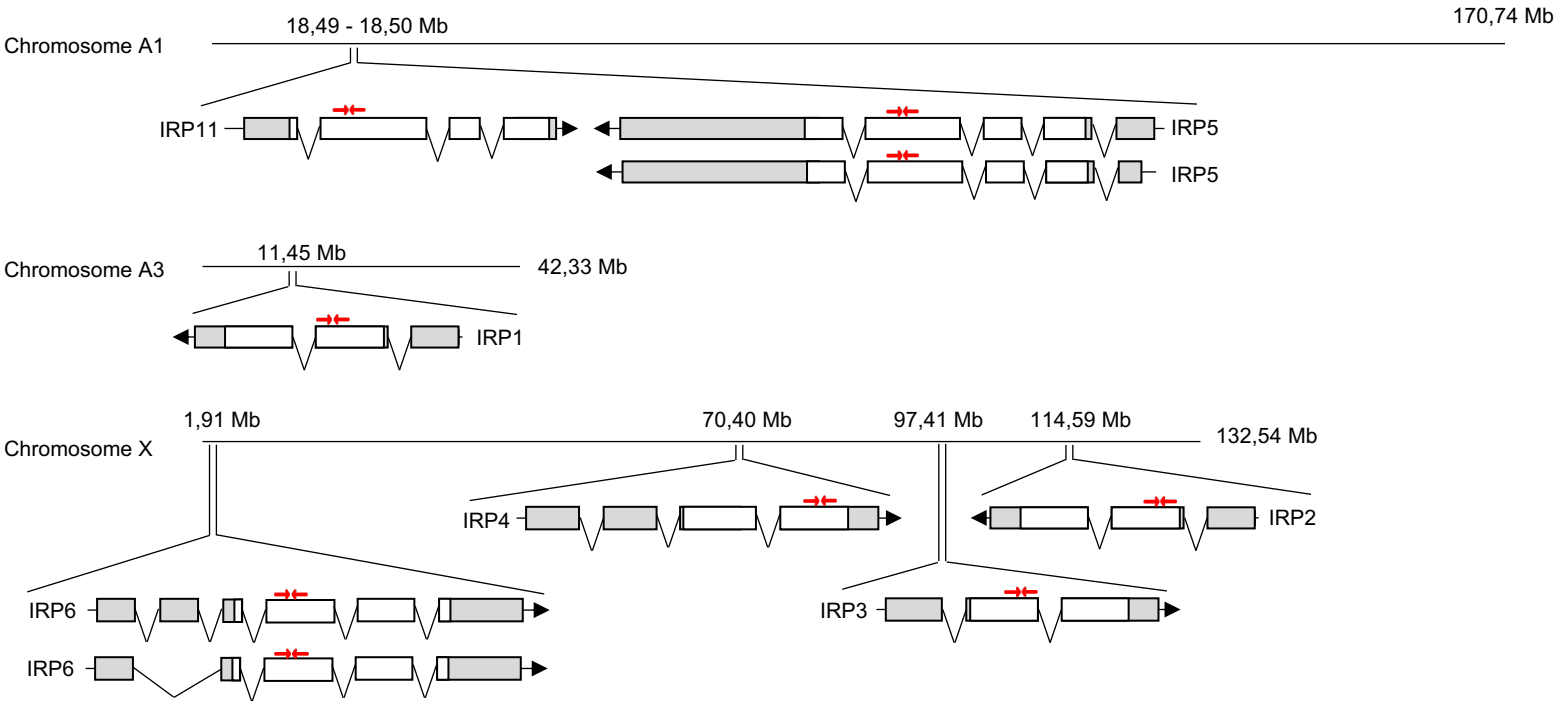
Table 1. Putative IRP-encoding genes in the *A. pisum* genome

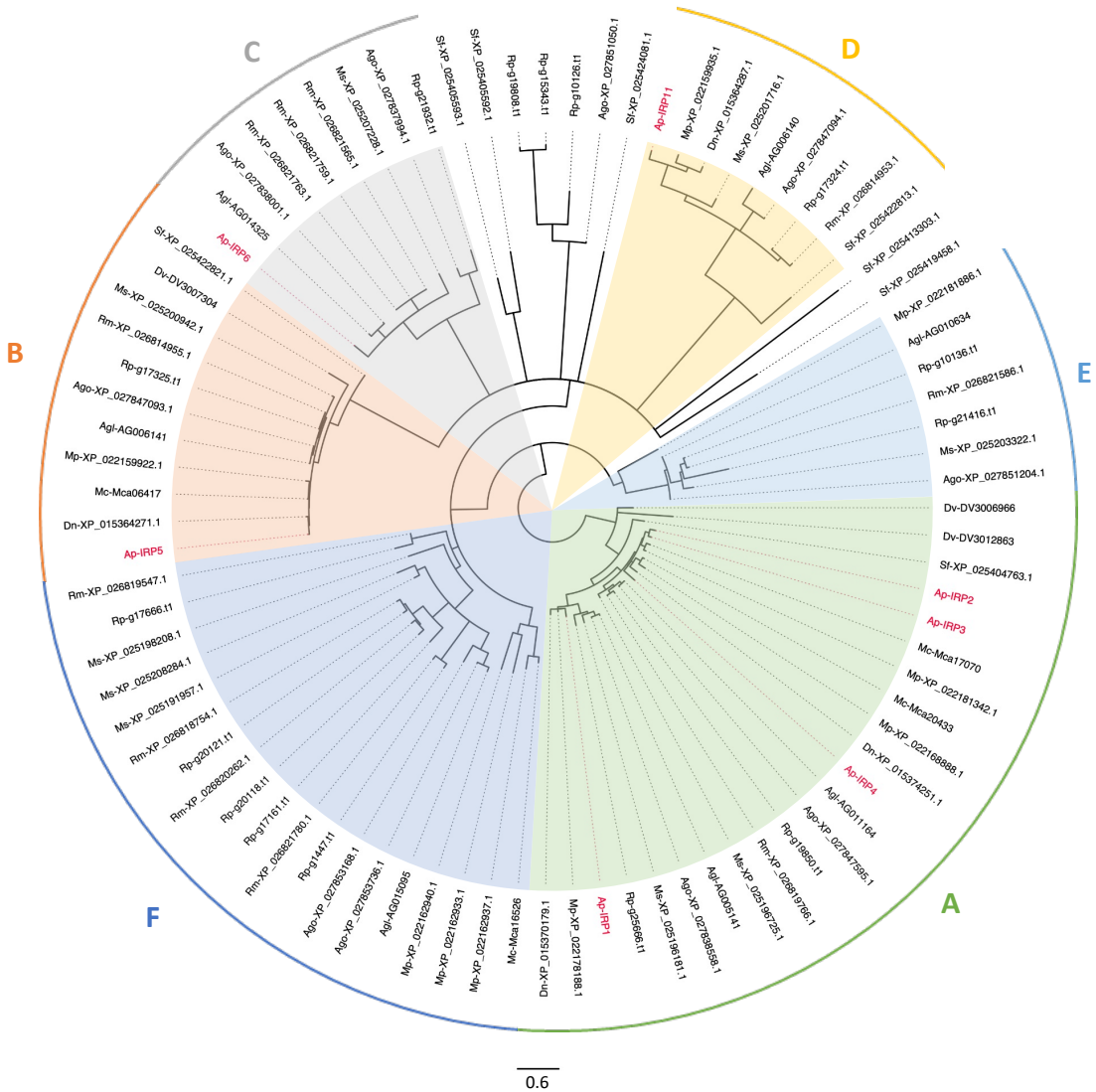
Table 2. Similarity between *A. pisum* IRP protein sequences.

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

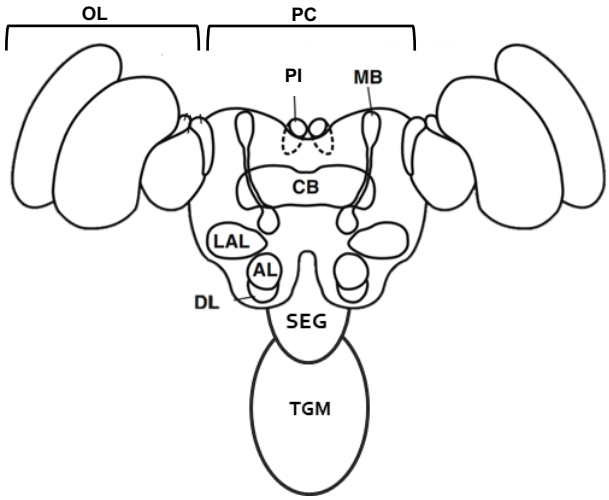
Table 4. qRT-PCR analyses of *A. pisum* IRPs expression within various tissues relative to whole body.


A)
SP
B
C
S-S
A
S-S
S-S
IRP1-4
Insulin
Insulin-like peptides
IRP1
IRP2
IRP3
IRP4
B)
SP
B
C
S-S
A
D
E
S-S
S-S
IRP5-6
IGFs
SP
B
C
S-S
A
S-S
S-S
IGF-like peptides
IRP5
IRP6
C)
SP
B
C
S-S
A
S-S
S-S
IRP11
IRP11


Chromosome A1
18,49 - 18,50 Mb
170,74 Mb
IRP11
IRP5
IRP5
Chromosome A3
11,45 Mb
42,33 Mb
IRP1
Chromosome X
1,91 Mb
70,40 Mb
97,41 Mb
114,59 Mb
132,54 Mb
IRP4
IRP2
IRP6
IRP6
IRP3

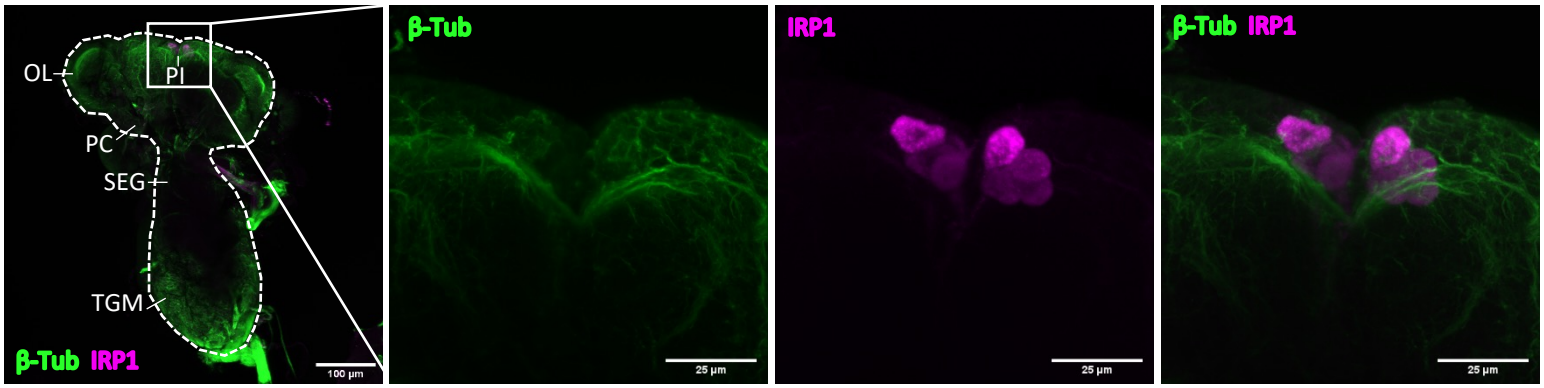

A
B
C
D
E
F
Ap-IRP1
Ap-IRP2
Ap-IRP3
Ap-IRP4
Ap-IRP5
Ap-IRP6
Ap-IRP11
Mp-XP_022178188.1
Rp-g25686.t1
Ms-XP_025196181.1
Ago-XP_027838558.1
Agl-AG035141
Ms-XP_025196725.1
Rm-XP_026819766.1
Rp-g19850.t1
Ago-XP_027847595.1
Agl-AG011164
Dn-XP_015374251.1
Mp-XP_022168888.1
Mc-Mca20433
Mp-XP_022181342.1
Mc-Mca17070
Sf-XP_025404763.1
Dv-DV3012863
Dv-DV3006966
Ago-XP_027851204.1
Ms-XP_025203322.1
Rp-g21416.t1
Rm-XP_026821586.1
Rp-g10136.t1
Agl-AG010634
Mp-XP_022181886.1
Sf-XP_025419458.1
Sf-XP_025413303.1
Sf-XP_025422813.1
Rm-XP_026814953.1
Rp-g17324.t1
Ago-XP_027847094.1
Agl-AG006140
Ms-XP_025201716.1
Dn-XP_015364287.1
Mp-XP_022159935.1
Dn-XP_015370179.1
Mc-Mca16526
Mp-XP_022162937.1
Mp-XP_022162933.1
Mp-XP_022162940.1
Agl-AG015095
Ago-XP_027853736.1
Ago-XP_027853168.1
Rp-g1447.t1
Rm-XP_026821780.1
Rp-g17161.t1
Rp-g20118.t1
Rm-XP_026820262.1
Rp-g20121.t1
Rm-XP_026818754.1
Ms-XP_025191957.1
Ms-XP_025208284.1
Ms-XP_025198208.1
Rp-g17666.t1
Rm-XP_026819547.1
Dn-XP_015364271.1
Mc-Mca06417
Mp-XP_022159922.1
Agl-AG006141
Ago-XP_027847093.1
Rp-g17325.t1
Rm-XP_026814955.1
Ms-XP_025200942.1
Dv-DV3007304
Sf-XP_025422821.1
Ap-IRP6
Agl-AG014325
Ago-XP_027838001.1
Rm-XP_026821763.1
Rm-XP_026821759.1
Rm-XP_026821565.1
Ms-XP_025207228.1
Ago-XP_027837894.1
Rp-g21992.t1
Sf-XP_025405593.1
Sf-XP_025405592.1
Rp-g19808.t1
Rp-g15343.t1
Rp-g10120.t1
Ago-XP_027851050.1
Sf-XP_025424081.1
Ap-IRP11
Mp-XP_022159935.1
Dn-XP_015364287.1
Ms-XP_025201716.1
Agl-AG006140
Ago-XP_027847094.1
Rp-g17324.t1
Rm-XP_026814953.1
Sf-XP_025422813.1
Sf-XP_025413303.1
Sf-XP_025419458.1
Mp-XP_022181886.1
Agl-AG010634
Rp-g10136.t1
Rm-XP_026821586.1
Rp-g21416.t1
Ms-XP_025203322.1
Ago-XP_027851204.1
Dv-DV3006966
Dv-DV3012863
Sf-XP_025404763.1
Ap-IRP2
Ap-IRP3
Mc-Mca17070
Mp-XP_022181342.1
Mc-Mca20433
Mp-XP_022168888.1
Dn-XP_015374251.1
Ap-IRP4
Agl-AG011164
Ago-XP_027847595.1
Rp-g19850.t1
Rm-XP_026819766.1
Ms-XP_025196725.1
Agl-AG035141
Ago-XP_027838558.1
Ms-XP_025196181.1
Rp-g25686.t1
Ap-IRP1
Mp-XP_022178188.1
Dn-XP_015370179.1
Mc-Mca16526
Mp-XP_022162937.1
Mp-XP_022162933.1
Mp-XP_022162940.1
Agl-AG015095
Ago-XP_027853736.1
Ago-XP_027853168.1
Rp-g1447.t1
Rm-XP_026821780.1
Rp-g17161.t1
Rp-g20118.t1
Rm-XP_026820262.1
Rp-g20121.t1
Rm-XP_026818754.1
Ms-XP_025191957.1
Ms-XP_025208284.1
Ms-XP_025198208.1
Rp-g17666.t1
Rm-XP_026819547.1
Dn-XP_015364271.1
Ap-IRP5
Mc-Mca06417
Mp-XP_022159922.1
Agl-AG006141
Ago-XP_027847093.1
Rp-g17325.t1
Rm-XP_026814955.1
Ms-XP_025200942.1
Dv-DV3007304
Sf-XP_025422821.1
Ap-IRP6
Agl-AG014325
Ago-XP_027838001.1
Rm-XP_026821763.1
Rm-XP_026821759.1
Rm-XP_026821565.1
Ms-XP_025207228.1
Ago-XP_027837894.1
Rp-g21992.t1
Sf-XP_025405593.1
Sf-XP_025405592.1
Rp-g19808.t1
Rp-g15343.t1
Rp-g10120.t1
Ago-XP_027851050.1
Sf-XP_025424081.1
Ap-IRP11
0.6

A

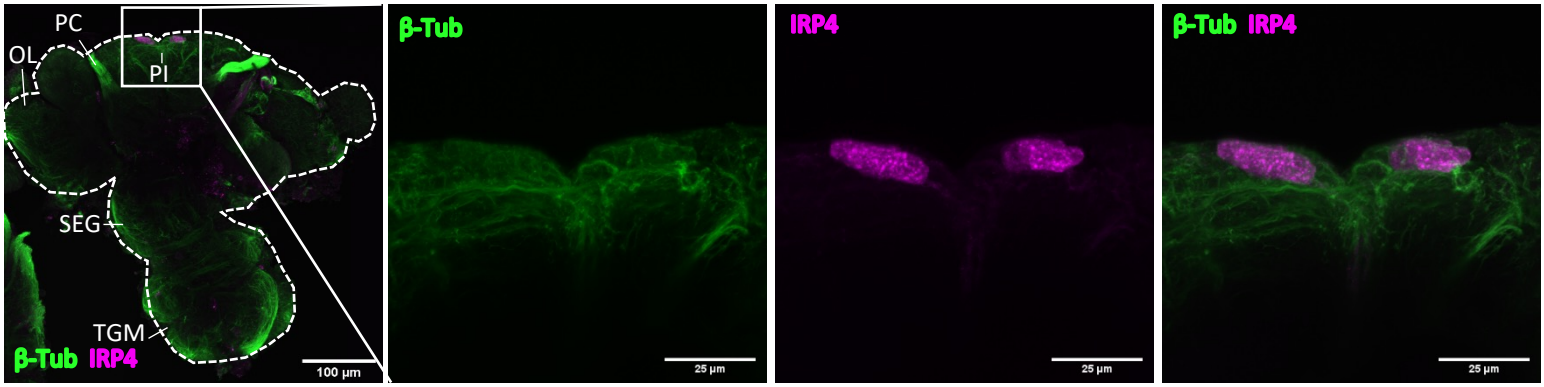


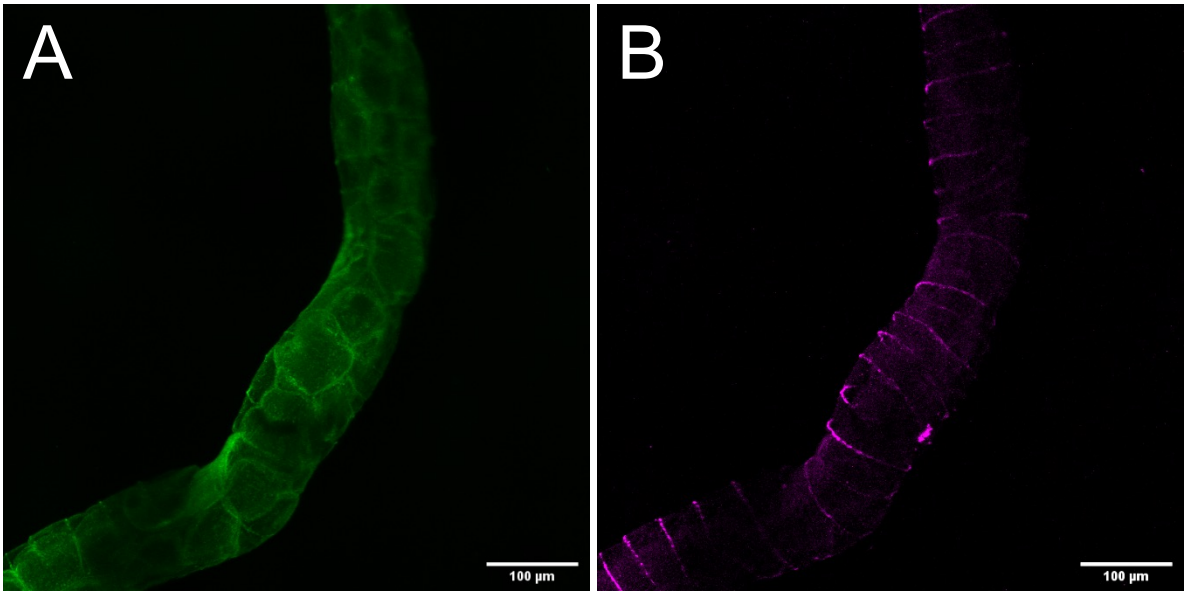
OL - Optic Lobe
PC - Protocerebrum
MB - Mushroom Body
PI - Pars Intercerebralis
CB - Central Body
LAL - Lateral Accesory Lobe
AL - Antennal Lobe
DL - Dorsal Lobe
SEG - Subesophageal Ganglion
TGM - Thoracic Ganglionic Mass

B



C




A
100 µm
B
100 µm

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**	**Id**	**Cov**	**Id**	**Cov**	**Id**	**Cov**	**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**	**Relative expression (SD)**	**Variation**	**Relative expression (SD)**	**Variation**	**Relative expression (SD)**	**Variation**	**Relative expression (SD)**	**Variation**	**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

C group

D group

B group

E group

F group

A group

Fat body

Carcass

Fat body

Intestinal cells

Neurosecretory cells

Male

TISSUE EXPRESSION

EVOLUTION

A. pisum IRP6

A. pisum IRP11

A. pisum IRP5

A. pisum IRP2

A. pisum IRP3

A. pisum IRP4

A. pisum IRP1