Digestive proteases in bodies and faeces of the two-spotted spider mite, *Tetranychus urticae*

María E. Santamaría [a,b,c]; Joel González-Cabrera [a1]; Manuel Martínez [b]; Vojislava Grbic [c]; Pedro Castañera [a]; Isabel Díaz [b]; Félix Ortego [a*]

[a] Departamento de Biología Medioambiental, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu, 9, 28040, Madrid, Spain.
[b] Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid, Campus Montegancedo, Autovía M40 (Km 38), 28223-Pozuelo de Alarcón, Madrid, Spain.
[c] Department of Biology WSC 339/341, The University of Western Ontario, 1151 Richmond St, London, ON N6A 5B7, Canada.

* Corresponding Author: Félix Ortego, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu, 9, 28040, Madrid, Spain. Tel.: +34- 918373112; fax: +34- 91536043; E-mail: ortego@cib.csic.es

[1] Present Address: Biological Chemistry and Crop Protection Department, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom.

Abstract

Digestive proteases of the phytophagous mite *Tetranychus urticae* have been characterized by comparing their activity in body and faecal extracts. Aspartyl, cathepsin B- and L-like and legumain activities were detected in both mite bodies and faeces, with a specific activity of aspartyl and cathepsin L-like proteases about 5- and 2-fold higher, respectively, in mite faeces than in bodies. In general, all these activities were maintained independently of the host plant where the mites were reared (bean, tomato or maize). Remarkably, this is the first report in a phytophagous mite of legumain-like activity, which was characterized for its ability to hydrolyse the specific substrate Z-VAN-AMC, its activation by DTT and inhibition by IAA but not by E-64. Gel free nanoLC-nanoESI-QTOF MS/MS proteomic analysis of mite faeces resulted in the identification of four cathepsins L and one aspartyl protease (from a total of the 29 cathepsins L, 27 cathepsins B, 19 legumains and two aspartyl protease genes identified the genome of this species). Gene expression analysis reveals that four cathepsins L and the aspartyl protease identified in the mite faeces, but also two cathepsins B and two legumains that were not detected in the faeces, were expressed at high levels in the spider mite feeding stages (larvae, nymphs and adults) relative to embryos. Taken together, these results indicate a digestive role for cysteine and aspartyl proteases in *T. urticae*. The expression of the cathepsins B and L, legumains and aspartyl protease genes analyzed in our study increased in female adults after feeding on Arabidopsis plants over-expressing the HvCPI-6 cystatin, that specifically targets cathepsins B and L, or the CMe trypsin inhibitor that targets serine proteases. This unspecific response suggests that in addition to compensation for inhibitor-targeted enzymes, the increase in the expression of digestive proteases in *T. urticae* may act as a first barrier against ingested plant defensive proteins.

Keywords: Phytophagous mites, cathepsin, legumain, cystatin, protease inhibitors, proteomics.

1. Introduction

Proteolytic digestion in Acari (mites and ticks) has been characterized in only a few taxonomic groups. In stored products and house dust mites, the hydrolysis of dietary proteins mostly relies on cysteine and serine proteases (Stewart et al., 1991, 1992; Ortego et al., 2000; Erban and Hubert, 2010a). These enzymes are synthesized in cells lining the gut epithelium and secreted into the gut lumen for extracellular digestion (Thomas et al., 1991; Erban and Hubert, 2011; Herman et al., 2014). It has been reported that, at least in house dust mites, serine and cysteine proteases are synthesized as inactive zymogens that require enzymatic or autocatalytic activation, respectively, by removal of N-terminal pro-peptide regions (Herman et al., 2014; Dumez et al., 2014). The gut contents showed a gradient in pH for most species, ranging from 4.5-5.0 at the anterior midgut to 5.5-7.0 in the posterior midgut (Erban and Hubert, 2010b), which may be related with the activation and/or the compartmentalization of the activity for some of these proteases (Erban and Hubert, 2011; Dumez et al., 2014). Ingested food and digestive enzymes are packed into a food bolus surrounded by a peritrophic membrane formed in the midgut and compacted as faecal pellets in the posterior midgut and the hindgut (Tovey et al., 1981; Sobotnik et al., 2008), being considered the presence of specific proteases in the faeces as an indication of their involvement in digestion (Stewart et al., 1992; Erban and Hubert, 2015). Observations of the digestive tract of house dust mites showed that some cells lining the gut epithelium can be detached in the gut lumen to become free-floating digestive cells that are also incorporated to the faecal pellets (Colloff, 2009). Another groups whose digestive physiology has been analyzed in detail are blood-feeding ticks and scab mites. Both utilize a digestive system reliant upon acidic cysteine and aspartyl proteases functioning in an intracellular environment in the digestive vacuoles of the midgut cells with further processing by cytosolic and/or membrane bound dipeptidases and exopeptidases (Hamilton et al., 2003; Sojka et al., 2013).

The proteolytic profile of the plant feeding tetranichid mite *Tetranychus urticae* Koch has been analyzed using whole mite extracts (Carrillo et al., 2011a). Cysteine (cathepsins B- and L-like) and aspartyl protease activities were identified, being consistent with the proliferation of C1A cysteine protease genes (includes 27 cathepsins B and 29 cathepsins L) and the presence of two aspartyl protease genes in the genome of *T. urticae* (Grbic et al., 2011). A proliferation of C13 cysteine protease genes (19 legumains) was also identified in the genome of *T. urticae* (Grbic et al., 2011), but this activity has not been

tested. Both C1A and C13 cysteine proteases are expressed at a higher rate in *T. urticae* feeding stages than in embryos (Santamaría et al., 2012a,b) and the mites modulate their expression when feeding on different host plants (Grbic et al., 2011), which fits with their putative role in the proteolytic digestion of dietary proteins. However, though faecal pellets surrounded by a peritrophic membrane have been described in *T. urticae* (Evans, 1992), they have not been enzymatically characterized. A large S1 serine protease gene family was also identified in the *T. urticae* genome (Grbic et al., 2011). However, trypsin- and chymotrypsin-like activities were not detected in *T. urticae* (Carrillo et al., 2011a) and serine protease genes did not show a clear developmental pattern of expression, indicating that they are probably not directly involved in the hydrolytic digestion of dietary proteins (Santamaría et al., 2012b).

It has been proposed that the proliferation of cysteine proteases in *T. urticae* could be a strategy to counteract the effect of plant defense proteins synthesized by different host plants (Santamaría et al., 2012a). Polyphagous herbivorous are expected to be exposed to a broad array of defense proteins of plant origin, representing their digestive enzymes the first-line of defenses to deal with these noxious plant proteins. In the case of phytophagous insects, it is established that the success to cope with plant defense proteins depends on the regulation of complex multigene families of digestive enzymes (Bown et al., 1997; Chougule et al., 2005; Chi et al., 2009). Indeed, insect feeding on plants are adapted to circumvent the adverse effects of plant protease inhibitors by the up-regulation of targeted proteases, the induced expression of novel insensitive proteases, and/or the proteolytic degradation of protease inhibitors (Jongsma and Bolter, 1997; Ortego, 2012). However, little is known about the ability of phytophagous mites to regulate gene expression of digestive enzymes in response to the ingestion of protease inhibitors. Transcriptome analysis of *T. urticae* showed that about one third of cysteine proteases were differentially expressed when feeding on different host plants (Grbic et al., 2011). Interestingly, Santamaría et al. (2012b) reported that proteolytic activities in *T. urticae* were altered when mites fed on transgenic Arabidopsis plants that over-expressed different types of protease inhibitors, though this study did not examine gene expression profiles.

In this article we identified proteases in *T. urticae* faeces by proteomic analysis, characterized their activities in mite bodies and faeces, and analyzed the expression of selected protease genes at different mite developmental stages and in response to feeding on Arabidopsis plants over-expressing protease inhibitors. Our results support the role of

cathepsin B, cathepsin L, legumain and aspartyl proteases in the digestive process of *T. urticae*.

2. Materials and Methods

2.1. Spider mites

A colony of *T. urticae,* London strain (Acari: Tetranychidae), provided by Dr. Miodrag Grbic (UWO, Canada), was reared for several generations on beans (*Phaseolus vulgaris*) var California red kidney (Stokes Seed Co., St. Catharines, Canada), tomato (*Solanum lycopersicum*) var Moneymaker (Lidl Supermercados S.A.U., Barcelona, Spain) or maize (*Zea mays*) var DKC4795 (Monsanto Agricultura España S.L., Madrid, Spain) maintained in growth chambers (Sanyo MLR-350-H, Sanyo, Japan) at 25 ± 1ºC, 70 ± 15% relative humidity and a 16h/8h day/night photoperiod.

2.2. Plant Material

Arabidopsis plants expressing the barley *Icy6* gene encoding the HvCPI-6 cystatin without its signal peptide (35S-*Icy6*-plant, in this article CPI6-plants) or the barley *Itr1* gene encoding the BTI-CMe trypsin inhibitor (35S-*Itr1*-plants, in this article CMe-plants), previously described by Carrillo et al. (2011b) and Santamaría et al. (2012b), were used in this study. Transgenic and non-transformed plants were grown in a Sanyo MLR-350-H growth chamber at 23±1ºC, >70% relative humidity and a 16h/8h day/night photoperiod.

2.3. Enzymatic assays

Adult mite bodies and faeces were collected from infested bean, tomato and maize plants using a vacuum pump with vacuum regulator (Dinko D-95, Dinko Instruments, Barcelona, Spain) under a Leica MS5 stereomicroscope (Leica Microsystems S.A., Barcelona, Spain) and were stored frozen at -80 ºC. Three biological replicates for bodies and two biological replicates for faeces were obtained from each host plant. Total proteins were extracted by homogenising mite bodies (2-3mg) or faeces (0.5-1mg) in 150 µl of ice-cold 150 mM NaCl using a pellet pestle motor hand held homogenizer (Sigma-Aldrich Corp., St. Louis, MO, USA.). The mixtures were centrifuged at 16000 × g and the supernatants containing the protein extracts were used to determine enzymatic activities. Total protein concentration was determined by the method of Bradford (Bradford, 1976) using Bovine Serum Albumin (Sigma-Aldrich Corp.) as standard.

Cathepsin B-, cathepsin L- and legumain-like activities were assayed using Z-RR-AMC (N-carbobenzoxyloxy-Arg-Arg-7-amido-4-methylcoumarin) (Calbiochem, San Diego, CA, USA), Z-FR-AMC (N-carbobenzoxyloxy-Phe-Arg-7-amido-4-methylcoumarin) (Calbiochem) and Z-VAN-AMC (N-carbobenzoxyloxy-Val-Ala-Asn-7-amido-4-methylcoumarin) (Bachem, Bubendorf, Switzerland) as substrates, respectively. The assays were performed in 100 µl reaction buffer (100 mM sodium citrate, 150 mM NaCl, 5mM magnesium chloride), containing 1mM dithiothreitol (DTT) (Sigma-Aldrich Corp.) and 20 µM of substrate (added in 5 µl of DMSO). Total protein (homogenate extracts) in the reaction was 1 (cathepsin L), 2 (cathepsin B) or 5 µg (legumain) for mite bodies, and 0.2 (cathepsin L), 0.3 (cathepsin B) or 0.5 µg (legumain) for mite faeces. The pH of the reaction buffer was set to 5.5 to measure cathepsins B and L-like activities (Carrillo et al., 2011a). The optimum pH for legumain-like activity was determined to be 4.5 after testing this activity using sodium citrate buffer pH 3.0 to 6.5 with 0.5 pH unit increments (Supplementary Figure S1). The reactions were incubated at 30ºC and emitted fluorescence was measured with a 340 nm excitation wavelength filter and a 460 nm emission wavelength filter in a Varioskan™ Flash Multimode Reader (Thermo Scientific, Vantaa, Finland) or a POLARstar Omega fluorometer (BMG Labtech GmbH, Germany). The calibration was performed with known amounts of AMC (7-Amino-4-methylcoumarin) (Bachem) in a standard reaction. Aspartyl protease-like activity was determined using the fluorescence-quenching substrate MocAc-GKPILFFRLK(Dnp)-D-R-NH_2 (MocAc-Gly-Lys-Pro-Ile-Leu-Phe-Phe-Arg-Leu-Lys(Dnp)-D-Arg-NH_2) (PeptaNova GmbH, Sandhausen, Germany). The assays were performed in 100 µl reaction buffer (100 mM sodium citrate pH 3.5, 150 mM NaCl, 5mM magnesium chloride), containing: 10 µM E-64 (L-trans-epoxysuccinyl-leucylamido-(4-guanidino)-butane) (Sigma-Aldrich Corp.) that inhibits cysteine proteases; 20 µM of substrate (added in 5 µl of DMSO); and 2 or 0.2 µg of total protein from mite bodies or faeces, respectively. The reactions were incubated at 30ºC and emitted fluorescence was measured with a 328 nm excitation wavelength filter and a 393 nm emission wavelength filter in a Varioskan™ Flash Multimode Reader (Thermo Scientific). The calibration was performed with known amounts of MCA (MocAc-Pro-Leu-Gly) (Peptanova), the reference compound for MocAc-substrates. All experiments were performed in 96-well microtiter plates, and blanks were used to account for spontaneous breakdown of substrates. The specific activities were expressed as nmol of substrate hydrolysed per minute per mg protein.

To test the inhibitory activity of IAA (iodoacetamide) (Sigma-Aldrich Corp.) and E-64 on cathepsin B-, L- and legumain-like activities, mite body extracts were incubated for 10 min at room temperature with the reaction buffer containing 2 mM IAA or 2 μM of E64 prior to the addition of the corresponding substrate. The effect on legumain-like activity of removing the cysteine protease activator DTT from the reaction buffer and/or adding the cell lysis detergent Brij35 (0.01 %) to the homogenate was also tested for mite body extracts.

2.4. nanoLC-nanoESI-QTOF MS/MS proteomic analysis

Mite faeces were collected from infested maize plants and stored frozen at -80 ºC. Total proteins were extracted by homogenising faeces in iced-cold 150 mM NaCl. The mixture was centrifuged at 16000 × g and the supernatant containing the protein extract was stored frozen (-80 ºC) and sent in dry ice to the Proteomics Facility at the Centro de Investigación Príncipe Felipe (Valencia, Spain). A sample volume of 50 μl (containing 5.5 μg of protein) was sonicated on ice for 5 min, shaked for 30 min, centrifuged at 12000 × g for 30 min, and the supernatant filtered using a Teflon syringe filter (0.45 μm pore size). Proteins were precipitated overnight with trichloroacetic acid (final concentration of 10%) at 4ºC. The pellet was washed with acetone at -20ºC and was dissolved in 20 μl of 50 mM ammonium bicarbonate (ABC) containing 50% trifluoroethanol (TFE). The sample was sequentially treated with 2 mM DTT at 60ºC for 30 min, 5 mM IAA at room temperature in the dark for 30 min, and 10 mM DTT at room temperature for 30 min. After dilution of TFE (< 5%) with 50 mM ABC, the sample was treated with 500 ng trypsin (Promega, Madison, WI) at 37ºC overnight. The digestion was stopped with 20 μl of 10% trifluoroacetic acid (TFA) (pH 1), and the sample dried under a vacuum centrifuge. The pellet was resuspended in 9 μl of 0.1% TFA.

Five microliters of the sample were delivered to a trap column (LC Packings C18, 300 μm x 5 mm, Amsterdam, The Netherlands) via an isocratic flow of 0.1% TFA in water at a rate of 30 μl/min for 3 min. After trapping, peptides were eluted onto an analytical nano liquid chromatography (nanoLC) column (PepMap C18, 3 μm, 100 Å, 75 μm x 50 cm) and separated with a gradient: 5–40% of buffer B (0.1% formic acid in 95% acetonitrile) in buffer A (0.1% formic acid in water) over 120 min at 300 nl/min. The eluted peptides were analyzed by nano electrospray ionisation quadrupole time-of-flight tandem mass spectrometry (nanoESI-QTOF MS/MS) (QSTAR XL, Applied Biosystems) in an information-dependent acquisition mode (IDA), in which a 1-s TOF MS scan from 350 to

2000 m/z was performed, followed by 3-s product ion scans from 65 to 2000 m/z on the three most intense doubly- or triply-charged ions.

Database searches were performed at the OrcAE *T. urticae* database (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur; 32527 sequences; 9591014 residues) using the MASCOT DAEMON (Matrix-Science) search engine. Basically, MASCOT searches were done with tryptic specificity, allowing one missed cleavage and a tolerance on the mass measurement of 0.8 Da. The carbamidomethylation of cysteine was set as a fixed modification, and the oxidation of methionine and the deamidation of asparagine and glutamine were selected as variable modifications. The significance threshold was set at p value > 0.05. Only proteins with a mascot identity score threshold $\geq$ 37 were selected and shown in the results. Subsequently, the peptide sequences reported in the proteomic analysis of the proteases identified in the faeces were searched using BlastP against the non-redundant NCBI protein database of maize to discard any bias coming from the host/food proteome.

2.5. Gene expression in T. urticae

To study gene expression of different *T. urticae* developmental stages, 100 mg of eggs, larvae, nymphs and adults were collected from infested bean plants using sieves of different sizes and a vacuum pump, frozen in liquid N_2 and stored at -80ºC until used for RNA isolation. Gene expression in adult mites feeding on transgenic and non-transformed Arabidopsis leaf disks was also analyzed. Leaf disks (0.5 cm diameter) were maintained on wet cotton in Petri dishes. Each leaf disk was inoculated with 10 female adults of *T. urticae*. Dishes were maintained in controlled conditions as described above. Mites were feeding on the leaf disk for 48h, and then collected using a vacuum pump, frozen into liquid N_2 and stored at -80ºC until used for RNA isolation. Three replicates were done for each developmental stage and/or plant genotype.

Total RNA was extracted by the TRIZOL® Reagent following manufacturer instructions (Ambion Austin TX), but adding two extra steps to the basic protocol to enhance the RNA yield. Prior to precipitation, 10 µg RNase-free glycogen (Ambion) has been added as carrier to the aqueous phase and the final RNA pellet has been suspended in RNA secure (Ambion).

cDNAs were synthesized from 2 µg of RNA using the Revert Aid™ H Minus First Strand cDNA Synthesis Kit (Fermentas, St. Leon-Roth, Germany) following manufacturer's instructions. The RT-qPCR conditions were 15 sec at 95ºC, 1 min at 55ºC and 5 sec at

65ºC over 40 cycles. FastStart Universal SYBR Green Master (Rox) (Roche Diagnostics GmbH, Germany) was used in a total volume of 20 µl. The reactions were carried out in a C1000^{TM} thermal cycler with CFX96^{TM} optical reaction module (BioRad, Hercules, CA, USA) using multiplate PCR plates, and the results were analyzed using CFX Manager Software 2.0 (BioRad, Hercules, CA, USA). For negative controls, 1 µl of water was used instead of cDNA, as well as an RNA sample without reverse transcription (no-RT). Primer efficiency was tested using a standard curve for each gene. After amplification, a melting curve analysis was performed to verify gene specificity. The absence of genomic DNA was validated by the no-RT control. Reactions were performed per duplicate for each sample. Values were normalized to ribosomal protein Rp49 expression levels. Gene expression values were referred as fold change (2^{-ddCt}) relative to embryos (for different mite developmental stages) or to adults feeding on non-transformed plants (for adult mites feeding on transgenic Arabidopsis). The oligonucleotides used as primers for RT-qPCR amplification are shown in Table 1.

2.6. Statistical analysis

Statistical differences of enzymatic activities and gene expression levels were analyzed by one-way ANOVA followed by pair-wise multiple comparisons (Student-Newman-Keuls method).

3. Results

3.1. Enzymatic assays

Our results indicate that both *T. urticae* body and faecal extracts hydrolyzed specific substrates for aspartyl (MocAc-GKPILFFRLK(Dnp)-D-R-NH_2), cathepsin B (Z-RR-AMC), cathepsin L (Z-FR-AMC) and legumain (Z-VAN-AMC) proteases (Table 2). The specific activity of aspartyl- and cathepsin L-like proteases were about 5- and 2-fold higher, respectively, in mite faeces than in bodies, whereas no significant differences between bodies and faeces were obtained for cathepsin B- and legumain-like proteases. These specific activities and proportions were maintained independently of the host plant where the mites were reared (bean, tomato or maize), with the only exception of the bodies of mites reared on maize that showed significantly higher cathepsin B-like activity that when reared on beans.

Further characterisation of these enzymatic activities was performed by testing the inhibitory effect of IAA and E-64 on mite body extracts (Fig. 1A). Both inhibitors strongly inhibited cathepsin B-like activity, whereas cathepsin L-like activity was differentially inhibited with about 20% reduction when IAA was present in the reaction and more than 80% reduction with E64. Legumain-like activity was also differentially affected, but in this case IAA inhibited about 50% of the activity while the presence of E64 increased the activity by more than 40%. We have also determined that legumain-like activity was significantly reduced when DTT was not added to the reaction buffer and that the addition of the detergent Brij 35 during the homogenization of mite bodies did not have any effect (Fig. 1B).

3.2. Proteomic analysis

Five different proteases were identified in faecal extracts of *T. urticae* by nanoLC-nanoESI-QTOF MS/MS, corresponding to four cathepsins L and one aspartyl protease (Table 3). When the peptide sequences of the proteases identified in faeces were searched against the non-redundant NCBI protein database, complete sequence identity was not observed with any protein from maize (host/food for the mites used to collect the mite faeces), dismissing the possibility of miss-identification of mite proteases in our analysis.

Other proteins present in the faecal extracts were identified as: enzymes involved in glycoside hydrolysis (beta-galactosidase, fructose-bisphosphate aldolase and alpha-N-acetylgalactosaminidase), detoxification (glutathione S-transferases), anti-oxidant systems (thioredoxin peroxidase) and cell metabolism (ATP synthases, ATPase and ADP/ATPtranslocase); storage (vitellogenins), lipid-binding (apolipoproteins, fatty acid-binding proteins, lipoprotein Blc and calycin-like), cuticular and structural (actin, tubulin and clathrin) proteins; proteins involved in other physiological processes (cell wall surface anchor family protein, low complexity secreted protein, elongation factor 1-alpha, TEP1, C-factors, histone H4, adhesion molecules, heat shock protein hsp70, guanylate-binding protein and telomerase-binding protein EST1A); and proteins of unknown function (Supplementary Table S1).

3.3. Expression of protease genes

The following protease genes were selected to study their expression by RT-qPCR: a) the four cathepsins L (TuPap-49, 38, 41 and 42) and the aspartyl protease (TuPep-2)

identified in *T. urticae* faecal extracts by proteomic analysis; and b) two cathepsins B (TuPap-17 and 12), two legumains (TuLeg-10 and 5) and four serine proteases (TuSP7, 28, 29 and 35) selected from the *T. urticae* genome at the OrcAE website (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur). In general, the expression of cathepsin L (TuPap-49, 38, 41 and 42), cathepsin B (TuPap-17 and 12), legumain (TuLeg-10 and 5) and aspartyl (TuPep-2) proteases was significantly higher in the feeding stages (larvae, nymph and adult) than in the embryo for most of the genes analyzed (Fig. 2A). In the case of the four serine protease genes analyzed, the expression was higher in embryos than in adults for TuSP28 and 29, equal for TuSP7, and lower in embryos for TuSP35 (Fig. 2B).

Gene expression was also analyzed in adult female mites after feeding on Arabidopsis plants that over-express the HvCPI-6 cystatin or the BTI-CMe trypsin inhibitor (Figure 3). All cathepsin L (TuPap-49, 38, 41 and 42), cathepsin B (TuPap-17 and 12), legumain (TuLeg-10 and 5) and aspartyl (TuPep-2) protease genes analyzed showed statistically higher expression in mites after feeding on transgenic plants that over-express either cysteine or serine protease inhibitors. However, different responses were obtained for the four genes encoding for serine proteases. The expression of TuSP7 and 28 was higher in mites that were fed on plants over-expressing serine protease inhibitors, the expression of TuSP35 was higher in mites feeding on both types of transgenic plants, and the expression of TuSP29 was not affected by the ingestion of inhibitors.

4. Discussion

In this study we have shown that cysteine (cathepsin B-, cathepsin L- and legumain-like) and aspartyl protease activities are detected in both *T. urticae* bodies and faeces. Some of these proteolytic activities have been previously identified in *T. urticae* body extracts (Nisbet and Billingsley et al., 2000; Carrillo et al., 2011a), but their presence in the spider mite faeces is indicative of their involvement in the digestive process. Cysteine proteases in *T. urticae* constitute a large family that includes 27 cathepsins B, 29 cathepsins L and 19 legumains, whereas only two aspartyl protease genes have been identified in this species (Grbic et al., 2011). The majority of these protease genes contain complete catalytic site motifs, though there are also a number of predicted proteolytically inactive proteases homologous with modified catalytic sites (1 legumain, 1 cathepsin L and 12 cathepsins B genes). The proteomic analysis of the faeces resulted in the identification

of four different cathepsins L and one aspartyl protease, indicating that they represent the most abundant proteases in the gut lumen that are ultimately incorporated into the mite faecal pellets. Besides, all of them were predicted to be proteolytically active, whereas none of the protease homologous was identified in the faeces. Cathepsins B and legumains, as well as those cathepsins L and the aspartyl protease not identified in the proteomic analysis, may be quantitatively less abundant in the faecal pellets or may localize in other tissues. Interestingly, the specific activity of cathepsin L- and aspartyl protease-like proteases was about 2- and 5-fold higher in faeces than in bodies, corroborating their abundance in the faeces. Erban and Hubert (2015) have recently identified multiple isoforms of the allergens Der f 1 (a cysteine protease) and Der f 3 (a trypsin-like protease) when analyzing the faeces of the house dust mite *Dermatophagoides farinae* by two-dimensional gel proteomic analysis, and suggested the presence of both inactive zymogens and active forms of these enzymes in the faeces. However, the gel free proteomic analysis performed in our study does not allow the identification of isoforms, and therefore enzyme activity determined using substrates does not necessarily correspond to the total protease forms present in the faeces. In any case, the functional significance of the different types of cysteine and aspartyl proteases in the digestion of dietary proteins has yet to be determined. Indeed, gene expression analysis reveals that the four cathepsins L and the aspartyl protease identified in the mite faeces, but also two cathepsins B and two legumains that were not detected in the faeces and selected for this study were expressed at high levels in the mite feeding stages (larvae, nymphs and adults) relative to embryos, in agreement with their potential involvement in digestion. Besides, the activity of the different types of digestive proteases may depend on a variety of factors such as the host plant, as suggested by the higher specific activity of cathepsin B-like proteases in the bodies of mites reared on maize compared to those reared on beans. A total of 70 serine proteases have also been identified in the genome of *T. urticae* (Grbic et al., 2011), but the lack of detection of trypsin-, chymotrypsin- and elastase-like activity in mite extracts (Carrillo et al., 2011a), together with the different expression patterns of the four serine protease genes analyzed in this study, suggest that they are probably associated with physiological processes other than the hydrolysis of dietary proteins.

We report here for the first time the characterization of legumain-like activity in *T. urticae*. Our results show that legumain-like activity was enhanced by DTT and inhibited by the cysteine protease inhibitor IAA, whereas E-64 was ineffective, as already reported for

legumains from different organisms (Chen et al., 1997; Caffrey et al., 2000; Alim et al., 2007). This inhibition profile contrasts with those obtained for cathepsin B- and L-like activities in *T. urticae*, which were inhibited by both IAA and E-64 or only by E-64, respectively, indicating that the substrate used to measure legumain-like activity was not hydrolyzed by these other cysteine proteases. In both plants and animals, legumains are enzymes with high specificity for peptide hydrolysis on the carboxyl side of asparagine residues that act as lysosomal/vacuolar processing enzymes of biologically relevant protein substrates (Chen et al., 2000; Müntz and Shutov, 2002; Julian et al., 2013). In addition, legumains have been shown to be involved in the digestion of dietary proteins in haematophagous ticks (Sojka et al., 2007, 2013) and platyhelminth parasites (Delcroix et al., 2006), being in both cases compartmentalized in acidic pH endosomal/lysosomal vesicles in the gut cells. The location of legumains in the phytophagous *T. urticae* is unknown, but its specific activity was similar in faeces and bodies indicating that they might be secreted to the lumen or enclosed in detached digestive cells that are incorporated to the faecal pellets. We have found that legumain-like activity was not increased when *T. urticae* body extracts were homogenized with Brij-35 to favour cell lysis, supporting that they may be secreted, though further evidences are needed to settle this issue. Legumains have been reported to contribute to blood-meal digestion in ticks by: 1) the direct hydrolysis of hemoglobin and serum proteins; and 2) the activation of the zymogen forms of other proteases (clan CA cathepsins) involved in digestion (Alim et al., 2008; Horn et al., 2009; Sojka et al., 2013). Interestingly, the four cathepsins L identified in *T. urticae* faecal extracts are pre-pro-enzymes that share a conserved asparagine near the predicted cleavage site of the pro-region of the enzyme (Supplementary Figure S2), which could be a potential cleavage site for legumains. Additional studies will be necessary to determine if legumains contribute directly to the cleavage of dietary proteins and/or to the processing of other digestive proteases. An intriguing aspect is the proliferation of legumains (19 genes identified) in the genome of *T. urticae* (Grbic et al., 2011), whereas in other arthropods, including ticks and the phytoseiid mite *M. occidentalis*, the number of legumain sequences is restricted to 1 to 3 (Alim et al., 2007, 2008; Anderson et al., 2008; Santamaría et al., 2012a; Hoy et al., 2013). However, Schicht et al (2013) have recently identified 11 predicted legumains in the transcriptome of the hematophagous poultry red mite *Dermanyssus gallinae,* suggesting that extensive expansion of this protein family amongst mites may not be unusual.

Plant-feeding insects have developed a remarkable diversity of physiological adaptations to circumvent the adverse effects of plant defenses, including the regulation of multigene digestive enzymes families in response to the ingestion of hydrolytic enzyme inhibitors (Ortego, 2012). The mite *T. urticae* is one of the most polyphagous species of herbivorous arthropods, being potentially exposed to a broad array of protease inhibitors of plant origin. In a previous study, we demonstrated that proteolytic activities in *T. urticae* were altered when mites fed on Arabidopsis plants that over-expressed cysteine and serine protease inhibitors from barley (Santamaría et al., 2012b). Using the same plant material, we have shown here that gene expression of all cysteine (cathepsins B and L and legumains) and aspartyl protease genes analyzed in this study increased in female adults of *T. urticae* after feeding on CPI6-5.4 and CPI6-9.8 Arabidopsis plants over-expressing the HvCPI-6 cystatin. This barley cystatin targets cathepsins B and L of endogenous (Martínez et al., 2009) and exogenous origin, including those from *T. urticae* (Carrillo et al., 2011a), but have no *in vitro* inhibitory effect on legumains (Martínez et al., 2007). Thus, mites seem to be responding to the ingestion of the HvCPI-6 cystatin by increasing the expression of both inhibitor-targeted (cathepsins B and L) and non-targeted (legumains and aspartyl proteases) enzymes, a strategy already reported for other phytophagous arthropods (Ortego et al., 2001; Zhu-Salzman et al., 2003). The resulting balance of the protease-protease inhibitor interaction can result in a net increase or decrease in the activity of targeted proteases, as previously reported (Santamaría et al., 2012b). Remarkably, cysteine and aspartyl protease genes also increased their expression when *T. urticae* was fed on CMe-3.4 and CMe-8.9 Arabidopsis plants over-expressing the CMe trypsin inhibitor, though this inhibitor was shown to lack an inhibitory effect on *T. urticae* cathepsins B- and L-like activities *in vitro* (Santamaría et al., 2012b). This unspecific response may have an adaptive value for polyphagia, since gut proteases have the potential to inactivate plant defence proteins targeting the digestive system or other tissues, as has been demonstrated in phytophagous insects (Yang et al., 2009). Information about similar mechanisms in mites is scarce, but it has been suggested that stored product mites can also adapt to the ingestion of α-amylase inhibitors by proteolytic degradation of these plant defence proteins (Hubert et al., 2014). It is important to note that although we have only analyzed 6 out of the 56 cathepsins B and L, 2 of the 19 legumains, and 1 of the 2 aspartyl protease genes identified in the genome of *T. urticae* (Grbic et al., 2011), all of them showed a similar pattern of expression in response to the ingestion of protease inhibitors. On the contrary, three different expression patterns were

obtained for the four serine protease genes (of a total of 70 identified in the genome of *T. urticae*) analyzed in this study: higher expression in mites ingesting the cystatin HvCPI-6 or the CMe trypsin inhibitor for one gene; higher expression of two genes only when mites ingested the CMe trypsin inhibitor; and one gene whose expression was not affected by the ingestion of any of the two inhibitors. As indicated above, serine proteases do not appear to be directly involved in the hydrolysis of dietary proteins in *T. urticae* and, therefore, it is expected that they would be differently regulated depending on their specific physiological roles. Nevertheless, we cannot discard the possibility that some serine proteases may participate in the processing of other proteases involved in digestion. If this is the case, the inhibition of these serine proteases by the CMe trypsin inhibitor would have an effect in the processing of other digestive proteases, being a plausible explanation for the reduction in cathepsin L- and B-like activities observed in mites that have fed on Arabidopsis plants expressing CMe (Santamaría et al., 2012b).

Our results indicate that cysteine and aspartyl proteases are involved in proteolytic digestion in *T. urticae*, being legumain-like activity characterized for the first time in a phytophagous mite. Besides, we have shown that *T. urticae* regulates the expression of these genes in response to the ingestion of protease inhibitors to compensate for inhibitor-targeted enzymes, though they may also act as a direct barrier against ingested plant defensive proteins. Further studies are required to establish the mechanisms by which these different enzymes contribute to dietary digestion.

Acknowledgements

This work was supported by a grant from CSIC (grant CSIC-201040E049 to F.O), Ministerio de Economía e Innovación (grant AGL11-23650 to I.D.) and the Government of Canada through Genome Canada and the Ontario Genomics Institute (grant OGI–046 to V.G.), and the Ontario Research Fund-Global Leadership in Genomics and Life Sciences (grant GL2–01–035 to V.G.). M.E.S and J. G. C. were recipient of post-doctoral grants from the subprogram Juan de la Cierva 2012 and the CSIC (JAE doc), respectively. We are grateful to Dr. Miodrag Grbic (UWO, Canada) for critical review of the manuscript. Thanks are also due to Miguel González Ximénez de Embún (CIB, CSIC) for the isolation of mite faeces and to the Proteomics and Genomics facility at CIB for their assistance in the interpretation of the proteomic results.

References

Alim, M.A., Tsuji, N., Miyoshi, T., Islam, M.K., Huang, X., Motobu, M., Fujisaki, K. 2007. Characterization of asparaginyl endopeptidase, legumain induced by blood feeding in the ixodid tick *Haemaphysalis longicornis*. Insect Biochem. Mol. Biol. 37, 911-922.

Alim, M.A., Tsuji, N., Miyoshi, T., Islam, M.K., Huang, X., Hatta, T., Fujisaki, K. 2008. HlLgm2, a member of asparaginyl endopeptidases/legumains in the midgut of the ixodid tick *Haemaphysalis longicornis*, is involved in blood-meal digestion. J. Insect Physiol. 54, 573-585.

Anderson, J.M., Sonenshine, D.E., Valenzuela, J.G. 2008. Exploring the mialome of ticks: an annotated catalogue of midgut transcripts from the hard tick, *Dermacentor variabilis* (Acari: Ixodidae). BMC Genom. 9, 552.

Bown, D.P., Wilkinson, H.S., Gatehouse, J.A. 1997. Differentially regulated inhibitor-sensitive and insensitive protease genes from the phytophagous insect pest, *Helicoverpa armigara*, are members of complex multigene families. Insect Biochem. Molec. Biol. 27, 625-638.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248-254.

Caffrey, C.R., Mathieu, M.A., Gaffney, A.M., Salter, J.P., Sajid, M., Lucas, K.D., Franklin, C., Bogyo, M., McKerrow, J.H. 2000. Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*. FEBS Lett. 466, 244-248.

Carrillo, L., Martinez, M., Ramessar, K., Cambra, I., Castañera, P., Ortego, F., Diaz, I. 2011a. Expression of a barley cystatin gene in maize enhances resistance against phytophagous mites by altering their cysteine-proteases. Plant Cell Rep. 30, 101–112.

Carrillo, L., Martinez, M., Álvarez-Alfageme, F., Castañera, P., Smagghe, G., Diaz, I., Ortego, F. 2011b. A barley cysteine-proteinase inhibitor reduces the performance of two aphid species in artificial diets and transgenic arabidopsis plants. Transgenic Res. 20, 305-319.

Chen, J.M., Dando, P.M., Rawlings, N.D., Brown, M.A., Young, N.E., Stevens, R.A., Hewitt, E., Watts, C., Barrett, A. J. 1997. Cloning, isolation, and characterization of

mammalian legumain, an asparaginyl endopeptidase. J. Biol. Chem. 272, 8090-8098.

Chen, J.M., Fortunato, M., Barrett, A.J., 2000. Activation of human prolegumain by cleavage at a C-terminal asparagine residue. Biochem J. 352, 327–334.

Chi, Y.H., Salzman, R.A., Balfe, S., Ahn, J.E., Sun, W., Moon, J., Yun, D.J., Lee, S.Y., Higgins, T.J.V., Pittendrigh, B., Murdock, L.L., Zhu-Salzman, K. 2009. Cowpea bruchid midgut transcriptome response to a soybean cystatin – costs and benefits of counter-defence. Insect Mol. Biol. 18, 97-110.

Chougule, N.P., Giri, A.P., Sainani, M.N., Gupta, V.S. 2005. Gene expression patterns of *Helicoverpa armigera* gut proteases. Insect Biochem. Molec. Biol. 35, 355-367.

Colloff, M. J. 2009. Dust Mites. CSIRO Publishing. Springer, Dordrecht, The Netherlands. 583 pp.

Delcroix, M., Sajid, M., Caffrey, C.R., Lim, K.C., Dvorák, J., Hsieh, I., Bahgat, M., Dissous, C., McKerrow, J.H., 2006. A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. J. Biol. Chem. 281, 39316–39329.

Dumez, M.E., Herman, J., Campizi, V., Galleni, M., Jacquet, A., Chevigné, A. 2014. Orchestration of an uncommon maturation cascade of the house dust mite protease allergen quartet. Front. Immunol. 5: 138.

Erban, T., Hubert, J. 2010a. Comparative analysis of proteolytic activities in seven species of synanthropic acaridid mites. Arch. Insect Biochem. Physiol. 75, 187-206.

Erban, T., and Hubert, J. 2010b. Comparative analysis of proteolytic activities in seven species of synanthropic acaridid mites. Arch. Insect Biochem. Physiol. 75, 187-206.

Erban, T., Hubert, J. 2011. Visualization of protein digestion in the midgut of the acarid mite *Lepidoglyphus destructor*. Arch. Insect Biochem. Physiol. 78, 74-86.

Erban, T., Hubert, J. 2015. Two-dimensional gel proteomic analysis of *Dermatophagoides farinae* feces. Exp. Appl. Acarol. 65, 73-87.

Evans, G.O. 1992. Principles of Acarology. CAB International, Wallingford, UK.

Grbić, M., Van Leeuwen, T., Clark, R., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E., Dermauw, W., Thi Ngoc, P.C., Ortego, F., Hernández-Crespo, P., Diaz, I., Martinez, M., Navajas, M., Sucena, E., Magalhães, S., Nagy, L., Pace, R.M., Djuranović, S., Smagghe, G., Iga, M., Christiaens, O., Veenstra, J.A. Ewer, J., Mancilla-Villalobos, R., Hutter, J.L., Hudson, S.D., Velez, M., Yi, S.V., Zeng, J., Pires-daSilva, A., Roch, F., Cazaux, M., Navarro, M., Zhurov, V., Acevedo, G., Bjelica, A., Fawcett, J.A., Bonnet, E., Martens, C., Baele, G., Wissler, L., Sanchez-Rodriguez, A., Tirry, L.,

Blais, C., Demeestere, K., Henz, S.R., Gregory, R., Mathieu, J., Verdon, L., Farinelli, L., Schmutz, J., Lindquist, E., Feyereisen, R., Van de Peer, Y. 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. Nature 479, 487–492.

Hamilton, K.A., Nisbet, A.J., Lehane, M.J., Taylor, M.A., Billingsley, P.F. 2003. A physiological and biochemical model for digestion in the ectoparasitic mite, *Psoroptes ovis* (Acari: Psoroptidae). Int. J. Parasitol. 33, 773-785.

Herman, J., Thelen, N., Smargiasso, N., Mailleux, A. C., Luxen, A., Cloes, M., De Pauw, E., Chevigné, A., Galleni, M., Dumez, M. E. 2014. Der p 1 is the primary activator of Der p 3, Der p 6 and Der p 9 the proteolytic allergens produced by the house dust mite *Dermatophagoides pteronyssinus*. Biochim Biophys Acta. 1840, 1117-1124.

Horn, M., Nussbaumerova, M., Sanda, M., Kovarova, Z., Srba, J., Franta, Z., Sojka, D., Bogyo, M., Caffrey, C. R., Kopacek, P., Mares, M. 2009. Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. Chem. Biol. 16, 1053-1063.

Hoy, M.A., Yu, F., Meyer, J.M., Tarazona, O.A., Jeyaprakash, A., Wu, K. 2013. Transcriptome sequencing and annotation of the predatory mite *Metaseiulus occidentalis* (Acari: Phytoseiidae): a cautionary tale about possible contamination by prey sequences. Exp. Appl. Acarol. 59, 283-296.

Hubert, J., Nesvorna, M., Erban, T. 2014. Growth-suppressive effect of the a-amylase inhibitor of *Triticum aestivum* on stored-product mites varies by the species and type of diet. Exp. Appl. Acarol. 62, 57-65.

Jongsma, M.A., Bolter, C. 1997. The adaptation of insects to plant protease inhibitors. J. Insect Physiol. 43, 885-895.

Julián, I., Gandullo, J., Santos-Silva, L.K., Diaz, I., Martinez, M. 2013. Phylogenetically distant barley legumains have a role in both seed and vegetative tissues. J. Exp. Bot. 64, 2929-2941.

Martinez, M., Diaz-Mendoza, M., Carrillo, L., Diaz, I. 2007. Carboxy terminal extended phytocystatins are bifunctional inhibitors of papain and legumain cysteine proteinases. FEBS Lett. 581, 2914-2918.

Martinez, M., Cambra, I., Carrillo, L., Diaz-Mendoza, M., and Diaz, I. 2009. Characterization of the entire cystatin gene family in barley and their target cathepsin L-like cysteine-proteases, partners in the hordein mobilization during seed germination. Plant Physiol. 151, 1531-1545.

Müntz, K., Shutov, A.D. 2002. Legumains and their functions in plants. Trends Plant Sci. 7, 340-344.

Nisbet, A.J., Billingsley, P.F. 2000. A comparative survey of the hydrolytic enzymes of ectoparasitic and free-living mites. Int. J. Parasitol. 30, 19–27.

Ortego, F. 2012. Physiological adaptations of the insect gut to herbivory. In: Smagghe, G., Diaz, I. (Eds.), Arthropod-plant interactions. Novel Insights and Approaches for IPM. Springer, Dordrecht, Volume 14, pp. 75-88.

Ortego, F., Sánchez-Ramos, I., Ruiz, M., Castañera, P. 2000. Characterization of proteases from a stored product mite, *Tyrophagus putrescentiae*. Arch. Insect Biochem. Physiol. 43, 116-124.

Ortego, F., Novillo, C., Sánchez-Serrano, J.J., Castañera, P. 2001. Physiological response of Colorado potato beetle and beet armyworm larvae to depletion of wound-inducible proteinase inhibitors in transgenic potato plants. J. Insect Physiol. 47, 1291-1300.

Santamaría, M.E., Hernández-Crespo, P., Ortego, F., Grbic, V., Grbic, M., Diaz, I., Martinez, M. 2012a. Cysteine peptidases and their inhibitors in *Tetranychus urticae*: a comparative genomic approach. BMC Genom. 13, 307.

Santamaría, M.E., Cambra, I., Martinez, M., Pozancos, C., González-Melendi, P., Grbic, V., Castañera, P., Ortego, F., Diaz, I. 2012b. Gene pyramiding of peptidase inhibitors enhances plant resistance to the spider mite *Tetranychus urticae.* PLoS One 7, e43011.

Schicht, S., Qi, W., Poveda, L., Strube, C. 2013. The predicted secretome and transmembranome of the poultry red mite *Dermanyssus gallinae*. Parasit. Vectors 6, 259.

Sobotnik, J., Alberti, G., Weyda, F., Hubert, J. 2008. Ultrastructure of the digestive tract in *Acarus siro* (Acari: Acaridida). J. Morphol. 269, 54-71.

Sojka, D., Hajdusek, O., Dvorak, J., Sajid, M., Franta, Z., Schneider, E. L., Craik, C. S., Vancova, M., Buresova, V., Bogyo, M., Sexton, K. B., McKerrow, J. H., Caffrey, C. R., Kopácek, P. 2007. IrAE - An asparaginyl endopeptidase (legumain) in the gut of the hard tick *Ixodes ricinus*. Int. J. Parasitol. 37, 713-724.

Sojka, D., Franta, Z., Horn, M., Caffrey, C. R., Mares, M., Kopácek, P. 2013. New insights into the machinery of blood digestion by ticks. Trends Parasitol. 29, 276-285.

Stewart, G.A., Lake, F.R., Thompson, P.J. 1991. Faecally derived hydrolytic enzymes from *Dermatophagoides pteronyssinus*: Physicochemical characterization of potential allergens. Int. Arch. Allergy Appl. Immunol. 95, 248-256.

Stewart, G.A., Bird, C.H., Krska, K.D., Colloff, M.J., Thompson, P.J. 1992. A comparative study of allergenic and potentially allergenic enzymes from *Dermatophagoides pteronyssinus*, *D. farinae* and *Euroglyphus maynei*. Exp. Appl. Acarol. 16, 165-180.

Thomas, B., Heap, P., Carswell, F. 1991. Ultrastructural localisation of the allergen Der p I in the gut of the house dust mite *Dermatophagoides pteronyssinus*. Int. Arch. Allergy Appl. Immunol. 94, 365-367.

Tovey, E.R., Chapman, M.D., Platts-Mills, T.A.E. 1981. Mite feces are a major source of house dust allergens. Nature 289, 592-593.

Yang, L., Fang, Z., Dicke, M., van Loon, J.J., Jongsma, M.A. 2009. The diamondback moth, *Plutella xylostella*, specifically inactivates Mustard Trypsin Inhibitor 2 (MTI2) to overcome host plant defence. Insect Biochem. Mol. Biol. 39, 55–61.

Zhu-Salzman, K., Koiwa, H., Salzman, R.A., Shade, R.E., Ahn, J. E. 2003. Cowpea bruchid *Callosobruchus maculatus* uses a three-component strategy to overcome a plant defensive cysteine protease inhibitor. Insect Mol. Biol. 12, 135-145.

Table 1

Oligonucleotide primers used for RT-qPCR of *T. urticae* genes.

Family	Gene Name	Tetur ID[1]	Oligonucleotide sequences (5' to 3')	
			Forward	Reverse
Ribosomal protein	Ribosomal protein 49	tetur18g03590	CTTCAAGCGGCATCAGAGC	CGCATCTGACCCTTGAACTTC
Cathepsin L	TuPap-49	tetur09g04400	GCTTTCTCTGCCATTGCTTC	TTTCCATGAGCAATCCATCA
	TuPap-38	tetur12g01820	CCATTGCCTCTGTTGAAGGT	CCTTGGTTTCCCTCATCGTA
	TuPap-41	tetur12g01860	CCGCTGTTGGTTACGGTACT	CTCTGGCCATTCGGAAGTAA
	TuPap-42	tetur25g00650	AAATCAACCGAATGCTCACC	AGCTCCCCAACTGTTCTTGA
Cathepsin B	TuPap-17	tetur01g05480	TCGATGCTCGTGTAGTTTGG	TCAGCAATTGAACCAACAGC
	TuPap-12	tetur08g05020	GAACTTGGTGGCCATGCTAT	CAATCGGTGTTCCAGGAGTT
Legumain	TuLeg-10	tetur05g04700	CAATTCTTTTCGCCCTTTCA	TTGACCCAGCAACCAATACA
	TuLeg-5	tetur28g01760	TGAAAATCCCCATCCAGGTA	GTTTTCGGCGTTGACTTCAT
Aspartyl protease	TuPep-2	tetur14g03010	CTTTGGGTTCCTTCGTCAAA	CCATTGGGTTTGTAGGTGCT
Serine protease	SP7	tetur07g00150	CGATCCAAGCCTCATTCAGT	CTGGTGAGGAAAGGGTCAAA
	SP28	tetur12g03940	CGCTTGTTCAGGTGATTCAG	GTCCATGAACCTGAGACAAC
	SP29	tetur12g03950	TGCTTGTTCAGGTGATTCAG	GTCCATGAACCAGAAACAAC
	SP35	tetur19g00740	GGAAAAGATTCATGCCAAGG	CTTCGCCCCAAGAAACAATA

[1] OrcAE website (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur).

Table 2

Hydrolysis of specific substrates for aspartyl and cysteine (cathepsin B, cathepsin L and legumain) proteases by bodies and faeces of *T. urticae* reared on different plants.

			Specific activity[1] (nmol/min/mg) ± SE					
			Bean		Tomato		Maize	
Protease	Substrate	pH	Bodies	Faeces	Bodies	Faeces	Bodies	Faeces
Aspartyl	MocAc-GKPILFFR LK(Dnp)-D-R-NH_2	3.5	9.8±3.4 ^{a}	49.6±6.1 ^{b}	9.2±1.6 ^{a}	43.9±4.3 ^{b}	10.9±1.3 ^{a}	51.3±4.2 ^{b}
Cathepsin L	Z-FR-AMC	5.5	15.5±2.7 ^{a}	32.0±4.5 ^{b}	16.2±2.7 ^{a}	32.7±1.2 ^{b}	22.3±1.2 ^{a}	35.7±0.3 ^{b}
Cathepsin B	Z-RR-AMC	5.5	2.99±0.10 ^{a}	4.36±0.36 ^{ab}	4.80±0.72 ^{ab}	4.37±0.05 ^{ab}	5.82±0.44 ^{b}	5.43±0.15 ^{b}
Legumain	Z-VAN-AMC	4.5	0.29±0.06 ^{a}	0.36±0.14 ^{a}	0.43±0.03 ^{a}	0.31±0.06 ^{a}	0.31±0.04 ^{a}	0.32±0.03 ^{a}

[1] Values are the mean ± SE of three and two biological replicates for bodies and faeces, respectively. Different letters indicate significant differences between samples for each proteolytic activity ($P \leq 0.05$, one-way ANOVA followed by Student-Newman-Keuls multiple comparison test).

Table 3

Proteases identified in *T. urticae* faecal extracts by nanoLC-nanoESI-QTOF MS/MS.

Family	Protease Gene	Tetur ID[1]	Mascot Protein Score	Number of Peptides	Peptide sequence	Sequence Coverage (%)
Aspartyl protease	TuPep-2	tetur14g03010	150	5	YYTIFDR	14.3
					FDGILGLGFK	
					LPDITFSIAGK	
					TEPALSFVLAK	
					FNVIFDTGSSNLWVPSSK	
Cathepsin L	TuPap-42	tetur25g00650	101	3	VPSSVDWR	13.1
					NSWGADFGEK	
					GASLSSYVDIPQGNENALLSAVAER	
	TuPap-41	tetur12g01860	47	2	DFWIIK	4.8
					NSWGAGWGEK	
	TuPap-49	tetur09g04400	39	1	DFGESGYFR	2.7
	TuPap-38	tetur12g01820	37	1	NSWGTTWGEAGYAR	4.3

[1] OrcAE website (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur

Figure legends

Figure 1. A) Effect of inhibitors on the hydrolysis of specific substrates for cathepsin B (Z-Arg-Arg-AMC), cathepsin L (Z-Phe-Arg-AMC) and legumain (Z-Val-Ala-Asn-AMC) by *T. urticae* body extracts. Protease inhibitors: IAA: Iodoaceamide, E-64: L-trans-epoxysuccinyl-leucylamido-(4-guanidino)-butane. **B)** Effect of removing DTT from the reaction buffer and/or adding 0.01 % Brij35 to the homogenate on the hydrolysis of a specific substrate for legumain (Z-Val-Ala-Asn-AMC) by *T. urticae* body extracts. Values are the mean ± standard error of at least triplicated experiments. Different letters indicate significant differences between treatments for each proteolytic activity ($P \leq 0.05$, one-way ANOVA followed by Student-Newman-Keuls multiple comparison test).

Figure 2. Expression of genes in *T. urticae* at different developmental stages. Genes coding for **A)**: cathepsin L (TuPap-49, 38, 41 and 42), cathepsin B (TuPap-17 and 12), legumain (TuLeg-10 and 5) and aspartyl (TuPep-2) proteases; and **B)** serine proteases (TuSP7, 28, 29 and 35). Data are the mean ± standard error of duplicate measurements of three replicates for each developmental stage. Different letters indicate significant differences between developmental stages in the analysis of each gene ($P \leq 0.05$, one-way ANOVA followed by Student-Newman-Keuls multiple comparison test).

Figure 3. Gene expression in adults of *T. urticae* feeding on control Arabidopsis plants (Col) and on Arabidopsis transgenic plants over-expressing the barley *Icy6* gene encoding the HvCPI-6 cystatin (CPI6, 5.4 and 9.8 lines) or the barley *Itr1* gene encoding the BTI-CMe trypsin inhibitor (CMe, 3.4 and 8.9 lines). Genes coding for cathepsin L (TuPap-49, 38, 41 and 42), cathepsin B (TuPap-17 and 12), legumain (TuLeg-10 and 5), aspartyl (TuPep-2) and serine (TuSP7, 28, 29 and 35) proteases. Data are the mean ± standard error of duplicate measurements of three replicates for each sample. Different letters indicate significant differences between treatments in the analysis of each gene ($P \leq 0.05$, one-way ANOVA followed by Student-Newman-Keuls multiple comparison test).

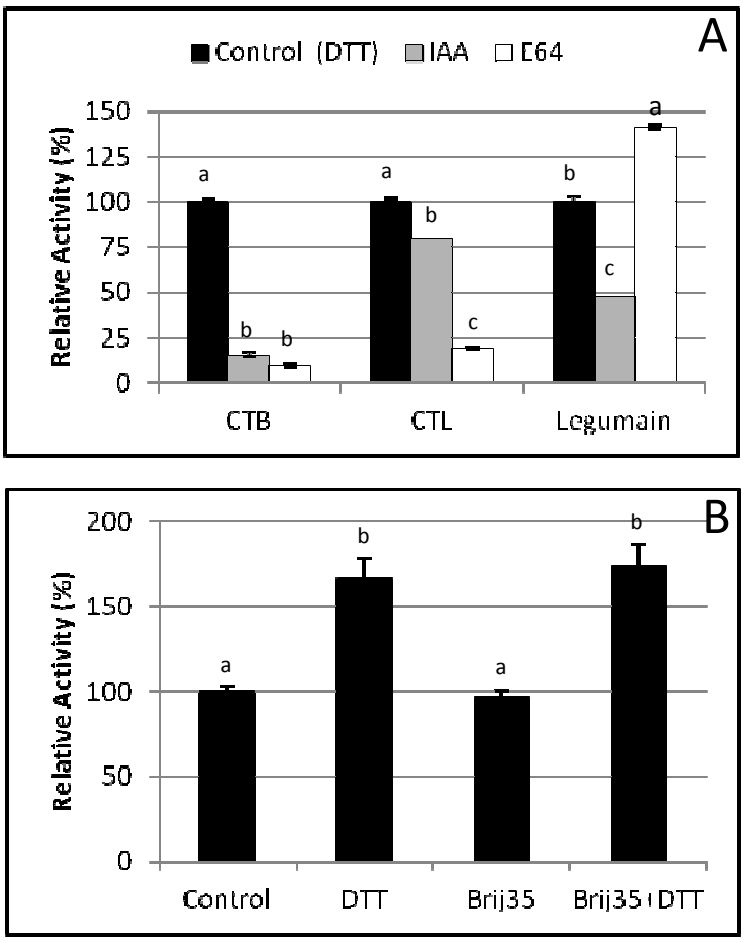

A
Control (DTT)
IAA
E64
Relative Activity (%)
150
125
100
75
50
25
0
a
b
b
a
b
c
b
c
a
CTB
CTL
Legumain
B
Relative Activity (%)
200
150
100
50
0
a
b
a
b
Control
DTT
Brij35
Brij35+DTT

Figure 1

‘

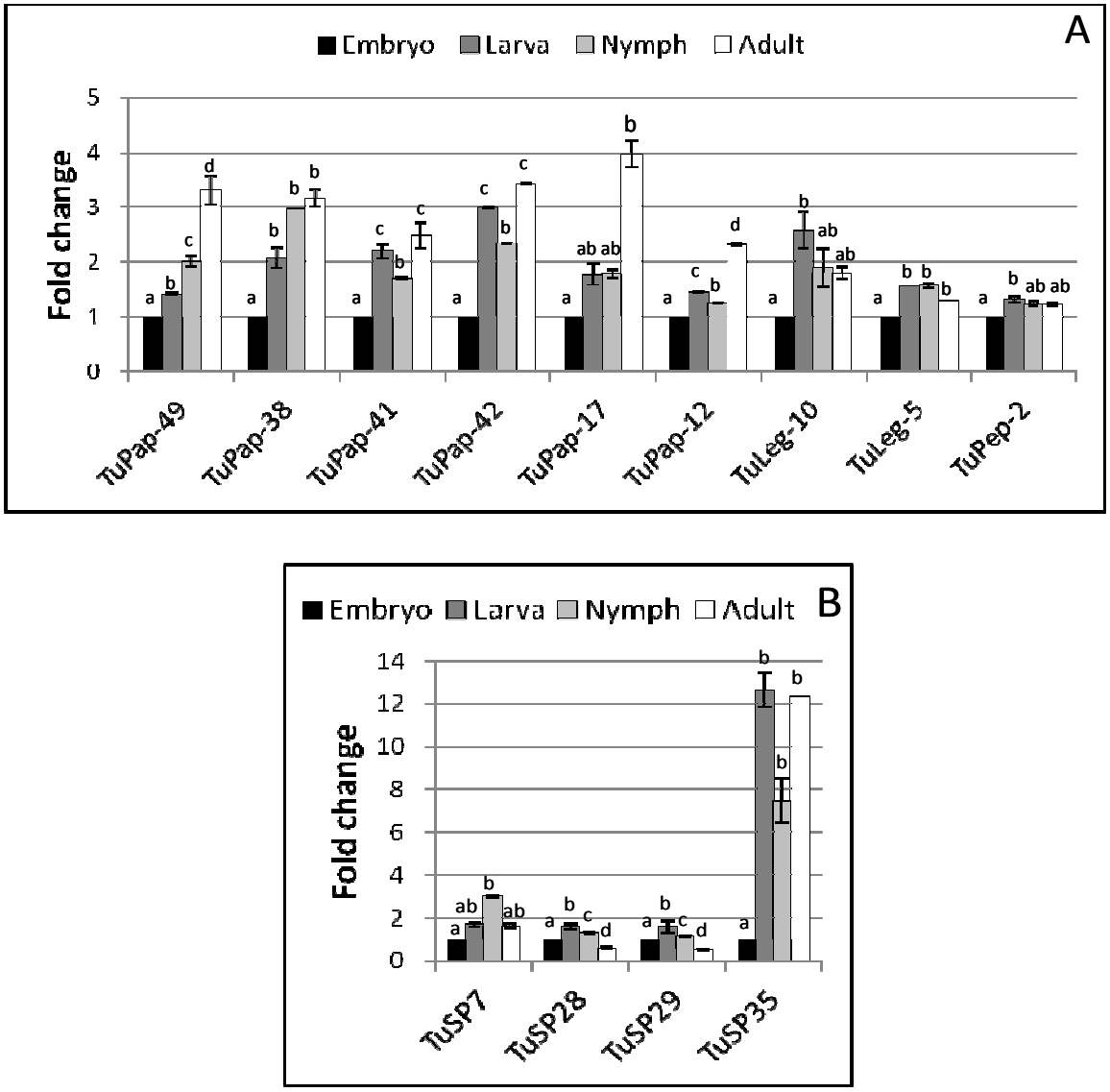

A
Embryo
Larva
Nymph
Adult
Fold change
0
1
2
3
4
5
TuPap-49
TuPap-38
TuPap-41
TuPap-42
TuPap-17
TuPap-12
TuLeg-10
TuLeg-5
TuPep-2
B
Embryo
Larva
Nymph
Adult
Fold change
0
2
4
6
8
10
12
14
TuSP7
TuSP28
TuSP29
TuSP35

Figure 2

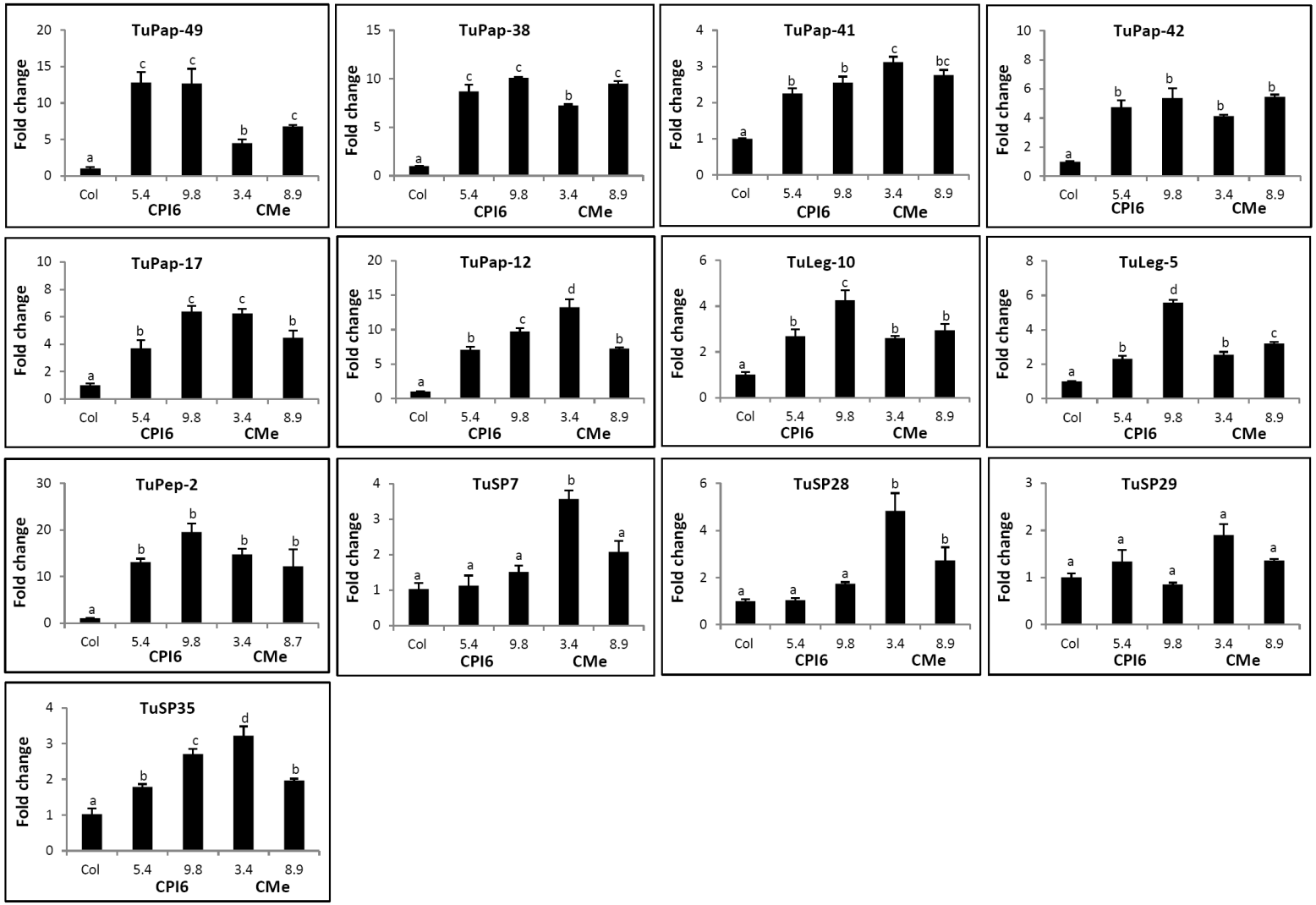

TuPap-49
TuPap-38
TuPap-41
TuPap-42
TuPap-17
TuPap-12
TuLeg-10
TuLeg-5
TuPep-2
TuSP7
TuSP28
TuSP29
TuSP35
Fold change
Col
5.4
9.8
3.4
8.9
8.7
CPI6
CMe

Figure 3

Supplementary Table S1

Other proteins identified in *T. urticae* faecal extracts by nanoLC-nanoESI-QTOF MS/MS.

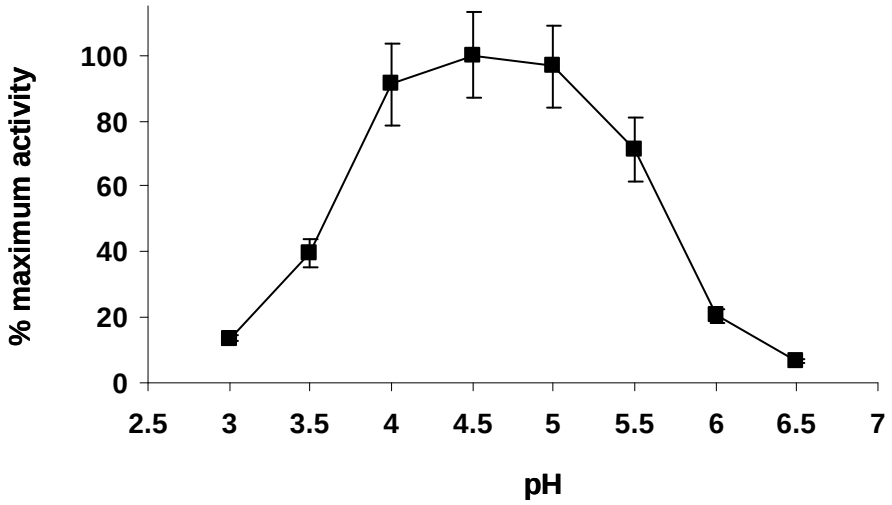
Tetur ID[1]	Mascot Protein Score	Number of Peptides	Sequence Coverage (%)	Description
Enzymes				
tetur41g00210	294	9	21.9	ATP synthase beta subunit
tetur07g07380	148	5	9.8	beta-galactosidase
tetur08g00400	145	3	16.8	Thioredoxin peroxidase
tetur03g07920	92	3	15.6	Glutathione S-transferase, delta class
tetur01g02510	76	3	13.9	Glutathione S-transferase, class delta
tetur07g03440	55	2	8.5	Fructose-bisphosphate aldolase, class I
tetur09g04140	52	2	3.4	V-type proton ATPase catalytic subunit A
tetur04g00130	45	1	2.3	alpha-N-acetylgalactosaminidase
tetur01g06130	45	2	5.7	ATP synthase subunit alpha, mitochondrial precursor
tetur06g03290	39	2	7.4	ADP/ATP translocase
Storage, lipid-binding and cuticular proteins				
tetur39g00810	1106	41	23	Vitellogenin 2
tetur43g00010	888	31	18	Vitellogenin 5
tetur39g00720	390	15	8	Vitellogenin 6
tetur30g01430	180	8	31.3	Apolipoprotein D precursor
tetur01g12820	179	4	30.2	Cuticle protein
tetur01g00130	176	5	65.3	Cuticle protein
tetur05g07070	130	3	15	Lipoprotein Blc
tetur12g02060	117	2	20.4	Cuticle protein
tetur01g05730	91	4	12.1	Apolipoprotein D precursor
tetur04g05980	75	4	18.3	Apolipoprotein D precursor
tetur06g05820	54	1	6.8	Calycin-like
Structural proteins				
tetur09g05350	324	12	26.9	Actin
tetur03g09480	177	7	16.2	Actin
tetur03g00230	117	4	12	alpha-tubulin

tetur03g00330	90	3	2.2	Similar to Clathrin heavy chain 1
Other				
tetur07g00160	398	15	11	Cell wall surface anchor family protein
tetur29g01360	159	8	11.8	Low complexity secreted protein
tetur08g02980	150	4	9.3	Elongation factor 1-alpha
tetur18g03030	144	5	6.7	TEP1
tetur12g00550	80	2	8.4	C-factor
tetur02g04230	71	2	19.4	Histone H4
tetur12g00600	69	2	9.2	C-factor
tetur25g00240	67	2	5.4	Adhesion molecule CD36
tetur04g02800	49	4	0.6	Novel Hypothetical Cell Adhesion Molecule
tetur08g01320	47	2	4.4	Heat shock protein Hsp70
tetur02g01180	45	2	2.3	Guanylate-binding protein, C-terminal
tetur02g08710	44	2	0.7	Telomerase-binding protein EST1A
Unknown				
tetur15g01970	62	2	9.2	Hypothetical protein
tetur21g01470	57	2	6.3	Hypothetical protein
tetur30g00140	46	2	11.5	Hypothetical protein
tetur15g01980	45	2	9.9	Hypothetical protein
tetur01g10180	38	1	5.2	Hypothetical protein
tetur04g07690	38	1	6.8	Protein of unknown function DM9

[1] OrcAE website (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur),

Supplementary Figure S1.

Legumain-like activity in *T. urticae* body extracts using Z-VAN-AMC (N-carbobenzoxyloxy-Val-Ala-Asn -7-amido-4-methylcoumarin) as substrate. The assays were performed at different pHs (100 mM sodium citrate, 150 mM NaCl), containing 1mM dithiothreitol (DTT) Values are the mean of at least three replicates and error bars indicate the standard error of mean.



Supplementary Figure S2.

Comparison of the amino acid sequences of the four cathepsins L identified in *T. urticae* faecal extracts. The alignment was generated using the MUSCLE program. The putative signal peptides and propeptide regions are coloured in red and blue, respectively. A conserved asparagine is highlighted in green.

```
TuPap-38        MFKTILILATLACAFATEVDFKS---EFNKFKATHNKLYTSEKEEAQRLKIFTSNLRKIN
TuPap-41        MLRFALLATLVALTYGYSISVDEIKQQFSAFKTNHSVVYADEKEEAKRFEIFAKNLEFIK
TuPap-49        MIRAAIFIALFALAAAASLSLDS---QWESFKTKYGKSYDSAEEETYRRNFFAQKLEKIK
TuPap-42        MIRTTCFIALFALAAAASLSYEN---EWESFKAKHHKVYAGESEEAWRKSVFLNNLKKIK
                *:.   : : .* : . .:. ..   ::. **:.:   * . .**: * ..* .:*  *:

TuPap-38        GHNERHARGMESYQMGVNAYTDLTFKEFSEKLLGYKANHDRVAPLVHNNSTVDGLPDSVD
TuPap-41        NHNEKANNGEHSFWLGVNKFADLTYEEFAARYLNYRGAEKKPAPLLHNSAYFGDLPTSVD
TuPap-49        AHNERAANGEVTYRLGVNKFSDLTAEEFKAKYLGYKPA-RRSAPLIHNVNYTVKAPASVD
TuPap-42        AHNERADNGEHTYRLGVNKFADLTFEEFKAAYLGFKKS-NRSAPLIHNVNKTVRVPSSVD
                 ***.  .*  ::.:*** ::*** :**    *.:.    . ***:**       * ***

TuPap-38        WRTKGIVNPIKDQAQCGSCWAFSTIASVEGAWAQAKGKLVSLSEQNLMDCSYDEGNQGCN
TuPap-41        WREKGAVTPVKDQGSCGSCWAFSAVAGIEGANFLKNGKLVSLSEQNLVDCAKDDVDQGCN
TuPap-49        WRTKGIVSEIKDQQQCGSCWAFSAIASIESANAQKNGKLVELSEQNLMDCSWKYGDNGCG
TuPap-42        WRTKGIVAEVKDQASCGSCWAFSAIASLESANAQKTGKLVELSEENLIECSWNYGNQGCN
                ** ** *  :*** .********::*.:*.*.   .****.***:**::*: .  ::**.

TuPap-38        GGLPDNAFEYIIKNGGVDTEDSYGYSAVSSNA---CKYSKANEGATISKYVDIPSRDEKA
TuPap-41        GGLMDHAFTYVINNKGIDTESSYPYKGADGS----CAFKSSDVGATISSFVDVESGNEKA
TuPap-49        GGLMDDAFSYVKANKGVDTEKSYPYISGDGEDYHKCRYTASNKGASISSFVDVPEADEKS
TuPap-42        GGLMDDAFNYIKANKGIDTEKSYPYTSGSGDDSGRCKYSASNKGASLSSYVDIPQGNENA
                *** * ** *:  * *:***.** * . ...    * :. :: **::*.:**: . :*::

TuPap-38        LKNAVAERVVSVAIHVGDSFQHYKSGIYN-DKFCPGFSFFLNHAVNVVGYGT----DHWI
TuPap-41        LATAVALQPVSVAIDASIFLQLYFGGVFDIGLLCSSSADKLNHGVTAVGYGT-GSKDFWI
TuPap-49        LLSAVAERVVSVAIDAG-PVQDYESGILN-TDECSSDPQDLDHGVAVVGYGTEDGTPYWI
TuPap-42        LLSAVAERVVSVAIDAEEDLQLYQSGVLK-STECS--PDALNHGVTAVGYGVDNGTPYWL
                * .*** . ***** .   .* * .*: .    *.  .  *:*.* .****.     .*:

TuPap-38        VRNSWGTTWGEAGYARFAINKNLCGIATKASYPLV
TuPap-41        IKNSWGAGWGEKGYFRMARGKNLCGVATQASYVVV
TuPap-49        VKNSWGKDFGESGYFRLYRGNNMCGIADSASYPIV
TuPap-42        IKNSWGADFGEKGYFRLYRGSNMCGVAESASYPIV
                :.****  :** ** *:  ..*:**:* .*** :*
```