The salivary proteome of Tetranychus urticae
Jonckheere, W.S.A.

Citation for published version (APA):
Jonckheere, W. S. A. (2018). The salivary proteome of Tetranychus urticae: Key to its polyphagous nature?

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
The salivary protein repertoire of *Tetranychus urticae*

Redrafted after:
Jonckheere, W., Dermauw, W., Zhurov, V., Wybouw, N., Van den Bulcke, J., Villaroel, C.,
Greenhalgh, R., Grbic, M., Schuurink, R.C., Tirry, L., Baggerman, G., Clark, R.M., Kant,
repertoire of the polyphagous spider mite *Tetranychus urticae* a quest for effectors.
*Molecular & Cellular Proteomics* 15, 3594-3613
3.0. Abstract

The two-spotted spider mite *Tetranychus urticae* is an extremely polyphagous crop pest. Alongside an unparalleled detoxification potential for plant secondary metabolites, it has recently been shown that spider mites can attenuate or even suppress plant defenses. Salivary constituents, notably effectors, have been proposed to play an important role in manipulating plant defenses and might determine the outcome of plant-mite interactions. Here, the proteomic composition of saliva from *T. urticae* lines adapted to various host plants – bean, maize, soy, and tomato – was analyzed using a custom-developed feeding assay coupled with nano-LC tandem mass spectrometry. About 90 putative *T. urticae* salivary proteins were identified. Many are of unknown function, and in numerous cases belonging to multi-membered gene families. RNAseq expression analysis revealed that many genes coding for these salivary proteins were highly expressed in the proterosoma, the mite body region that includes the salivary glands. A subset of genes encoding putative salivary proteins was selected for whole-mount *in situ* hybridization, and were found to be expressed in the anterior and dorsal podocephalic glands. Strikingly, host plant dependent expression was evident for putative salivary proteins, and was further studied in Chapter 4. This multidisciplinary approach revealed for the first time the salivary protein repertoire of a phytophagous chelicerate. The availability of this salivary proteome will assist in unraveling the molecular interface between phytophagous mites and their host plants, and may ultimately facilitate the development of mite-resistant crops. Furthermore, the technique used in this study is a time- and resource-efficient method to examine the salivary protein composition of other small arthropods for which saliva or salivary glands cannot be isolated easily.
3.1. Introduction

The family of spider mites (Chelicerata: Acari: Tetranychidae) comprises well over 1000 species, including several that are important pests on crops, and about 0.9 billion euro is being spent annually for their control worldwide (Migeon and Dorkeld, 2006-2016; Van Leeuwen et al., 2015). These minute herbivores – about 0.5 mm in size – use their stylets to pierce leaf mesophyll cells and to inject saliva, after which they suck out the cytoplasm. This results in cell death visible as chlorotic spots sometimes accompanied by necrosis, and ultimately in leaf abscission (Liesering, 1960; Tomczyk and Kropczynska, 1985). Among the spider mites, the two-spotted spider mite, *Tetranychus urticae*, is the most polyphagous, having been reported on more than 1000 host plant species in more than 140 different families (Migeon and Dorkeld, 2006-2016). However, not all these host plants are equally suitable to *T. urticae*, and host plant acceptance can even differ across mite populations (Kant et al., 2008; van den Boom et al., 2003; Yano et al., 1998).

Important factors determining host plant acceptance by the herbivore are plant defenses, including physical and molecular-chemical barriers that hamper herbivore feeding (Kant et al., 2015). Different herbivores can induce a different repertoire of defenses and these differential plant responses are set in motion via herbivore-specific signals, predominantly emanating from their saliva (Bonaventure et al., 2011). Plant defenses are regulated by a set of phytohormones, primarily jasmonates [such as jasmonic acid (JA)] (Wasternack and Hause, 2013), salicylic acid (SA) (Caarls et al., 2015; Mur et al., 2006) and ethylene (Pieterse et al., 2009). Hormonal interactions are believed to enable the plant to regulate and customize responses under variable biotic and abiotic stress conditions (Erb et al., 2012). Most spider mites induce a cocktail of JA- and SA-defenses (Alba et al., 2015; Glas et al., 2014; Kant et al., 2004; Matsushima et al., 2006; Zhurov et al., 2014) while a role for ethylene remains elusive (Kielkiewicz, 2002).

It is conceivable that some spider mites have evolved traits that enable them to resist (Dermauw et al., 2013b; Kant et al., 2008; Wybouw et al., 2014), attenuate (Wybouw et al., 2015) or suppress JA- (Kant et al., 2008) and SA-related defenses (Sarmento et al., 2011) to maintain a high fitness (Alba et al., 2015). Although it is largely unknown which terminal plant defenses determine resistance or susceptibility to mites, negative correla-
tions were found between mite fitness and several plant secondary metabolites (Bleeker et al., 2012; Chatzivasileiadis and Sabelis, 1997; Jared et al., 2015; Zhurov et al., 2014). How plants detect spider mite feeding is poorly understood, but analyses of transcriptional networks have suggested the involvement of receptor-like kinases reminiscent of other plant-herbivore interactions (VanDoorn and de Vos, 2013). These receptors may be involved in the recognition of molecules (elicitors) released during the onset of the plant-pathogen or plant-herbivore interaction (Martel et al., 2015).

Many herbivore elicitors emanate from saliva or regurgitation fluids released on or in the plant during feeding (Schmelz et al., 2009). Reminiscent of phytopathogens (Boller and Felix, 2009; Bonaventure et al., 2011; Jones and Dangl, 2006), herbivores evolved additional salivary molecules to counter the induction of defenses (Acevedo et al., 2015; Hogenhout and Bos, 2011; Kant et al., 2015). Such molecules, enhancing herbivore performance, were originally referred to as ‘effectors’. Some plant varieties have, however, evolved the means to recognize these effectors, effectively turning them into elicitors which activate plant defense responses (Hogenhout et al., 2009; Jones and Dangl, 2006; Lapin and Van den Ackerveken, 2013). Due to this context-dependency (Schmelz et al., 2012), a broader inclusive definition of the term effector was suggested (Hogenhout et al., 2009). Effectors are defined as pathogen- or herbivore-secreted proteins and small molecules that alter host-cell structure and function. Effectors are of high interest to the plant breeding industry because they can lead to the identification of resistance genes (R genes) (Bent and Mackey, 2007) and susceptibility genes (S genes) (van Schie and Takken, 2014). R genes code for immune receptors which confer recognition of pathogen- or herbivore-derived effectors or their modification inflicted on a host protein, eventually resulting in the activation of host defenses (van Schie and Takken, 2014). S genes, on the other hand, can considered to be all plant genes that facilitate infection and support compatibility (van Schie and Takken, 2014).

The vast majority of herbivore effectors emanate from saliva. Silencing salivary effectors in non-arthropod herbivores like nematodes has been shown to reduce their performance (Elling and Jones, 2014). Likewise, silencing salivary effectors in insects like aphids reduced their reproduction (Coleman et al., 2015). These studies indicate that salivary components are key players in the plant-herbivore molecular battlefield, and hence their

Proteomic composition of T. urticae saliva
identification is a high priority. Salivary proteins can be inferred from genomic, transcriptomic and/or proteomic data using a combination of criteria. For example, combining temporal and spatial gene expression data with the predicted presence of an N-terminal signal peptide (SP) in the corresponding proteins results in lists of putative salivary proteins (e.g., Bos et al., 2010; Carolan et al., 2011). For *T. urticae*, an annotated genome is available (Grbić et al., 2011b), but no salivary gland-specific transcriptome and/or proteome has been obtained yet. It is known that spider mites inject salivary substances into host plant leaves (Avery and Briggs, 1968; Rodriguez, 1954; Storms, 1971). However, the proteomic composition of these substances has yet to be elucidated. The generation of gland specific transcriptomes and proteomes is hampered by the extremely small size of spider mites and the complex morphology of the glands (Mothes and Seitz, 1981) (*T. urticae* adults have a body length of 400-500 μm with an approximate salivary gland length of 50 μm). Salivation of several eriophyid mite species has been achieved by soaking adult mites into immersion oil (De Lillo and Monfreda, 2004), and of *Varroa destructor* mites by topical application of cholinomimetic agents (Richards et al., 2011). Protein sequences were not obtained in these studies, however. A successful approach for obtaining sufficient amounts of salivary secretions suitable for protein analysis from non-mite arthropods has been to collect secretions from artificial diets encapsulated by a membrane on which feeding has taken place. For example, using this approach, multiple proteins, in a range from 10 to 100, have been identified in the secreted saliva of aphids (Nicholson et al., 2012; Vandermoten et al., 2014) and true bugs (Cooper et al., 2013).

We developed a set-up for collecting salivary secretions of *T. urticae* from artificial diet and analyzed the proteomic composition of these secretions. Our approach involved *T. urticae* lines that were reared on distinct economically important host plants for more than five generations, a period during which adaptation usually occurs (Fry, 1989). By including lines adapted to different hosts, we aimed to discover a broader spectrum of salivary proteins. Mite salivary secretions were harvested using a custom-developed mite feeding assay and subsequently investigated by nano-LC-MS/MS analysis. Additionally, a transcriptome of the proterosoma-har- boring the salivary glands- was constructed to validate proteomic data. Evidence for the salivary origin of a selection of identified proteins was obtained by whole-mount *in situ* hybridizations (ISHs). Furthermore, to
assess host-specificity of salivary gland productions, we investigated the host-dependency of expression of genes coding for the identified putative salivary proteins (Chapter 4). The results from this study lay the groundwork for an improved understanding of the molecular machinery behind induction or suppression of resistance during plant-mite interactions, and may open new opportunities for mite-resistance plant breeding.

3.2. Material and Methods

3.2.1. Establishment of T. urticae lines on different host plants

The T. urticae London strain has been maintained under laboratory conditions on bean plants (Phaseolus vulgaris cv. ‘Prelude’, Fabaceae) for many years. The genome of this London strain has been sequenced (Grbić et al., 2011b). Lines on alternative host plants were established by transferring approximately 250 adult female mites from the London strain on bean to new hosts. These new host plants were cotton (Gossypium hirsutum, Malvaceae), maize (Zea mays, cv. ‘Ronaldino’, Poaceae), soy (Glycine max, cv. ‘Merlin’, Fabaceae) and tomato (Solanum lycopersicum, cv ‘Moneymaker’, Solanaceae). Three independent lines were generated for cotton and tomato, while 4 independent lines were obtained for maize and soy. The mite lines were maintained in a climatically controlled environment at 26°C with 60% RH, and a light:dark (L:D) photoperiod of 16:8 h. Mites were offered fresh plants as needed, and were used in experiments after five generations for all hosts, except tomato, where replicate lines derived from London were adapted and maintained on tomato for over 30 generations (Wybouw et al., 2015).

3.2.2. Collection of artificial diet enriched with T. urticae saliva

To collect saliva, spider mites were allowed to feed on an artificial diet. Briefly, a pocket-like invagination was made in stretched Parafilm® M using a custom-built vacuum device (see Figure S3.1), consisting of a 96 well plate (plate thickness 4.2 mm, hole diameter 4.5 mm) fitting on a vacuum manifold plate (Analytical Research Systems, Florida, USA) connected to a vacuum pump (model N 035.1.2 A_.18, KNF Neuberger, Germany). Next, 70 μl sterile holdic artificial diet (1/30 diluted aphid diet; Febvay et al., 1988) supplemented with the antibiotic rifampicin (0.05 mg/ml) was added,
after which the pocket was sealed with packaging tape (Scotch Packaging Tape, Extra Strong, Belgium). The parafilm, with the diet-filled hemisphere side directed upwards, was cut to size (approximately 4 x 4 cm), mounted on the back of a small petri dish (90 mm diameter), and placed in a large petri dish (135 mm diameter) filled with water. Using cotton wool and paper tissue, a water barrier was created, confining the spider mites in close proximity of the diet hemisphere (Figure 3.1). Thirty to 40 adult female mites were transferred to each feeding arena. These mites originated from replicate lines adapted to bean, maize, soy or tomato (mites adapted to cotton were not used for the collection of saliva). For each host plant-specific sample, mites originating from each replicated host plant line were pooled. Addition of a blue colorant (0.05 mg/ml erioglaucine, Sigma) to control diet hemispheres was used to verify spider mite feeding as assessed by staining of gut contents (Figure 3.1). Petri dishes with feeding hemispheres were placed in an incubator (Panasonic MLR-352H) at 26°C with 60% RH and 16:8 L:D. After 24 h, the remaining content of the feeding hemispheres was collected using a Hamilton microsyringe under sterile conditions (ESCO Laminar Flow cabinet). Samples were stored at -80°C until enough sample (about 0.3 ml, 10-15 diet hemispheres per sample) was collected for nano-LC-MS/MS analysis. Feeding hemispheres that did not receive spider mites were treated identically and served as reference sample during nano-LC-MS/MS analysis.

3.2.3. Nano-LC-MS/MS analysis

Twenty μg of total protein (Pierce BCA Protein Assay Kit, Thermo Scientific) of the sample was reduced using 1.25 μl of 500 mM TRIS (2-carboxyethyl) phosphine in a volume of 100 μl 100 mM TEAB, and incubated for 1 h at 55°C. Next, the samples were processed using the filter-aided sample preparation (FASP) procedure (FASP Protein digestion kit, Protein discovery, Knoxville, TN) according to manufacturer’s instructions. In short, the samples were diluted in a urea buffer and processed on a FASP filter, alkylated with iodoacetamide and digested with trypsin (enzyme-protein ratio = 1:50) overnight. Liquid chromatography mass spectrometric analysis was performed on a Waters nanoAquity LC-Ultra system connected to a Thermo Scientific LTQ Velos Orbitrap mass spectrometer. The equivalent of 2 μg of total protein of the digested sample was dissolved in 20 μl of 2% acetonitrile in HPLC-grade water. 10 μl of the sample (1 μg) was loaded on
Proteomic composition of *T. urticae* saliva

the trapping column (Pepmap C18 300 μm x 20 mm, Dionex) with an isocratic flow of 2% acetonitrile in water with 0.1% formic acid at a flow rate of 5 μl min⁻¹. After 2 min, the column-switching valve was switched, placing the pre-column online with the analytical capillary column, a Pepmap C18, 3 μm 75 μm x 150 mm nano column (Dionex). Separation was conducted using a linear gradient from 2% acetonitrile in water, 0.1% formic acid to 40% acetonitrile in water, 0.1% formic acid in 100 min. The flow rate was set at 400 nl min⁻¹. The LTQ Orbitrap Velos (Thermo Scientific) was set up in a data dependent MS/MS mode where a full scan spectrum (350–5000 m/z, resolution 60,000) was followed by a maximum of ten CID tandem mass spectra (100 to 2000 m/z). Peptide ions were selected as the 20 most intense peaks of the MS1 scan. CID scans were acquired in the LTQ iontrap part of the mass spectrometer. The normalized collision energy used was 35% in CID. We applied a dynamic exclusion list of 45 s.

3.2.4. Protein identification

Prior to protein identification, calibration of the data was performed using the methods described by Gibbons et al. (2015), correcting the systematic bias in mass measurement in the 2nd replicate. Peak lists obtained from MS/MS spectra were identified using OMSSA version 2.1.9 (Geer et al., 2004), X!Tandem version X! Tandem Vengeance (2015.12.15.2) (Craig and Beavis, 2004), and MS-GF+ version Beta (v10282) (Kim and Pevzner, 2014). The search was conducted using SearchGUI version 2.8.5 (Vaudel et al., 2011). Protein identification was conducted against a concatenated target/decoy version of the *T. urticae* protein database holding 17,907 target sequences (version of December 16th 2014, see Data S3.1) and the common Repository of Adventitious Proteins (cRAP) database (available at http://www.thegpm.org/crap/). Reversing the target sequences in SearchGUI created the decoy sequences. The identification settings were as follows: trypsin with a maximum of 2 missed cleavages; 10.0 ppm as MS1 and 0.5 Da as MS2 tolerances; variable modifications: carbamidomethyl c (+57.021464 Da), oxidation of m (+15.994915 Da), pyro-glu from n-term q (-17.026549 Da), acetylation of protein n-term (+42.010565 Da), pyro-cmc (-17.026549 Da) and pyro-glu from n-term e (-18.010565 Da). Peptides and proteins were inferred from the spectrum identification results using PeptideShaker version 1.10.2 (Vaudel et al., 2015) [see TABLES S3.1 and S3.2 for a detailed list of all features of the LC-MS/MS identifications at 98]
the protein and peptide level in the treatment samples (artificial diet with feeding mites), respectively, and Table S3.3 for a list of all features of the LC-MS/MS identifications at the protein level in the reference samples (artificial diet without feeding mites). Peptide Spectrum Matches (PSMs), peptides and proteins were validated at a 1.0% False Discovery Rate (FDR) estimated using the decoy/hit distribution.

The mass spectrometry data along with the identification results have been deposited to the ProteomeXchange Consortium (Vizcaino et al., 2014) via the PRIDE partner repository (Martens et al., 2005) with the dataset identifiers PXD003022 and 10.6019/PXD003022.

3.2.5. Experimental design and statistical rationale

Two technical replicates of artificial diets enriched with saliva of mites maintained on four different host plant species were analyzed in this nano-LC-MS/MS study. The different host plants to which the mite lines have been adapted were bean, maize, soy and tomato, as described in 3.2.1. Eight samples were analyzed in total, yielding a list of putative *T. urticae* salivary proteins. Care should be taken when comparing the proteomics data of the different host plant lines, as each host-specific dataset is based on one pooled biological replicate (each analyzed in two technical replicates). However, we believe that our approach was sufficiently rigorous as host plant specific production of *T. urticae* salivary proteins was complemented by a gene expression analysis of mites maintained on the four different host plants (see Chapter 4).

3.2.6. Proteomics data filtering and abundance ranking

Proteins identified in the artificial diet were pooled into one dataset, while retaining host plant specific information. In order to retain confident proteins identifications, only proteins identified in at least one of the samples with a mean PSM value of at least two, calculated over both technical replicates per host plant, were retained in the final putative salivary protein list. A MS1-intensity based Top3 analysis was performed to derive protein abundance (Silva et al., 2006). First, the moFF (modest Feature Finder, https://github.com/compomics/moFF) tool was used to extract the MS1 intensities from the Thermo raw files for the validated PSMs matching to non-decoy (true) sequences (see Table S3.4). Afterwards, the intensities of the three (or fewer) peptides with the highest intensities were averaged for
every protein detected. Next, a normalized abundance factor (rTop3) was calculated by dividing the Top3 value by the sum of all Top3 values of the quantified proteins in each experiment (Krey et al., 2014), excluding contaminants. Additionally, the proteins were ranked according to their maximal rTop3 value, as calculated across each host-plant specific sample.

3.2.7. X-ray sub-micron computed tomography

Adult female *T. urticae* specimens were fixed as described previously (Laumann et al., 2008). Briefly, live mites were collected and then incubated in a 6:3:1 mixture of 80% ethanol, 38% formaldehyde and 100% acetic acid for 72 h. Specimens were washed with 70% ethanol and dehydrated through graded concentrations of 70 to 100% ethanol (5% steps) for 10 minutes per step. The 100% ethanol step was done overnight at room temperature (RT). The dehydrated specimens were critical point dried with CO2 (Bal-Tec CPD 020, Liechtenstein) and glued with their opisthosoma to a 0.5 mm carbon pencil lead (Staedtler, Belgium). Spider mite specimens were scanned with Nanowood (Dierick et al., 2014), an X-ray sub-micron computed tomography (CT) system developed at the Ghent University Centre for X-ray Tomography (UGCT). The sample was scanned with an open-type nano-focus X-ray tube, reaching a focal spot size <1 μm. Samples were scanned for 1 h, resulting in scans with an isotropic voxel pitch of approximately 0.5 μm. Reconstructions were performed using Octopus Reconstruction, a tomography reconstruction package for parallel, cone-beam and helical geometry (Vlassenbroeck et al., 2007), licensed by InsideMatters (www.insidematters.eu). Filtering was performed using the single step phase-retrieval Paganin algorithm (Paganin et al., 2002). After reconstruction, a noise removal anisotropic diffusion filter was applied using Octopus Analysis, formerly known as Morpho+ (Brabant et al., 2011), also licensed by InsideMatters. All visualization was performed with Fiji (Schindelin et al., 2012).

3.2.8. RNAseq expression analysis of the proterosoma of *T. urticae* adult females

RNA was extracted from intact adult *T. urticae* females and from dissected proterosomas of adult females using the Qiagen RNeasy RNA extraction kit (Qiagen, The Netherlands). Three replicates of 100 intact adult females from the London strain were collected from bean plants. For dissection of the proterosoma, a dissection chamber was prepared by affixing a square piece of 12.7
mm width double sided Scotch tape (3M, Maplewood, MN) on the bottom of a 35 mm Petri plate and allowing tape to cure for 24 h. Adult female mites of the London strain were carefully attached to tape with their ventral side downwards and covered with 1x phosphate buffered saline solution (PBS). The dorsal side of the hysterosoma was opened using micro-dissecting needles and contents (gut, ovaries, eggs, dorsal parts of exoskeleton) were removed with dissecting needles and micropipette. Remaining tissue (proterosoma, 1st and 2nd pair of legs, nervous mass, salivary glands) was gently lifted from tape and transferred to Buffer RLT (Qiagen RNeasy RNA extraction kit) using a micropipette. In total, 250 mites were dissected and dissected tissue was pooled into one sample. RNA from this sample, as well as three samples collected from whole adult females, was used for Illumina library construction and subsequent sequencing on a HiSeq instrument (Fasteris, Switzerland). Briefly, 100 bp strand-specific paired-end reads were generated for all samples. The paired-end strand-specific Illumina RNA-seq reads were aligned to the *T. urticae* reference genome (Grbić et al., 2011b) using the two-pass alignment mode of STAR 2.5.0b (Dobin et al., 2013) with a maximum intron size of 20 kb (the results were splice-aware alignments made independently of the reference genome annotation). The resulting BAM files were subsequently sorted by read name using Samtools 1.2 (Li et al., 2009). Read counts per gene, based on the reference annotation (version October 29, 2015), were then obtained using the default settings of HTSeq 0.6.0 (Anders et al., 2014) with the ‘STRANDED’ flag set to ‘yes’ and the ‘FEATURE’ flag set to ‘exon’. Differentially expressed genes between the proterosoma of *T. urticae* females and intact *T. urticae* females were determined using the DESeq2 (version 1.6.3; Anders et al., 2013) and Bioconductor (http://bioconductor.org/) R-packages. The ‘unfiltered DESeq2 results’ settings \[dds <- DESeq(dds, minReplicatesForReplace=Inf) and res <- results(dds, cooksCutoff=FALSE, independent Filtering=FALSE)\] were used for differential expression analysis. Genes with a fold change (FC) ≥8 and a Benjamini-Hochberg adjusted \(p\) ≤0.05 were considered differentially expressed (DE). All gene expression data have been uploaded to the Gene Expression Omnibus with accession number GSE81128.

3.2.9. Validation of salivary proteins by whole-mount *in situ* hybridization

*In situ* hybridization in *T. urticae* was based on Dearden and Akam (2000) and Dearden et al. (2002). Briefly, RNA was extracted from *T. urticae* mites
(London strain) feeding on bean or tomato plants (Total RNA Isolation Mini Kit, Agilent), treated with TURBO DNA-free™ Kit (Ambion) to remove contaminating genomic DNA and used for cDNA synthesis using Maxima First Strand cDNA Synthesis Kit. Primers were designed using Primer3 (http://bioinfo.ut.ee/primer3/). A fragment with preferred length of about 300 bp was amplified (primers used are in Table S3.5). PCR products were cloned into pGEM-T plasmids (Promega) and transformed into E. coli. Plasmids from liquid cultures were purified after which insert orientation and nucleotide sequence were determined by sequencing (LGC Genomics, Germany). A PCR was performed on the plasmids using pUC/M13 primers (Table S3.5). PCR product, containing insert flanked by T7 and SP6 promoter sites from the plasmid backbone, was checked by agarose gel electrophoresis and purified using E.Z.N.A. Cycle Pure Kit (Omega Biotek, GA, USA). Depending on orientation, sense or anti-sense digoxigenin-labeled (DIG-labeled) probes were generated using T7 or SP6 RNA polymerase (Roche), using the pUC/M13 PCR product and DIG-UTPs (Roche) in the in vitro labeling reaction. Probes were then purified using SigmaSpin™ Sequencing Reaction Clean-Up Columns (Sigma), supplemented with hybridization buffer (50% formamide (Sigma), 4x SSC (Sigma), 1x Denhardt’s solution (Sigma), 250 μg/ml tRNA (wheat germ type V, Sigma), 250 μg/ml ssDNA (boiled salmon sperm DNA, Sigma), 50 μg/ml heparin (sodium salt, Sigma), 0.1% Tween-20 (Sigma), 5% dextran sulfate (sodium salt, Sigma) and stored at -20°C until used.

*Proteomic composition of T. urticae saliva*

*Tetranychus urticae* nymphs and adults (London strain) of both sexes were collected from bean and tomato plants and fixed overnight in a 1:1 mixture of heptane and PTw (PBS with 0.1% Tween-20) containing 4% formaldehyde. The mites were then washed in methanol and gradually rehydrated in PTw, followed by sonication in a sonic cleansing bath and treatment with 0.2 mg/ml Proteinase K during 10 min. The mites were then re-fixed with 4% formaldehyde in PTw. After washing in PTw, mites were prehybridized in hybridization buffer for 1 h at 52°C. Hybridization buffer (300 μl) was refreshed and probe (4 μl) was added. The mites were then incubated overnight at 52°C. Washing occurred at 53°C with wash buffer composed of 50% formamide, 2x SSC and 0.1% tween-20. After washing at room temperature with PBTw (PTw with 0.1 % BSA, Sigma), the mites were incubated at RT for 2 h with a 1:1000 dilution of anti-digoxigenin-AP (Fab fragments, Roche) in PBTw. The mites were then washed
with PTw and alkaline phosphatase (AP) buffer (100 mM Tris pH 9.5, 100 mM NaCl, 1 M MgCl₂, 0.1% Tween-20). AP buffer containing the FastRed substrate (SIGMAFAST™ Fast Red TR/Naphthol AS-MX tablets, Sigma) was added and mites were incubated at RT in the dark, until red staining was visible. Methanol was used to reduce background staining and the mites were eventually cleared in 70% glycerol in PTw (pH 8.5) after washing with pure PTw. The mites were then mounted on a microscopy glass for further microscopic investigation (Nikon A1R fluorescence confocal microscope; emission at 500-530 nm and acquisition at 488 nm for spider mite auto-fluorescence and emission at 570-620 nm and acquisition at 561.7 nm for FastRed signal). Z-stacks were created using 15 slices with 2-3 μm distance between slices. All images were processed with Fiji (Schindelin et al., 2012).

3.2.10. *Tetranychus urticae* protein family analyses

*Tetranychus urticae* proteins were assigned to OrthoMCL groups using the online OrthoMCL software tool (http://www.orthomcl.org/orthomcl/proteomeUpload.do) and the *T. urticae* proteome (version July 29, 2015, available via the ORCAE database) as query (Chen et al., 2006). Signal peptides were predicted with SignalP 4.1 (Petersen et al., 2011) and protein subcellular localization was predicted using WoLF PSORT (organism type: ‘Animal’) (Horton et al., 2007) at http://www.genscript.com/wolf-psort.html. Proteins with a sequence length less than 30 AA or not containing a start methionine (pseudogenes) were excluded from WoLF PSORT analysis. Transmembrane domains were predicted using TMHMM server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). A phylogenetic analysis was performed for proteins belonging to OrthoMCL cluster Tu_MCL_35 and Tu_MCL_36. Except for tetur55g00110 (Tu_MCL_36) which is encoded by a pseudogene, proteins from each cluster were aligned using MUSCLE (Edgar, 2004). Model selection was done with ProtTest 2.4 (Abascal et al., 2005) and according to the Akaike information criterion WAG+G and WAG+G+F were optimal for the phylogenetic reconstruction of Tu_MCL_35 and Tu_MCL_36 proteins, respectively. Finally, for each alignment a maximum likelihood analysis was performed using Treefinder (version 2011) (Jobb et al., 2004) bootstrapping with 1000 pseudoreplicates (LR-ELW). The resulting trees were midpoint rooted and edited with MEGA 6.0 software (Tamura et al., 2013).
**Proteomic composition of *T. urticae* saliva**

### 3.2.11. Evaluation of *T. urticae* salivary proteins using available databases

*T. urticae* salivary proteins identified by nano-LC-MS/MS were used as query in a BLASTp search (E-value threshold 1.0 E-5) against the proteome of the non-phytophagous American house dust mite, *Dermatophagoides farinae* (Chan et al., 2015). Furthermore, expression levels of genes coding for putative *T. urticae* salivary proteins were compared between feeding (mobile) *T. urticae* stages (larva, nymph and adult) and the embryo stage (Grbić et al., 2011b), while a nano-LC–nano-ESI-QTOF MS/MS proteomic analysis of mite faeces (Santamaría et al., 2015) was screened for the presence of putative salivary proteins of *T. urticae*. A set of differentially expressed genes in diapausing *T. urticae* females (Bryon et al., 2013) was also investigated for genes coding for putative *T. urticae* salivary proteins. Finally, an expression dataset of *T. urticae* genes across different time points of host plant transfer to tomato (Wybouw et al., 2015) was mined for putative *T. urticae* salivary genes.

### 3.3. Results

#### 3.3.1. *Tetranychus urticae* secretes proteins in artificial diet which can reliably be identified through mass spectrometry analysis

Spider mite lines propagated for at least 5 generations on different host plants (i.e., bean, maize, soy and tomato) were allowed to feed on an artificial diet encapsulated in parafilm (Figure 3.1). Mites fed on the artificial diet within hours, as assessed by blue staining of control mites being fed an erioglaucine-supplemented diet. After 24 h, the artificial diet was collected for each plant-adapted mite line and for reference samples (artificial diet without feeding mites) and the protein content was analyzed using nano-LC-MS/MS. The detected salivary proteins from all host plant-adapted mites were pooled into one dataset. Ninety-five proteins (some belonging to 12 ‘protein inference groups’ containing proteins identified by shared peptides) had a mean PSM of at least 2 and were retained as putative *T. urticae* salivary proteins (Tables 3.1 and S3.6). Subsequently we used the normalized (relative) abundance factor rTop3, based on MS1 intensity, for abundance ranking of these putative salivary proteins. This rTop3 value has been shown to correlate with the mole fraction of the protein of interest (Grossmann et al., 2010; Krey et al., 2014). Proteins with a high rTop3 fac-
tor are therefore assumed to be more prominent in *T. urticae* saliva. The majority (81%) of the putative *T. urticae* salivary proteins had a SP for secretion as predicted by SignalP. Only four (tetur03g08030, tetur10g00090, tetur10g00100, tetur22g00260) out of the 95 putative salivary proteins were predicted to have a transmembrane domain, while 76 (80%) were predicted to have an extracellular localization (Table 3.1). The OrthoMCL analysis grouped 13,558 *T. urticae* proteins into 6397 ortholog groups. The majority of these groups already existed in the OrthoMCL database (http://www.orthomcl.org/orthomcl/), while 401 ortholog groups (group names starting with ‘Tu_MCL’) were specific for *T. urticae* (Table S3.7). From the 95 putative *T. urticae* salivary proteins detected by nano-LC-MS/MS, 63 proteins could be assigned to an existing OrthoMCL group, 22 proteins were grouped into a *T. urticae* specific OrthoMCL group and 10 proteins could not be included into any group (Tables 3.1 and S3.6).

**FIGURE 3.1.** Feeding arena used to collect spider mite saliva. (A) Schematic drawing of the feeding arena. (B) Left panel: overview of the feeding arena; middle panel: top view of the diet hemisphere infested with spider mites; right panel: adult female spider mite feeding on a control hemisphere with a blue colorant added to the diet (note the blue color visible in the gut of the spider mite).
Table 3.1. List of putative *Tetranychus urticae* salivary proteins identified in artificial diet using nano-LC-MS/MS

Proteins are ranked based on normalized abundance factor (rTop3). See Table S3.6 for a detailed overview.

<table>
<thead>
<tr>
<th><em>T. urticae</em> protein ID</th>
<th>Functional description</th>
<th>Max. rTop3&lt;sup&gt;b&lt;/sup&gt;</th>
<th>SP&lt;sup&gt;c&lt;/sup&gt;</th>
<th>WP&lt;sup&gt;d&lt;/sup&gt;</th>
<th>ISHe</th>
<th>Dff</th>
<th>OrthoMCL (#)&lt;sup&gt;g&lt;/sup&gt;</th>
<th>Feeding Proterosoma</th>
<th>Proterosoma Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>tetur01g01850</td>
<td>Projectin</td>
<td>0.1484</td>
<td>N</td>
<td>C</td>
<td>*</td>
<td>OG5_126738 (10)</td>
<td>0.4</td>
<td>2.9*</td>
<td></td>
</tr>
<tr>
<td>tetur20g00560</td>
<td>Hypothetical protein</td>
<td>0.1137</td>
<td>Y</td>
<td>E</td>
<td>A</td>
<td>Tu_MCL_12 (35)</td>
<td>6.0</td>
<td>33.9*</td>
<td></td>
</tr>
<tr>
<td>tetur07g00150</td>
<td>Serine protease</td>
<td>0.0877</td>
<td>Y</td>
<td>E</td>
<td>A</td>
<td>OG5_135950 (7)</td>
<td>8.0</td>
<td>28.6*</td>
<td></td>
</tr>
<tr>
<td>tetur28g01330</td>
<td>Wannes-Thomas Secreted</td>
<td>0.0855</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>Tu_MCL_36 (13)</td>
<td>3.7</td>
<td>37.1*</td>
<td></td>
</tr>
<tr>
<td>tetur55g00040</td>
<td>Protein (WTSP) 1</td>
<td>0.0817</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>OG5_144177 (6)</td>
<td>30.4</td>
<td>30.9*</td>
<td></td>
</tr>
<tr>
<td>tetur03g03700</td>
<td>Hypothetical protein/PE-PGRS family protein</td>
<td>0.0816</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_168371 (5)</td>
<td>emb(0)</td>
<td>10.4*</td>
<td></td>
</tr>
<tr>
<td>tetur12g03940</td>
<td>Serine protease</td>
<td>0.0738</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>OG5_144177 (6)</td>
<td>17.9</td>
<td>38.9*</td>
<td></td>
</tr>
<tr>
<td>tetur03g03680</td>
<td>Hypothetical protein/PE-PGRS family protein</td>
<td>0.0691</td>
<td>Y</td>
<td>E</td>
<td>A</td>
<td>Tu_MCL_35 (14)</td>
<td>2.3</td>
<td>10*</td>
<td></td>
</tr>
<tr>
<td>tetur32g00050</td>
<td>Small Secreted Protein, Family F</td>
<td>0.0628</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_144177 (6)</td>
<td>15.3</td>
<td>40*</td>
<td></td>
</tr>
<tr>
<td>tetur03g03670,</td>
<td>Hypothetical protein/PE-PGRS family protein</td>
<td>0.0530</td>
<td>Y</td>
<td>E</td>
<td>A</td>
<td>Tu_MCL_45 (11)</td>
<td>emb(0)</td>
<td>( )</td>
<td></td>
</tr>
<tr>
<td>tetur06g02580</td>
<td>Cathepsin B</td>
<td>0.0469</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>OG5_127800 (28)</td>
<td>1.5</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td>tetur05g04580</td>
<td>Wannes-Thomas Secreted Protein (WTSP)</td>
<td>0.0429</td>
<td>Y</td>
<td>E</td>
<td>D</td>
<td>Tu_MCL_36 (13)</td>
<td>11.4</td>
<td>9.5*</td>
<td></td>
</tr>
<tr>
<td>tetur16g03470</td>
<td>Serine protease homologue</td>
<td>0.0387</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_152237 (2)</td>
<td>3.7</td>
<td>12.6*</td>
<td></td>
</tr>
<tr>
<td>tetur06g03320</td>
<td>Wannes-Thomas Secreted Protein (WTSP)</td>
<td>0.0328</td>
<td>Y</td>
<td>E</td>
<td>D</td>
<td>Tu_MCL_36 (13)</td>
<td>(-)</td>
<td>-3.3</td>
<td></td>
</tr>
<tr>
<td>tetur16g03420</td>
<td>Beta-mannosidase</td>
<td>0.0313</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>OG5_129300 (4)</td>
<td>2.5</td>
<td>18.4*</td>
<td></td>
</tr>
<tr>
<td>tetur29g01360*</td>
<td>Hypothetical protein</td>
<td>0.0233</td>
<td>Y</td>
<td>E</td>
<td>D</td>
<td>Tu_MCL_74 (6)</td>
<td>9.4</td>
<td>36.1*</td>
<td></td>
</tr>
<tr>
<td>tetur03g10083</td>
<td>Hypothetical protein/PE-PGRS</td>
<td>0.0219</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>Tu_MCL_74 (6)</td>
<td>9.4</td>
<td>36.1*</td>
<td></td>
</tr>
<tr>
<td>T. urticae protein IDa</td>
<td>Functional description</td>
<td>Max. SPc</td>
<td>WPd</td>
<td>ISHe</td>
<td>Dff</td>
<td>OrthoMCL (#)g</td>
<td>Feeding Proterosoma Fech</td>
<td>Proterosoma FCi</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------</td>
<td>---------</td>
<td>-----</td>
<td>------</td>
<td>-----</td>
<td>--------------</td>
<td>--------------------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>tetur01g02670</td>
<td>MD-2-related lipid-recognition domain protein</td>
<td>0.0217</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>* OG5_158831 (2)</td>
<td>4.6 -2*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur04g05980*</td>
<td>Apolipoprotein D precursor</td>
<td>0.0215</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>* OG5_130527 (50)</td>
<td>6.4 -1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur05g04560</td>
<td>Wannes-Thomas Secreted Protein (WTSP) 8</td>
<td>0.0209</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>Tu_MCL_36 (13)</td>
<td>10.3 10.5*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur07g00160*</td>
<td>Hypothetical protein</td>
<td>0.0200</td>
<td>Y</td>
<td>E</td>
<td>D</td>
<td></td>
<td>4.4 14.6*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur03g03680, family protein</td>
<td>Hypothetical protein/PE-PGRS</td>
<td>0.0196</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>OG5_144177, Tu_MCL_74 (6)</td>
<td>17.9 38.9*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur03g03700, family protein</td>
<td>Hypothetical protein</td>
<td>0.0196</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>OG5_144177, Tu_MCL_74 (6)</td>
<td>17.9 38.9*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur07g00392, tetur95g00080</td>
<td>Trypsin Inhibitor-like, cysteine rich domain protein</td>
<td>0.0167</td>
<td>Y</td>
<td>E</td>
<td>D</td>
<td>OG5_207753 (2)</td>
<td>3.3 45.3*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur28g00360</td>
<td>Beta-mannosidase</td>
<td>0.0167</td>
<td>Y</td>
<td>E</td>
<td>A</td>
<td>* OG5_129300 (4)</td>
<td>5.5 44*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur28g00510</td>
<td>Hypothetical protein</td>
<td>0.0162</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>* OG5_133467 (14)</td>
<td>9.1 -2.7*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur55g00040</td>
<td>Wannes-Thomas Secreted Protein (WTSP) 1b</td>
<td>0.0131</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>Tu_MCL_36 (13)</td>
<td>1.8 16.4*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur14g02090</td>
<td>Immunoglobulin E-set</td>
<td>0.0128</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>* OG5_133467 (14)</td>
<td>9.1 -2.7*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur07g01660</td>
<td>Hypothetical protein</td>
<td>0.0122</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>Tu_MCL_36 (13)</td>
<td>3.7 37.1*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur515g00030</td>
<td>Wannes-Thomas Secreted Protein (WTSP) 1a</td>
<td>0.0119</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>Tu_MCL_25 (17)</td>
<td>4.7 27.3*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur31g00630</td>
<td>Hypothetical protein with Armadillo-type fold</td>
<td>0.0118</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>Tu_MCL_25 (17)</td>
<td>4.7 27.3*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur03g08710</td>
<td>Hypothetical protein</td>
<td>0.0111</td>
<td>Y</td>
<td>E</td>
<td>D</td>
<td></td>
<td>3.0 36*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur14g02080</td>
<td>MD-2-related lipid-recognition domain protein</td>
<td>0.0108</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>* OG5_133467 (14)</td>
<td>8.9 -22.6*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur02g04310</td>
<td>Glycoside hydrolase, catalytic core</td>
<td>0.0100</td>
<td>Y</td>
<td>E</td>
<td></td>
<td>* OG5_127620 (2)</td>
<td>7.9 -1.8*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur516g00020</td>
<td>Vitellogenin 1, partial</td>
<td>0.0098</td>
<td>N</td>
<td>C</td>
<td></td>
<td>* OG5_132251 (7)</td>
<td>336.7 -3.9*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T. urticae protein ID</td>
<td>Functional description</td>
<td>Max. SP&lt;sup&gt;c&lt;/sup&gt;</td>
<td>WP&lt;sup&gt;d&lt;/sup&gt;</td>
<td>ISHe</td>
<td>Df&lt;sup&gt;f&lt;/sup&gt;</td>
<td>OrthoMCL (#)&lt;sup&gt;g&lt;/sup&gt;</td>
<td>Feeding FC&lt;sup&gt;h&lt;/sup&gt;</td>
<td>Proterosoma FC&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------</td>
<td>----------------------</td>
<td>-------------</td>
<td>------</td>
<td>----------</td>
<td>--------------------------</td>
<td>----------------------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td>tetur01g05230, tetur01g16473, tetur06g02570, tetur06g02580</td>
<td>Cathepsin B</td>
<td>0.0097</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_127800 (28)</td>
<td>6.5</td>
<td>-6.4*</td>
<td></td>
</tr>
<tr>
<td>tetur30g01440</td>
<td>Serine protease</td>
<td>0.0096</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_152337 (1)</td>
<td>5.1</td>
<td>46.4*</td>
<td></td>
</tr>
<tr>
<td>tetur32g01880</td>
<td>Hypothetical protein</td>
<td>0.0094</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>Tu_MCL_43 (12)</td>
<td>2.6</td>
<td>62.3*</td>
<td></td>
</tr>
<tr>
<td>tetur28g01720</td>
<td>Short-chain dehydrogenase/reductase SDR</td>
<td>0.0091</td>
<td>N</td>
<td>C</td>
<td>*</td>
<td>OG5_128170 (22)</td>
<td>4.9</td>
<td>-46.1*</td>
<td></td>
</tr>
<tr>
<td>tetur14g02070</td>
<td>MD-2-related lipid-recognition domain protein</td>
<td>0.0091</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_133467 (14)</td>
<td>7.8</td>
<td>-16.7*</td>
<td></td>
</tr>
<tr>
<td>tetur10g00090+</td>
<td>Glycine-rich secreted protein</td>
<td>0.0083</td>
<td>Y</td>
<td>E</td>
<td>D</td>
<td></td>
<td>5.0</td>
<td>31.5*</td>
<td></td>
</tr>
<tr>
<td>tetur09g04400*</td>
<td>Cathepsin L</td>
<td>0.0083</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_126607 (22)</td>
<td>8.0</td>
<td>-1.1</td>
<td></td>
</tr>
<tr>
<td>tetur14g00010*</td>
<td>Vitellinogen 5</td>
<td>0.0082</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_132251 (7)</td>
<td>53.6</td>
<td>-7.5*</td>
<td></td>
</tr>
<tr>
<td>tetur10g01010</td>
<td>Orphan Secreted protein</td>
<td>0.0078</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>Tu_MCL_212 (2)</td>
<td>8.9</td>
<td>21.9*</td>
<td></td>
</tr>
<tr>
<td>tetur07g07380*</td>
<td>Beta-galactosidase</td>
<td>0.0074</td>
<td>Y</td>
<td>M</td>
<td>*</td>
<td>OG5_128163 (4)</td>
<td>5.5</td>
<td>-2.3*</td>
<td></td>
</tr>
<tr>
<td>tetur10g00000+</td>
<td>Hypothetical protein</td>
<td>0.0072</td>
<td>Y</td>
<td>E</td>
<td>4.2</td>
<td></td>
<td>24.4*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tetur04g09479</td>
<td>Small Secreted Protein, Family F</td>
<td>0.0068</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>Tu_MCL_35 (14)</td>
<td>5.9</td>
<td>65.8*</td>
<td></td>
</tr>
<tr>
<td>tetur01g00950</td>
<td>Orphan secreted protein</td>
<td>0.0068</td>
<td>Y</td>
<td>E</td>
<td>A</td>
<td>Tu_MCL_211 (2)</td>
<td>7.5</td>
<td>28.6*</td>
<td></td>
</tr>
<tr>
<td>tetur06g00230</td>
<td>Hypothetical protein</td>
<td>0.0061</td>
<td>Y</td>
<td>E</td>
<td>A</td>
<td>Tu_MCL_25 (17)</td>
<td>5.1</td>
<td>29.3*</td>
<td></td>
</tr>
<tr>
<td>tetur12g01860*</td>
<td>Cathepsin L</td>
<td>0.0057</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_126607 (22)</td>
<td>8.2</td>
<td>-1.5*</td>
<td></td>
</tr>
<tr>
<td>tetur31g00830</td>
<td>Hypothetical protein</td>
<td>0.0056</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>Tu_MCL_25 (17)</td>
<td>emb(0)</td>
<td>14.4</td>
<td></td>
</tr>
<tr>
<td>tetur03g07920*</td>
<td>Glutathione S-transferase, delta class</td>
<td>0.0053</td>
<td>N</td>
<td>C</td>
<td>*</td>
<td>OG5_126942 (17)</td>
<td>5.7</td>
<td>-2.2*</td>
<td></td>
</tr>
<tr>
<td>tetur31g01040, tetur31g01060</td>
<td>Hypothetical protein with Armadillo-type fold</td>
<td>0.0051</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>Tu_MCL_25 (17)</td>
<td>8.6</td>
<td>47.7*</td>
<td></td>
</tr>
<tr>
<td>tetur03g08030+</td>
<td>Cathepsin B</td>
<td>0.0051</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_127800 (28)</td>
<td>13.3</td>
<td>-15.8*</td>
<td></td>
</tr>
<tr>
<td>tetur16g03190</td>
<td>Serine protease homologue</td>
<td>0.0050</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_141111 (2)</td>
<td>5.3</td>
<td>48.7*</td>
<td></td>
</tr>
<tr>
<td>T. urticae protein ID</td>
<td>Functional description</td>
<td>Max. rTop3</td>
<td>SP</td>
<td>WP</td>
<td>ISHe</td>
<td>Dff</td>
<td>OrthoMCL (#)</td>
<td>Feeding Proterosoma FC</td>
<td>Proterosoma FC</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------</td>
<td>-------------</td>
<td>----</td>
<td>----</td>
<td>------</td>
<td>-----</td>
<td>--------------</td>
<td>------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>tetur01g02510*</td>
<td>Glutathione S-transferase, class delta</td>
<td>0.0048</td>
<td>N</td>
<td>C</td>
<td>*</td>
<td></td>
<td>OG5_126942(17)</td>
<td>1.2</td>
<td>10.1*</td>
</tr>
<tr>
<td>tetur39g00810*</td>
<td>Vitellogenin 2</td>
<td>0.0045</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td></td>
<td>OG5_132251(7)</td>
<td>574.8</td>
<td>1.3*</td>
</tr>
<tr>
<td>tetur23g01640</td>
<td>Flavin reductase</td>
<td>0.0043</td>
<td>N</td>
<td>C</td>
<td></td>
<td></td>
<td>OG5_131746(2)</td>
<td>2.7</td>
<td>1.3</td>
</tr>
<tr>
<td>tetur26g02320</td>
<td>Superoxide dismutase, copper/zinc binding</td>
<td>0.0043</td>
<td>N</td>
<td>C</td>
<td>*</td>
<td></td>
<td>OG5_127584(8)</td>
<td>1.8</td>
<td>3.2*</td>
</tr>
<tr>
<td>tetur19g01900</td>
<td>Prolylcarboxypeptidase</td>
<td>0.0043</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td></td>
<td>OG5_130246(1)</td>
<td>3.5</td>
<td>-1.8*</td>
</tr>
<tr>
<td>tetur18g03030*</td>
<td>Alpha-2-macroglobulin</td>
<td>0.0042</td>
<td>N</td>
<td>C</td>
<td>*</td>
<td></td>
<td>OG5_126560(8)</td>
<td>9.3</td>
<td>-12.2*</td>
</tr>
<tr>
<td>tetur07g08034</td>
<td>Serine proteases, trypsin family</td>
<td>0.0042</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td></td>
<td>OG5_149533(13)</td>
<td>1.0</td>
<td>4</td>
</tr>
<tr>
<td>tetur09g03880</td>
<td>Serine protease</td>
<td>0.0040</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td></td>
<td>OG5_168371(5)</td>
<td>6.7</td>
<td>48.3*</td>
</tr>
<tr>
<td>tetur06g01060</td>
<td>Cystatin</td>
<td>0.0039</td>
<td>N</td>
<td>E</td>
<td>*</td>
<td></td>
<td>OG5_147492(6)</td>
<td>10.1</td>
<td>-2.9*</td>
</tr>
<tr>
<td>tetur14g03160</td>
<td>Hypothetical protein</td>
<td>0.0039</td>
<td>N</td>
<td>E</td>
<td></td>
<td></td>
<td>Tu_MCL_63(8)</td>
<td>emb(0)</td>
<td>3.1</td>
</tr>
<tr>
<td>tetur06g01640</td>
<td>Outer membrane lipoprotein Blc/Lipocalin</td>
<td>0.0031</td>
<td>Y</td>
<td>E</td>
<td></td>
<td></td>
<td>OG5_130527(50)</td>
<td>3.8</td>
<td>-2.3*</td>
</tr>
<tr>
<td>tetur20g01290</td>
<td>Hypothetical protein</td>
<td>0.0030</td>
<td>Y</td>
<td>E</td>
<td></td>
<td></td>
<td>OG5_126560(8)</td>
<td>9.3</td>
<td>-12.2*</td>
</tr>
<tr>
<td>tetur13g03820</td>
<td>Proteinase inhibitor I2, Kunitz metazoa</td>
<td>0.0030</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td></td>
<td>OG5_134456(1)</td>
<td>4.1</td>
<td>2.5*</td>
</tr>
<tr>
<td>tetur25g00650*</td>
<td>Cathepsin L</td>
<td>0.0030</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td></td>
<td>OG5_126607(22)</td>
<td>16.4</td>
<td>-1.4</td>
</tr>
<tr>
<td>tetur13g00600</td>
<td>Hypothetical protein</td>
<td>0.0029</td>
<td>Y</td>
<td>E</td>
<td></td>
<td></td>
<td>OG5_129423(34)</td>
<td>17.5</td>
<td>-1</td>
</tr>
<tr>
<td>tetur147g00020, tetur33g01260</td>
<td>Glycoside hydrolase, subgroup, catalytic core</td>
<td>0.0029</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td></td>
<td>OG5_129423(34)</td>
<td>17.5</td>
<td>-1</td>
</tr>
<tr>
<td>tetur01g00940</td>
<td>Orphan secreted protein</td>
<td>0.0029</td>
<td>Y</td>
<td>E</td>
<td>A</td>
<td></td>
<td>Tu_MCL_212(2)</td>
<td>7.0</td>
<td>23.4*</td>
</tr>
<tr>
<td>tetur09g03620, tetur09g03650, tetur09g03670</td>
<td>Cystatin</td>
<td>0.0028</td>
<td>Y</td>
<td>E</td>
<td></td>
<td></td>
<td>OG5_147492(6)</td>
<td>7.0</td>
<td>-1.8*</td>
</tr>
<tr>
<td>tetur22g00260*</td>
<td>Trypsin Inhibitor-like, cysteine rich domain protein</td>
<td>0.0027</td>
<td>Y</td>
<td>E</td>
<td></td>
<td></td>
<td>OG5_176862(1)</td>
<td>11.3</td>
<td>2.6*</td>
</tr>
<tr>
<td>Protein ID</td>
<td>Functional description</td>
<td>Max. $r$Top3$^b$</td>
<td>SP$^c$</td>
<td>WP$^d$</td>
<td>ISH$^f$</td>
<td>OrthoMCL (#)$^g$</td>
<td>Feeding</td>
<td>Proterosoma</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------------------</td>
<td>-----------------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
<td>-----------------</td>
<td>---------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>tetu02g12930</td>
<td>Glycoside hydrolase, subgroup, catalytic core</td>
<td>0.0027</td>
<td>Y</td>
<td>E.R.</td>
<td>*</td>
<td>OG5_129423 (34)</td>
<td>8.2</td>
<td>-3.9*</td>
<td></td>
</tr>
<tr>
<td>tetu02g11340</td>
<td>Aldo-keto reductase family B, member 1</td>
<td>0.0026</td>
<td>N</td>
<td>C</td>
<td>*</td>
<td>OG5_126583 (9)</td>
<td>5.0</td>
<td>8.1*</td>
<td></td>
</tr>
<tr>
<td>tetu02g11390</td>
<td>Short-chain dehydrogenase/reductase</td>
<td>0.0022</td>
<td>N</td>
<td>C</td>
<td>*</td>
<td>OG5_128075 (3)</td>
<td>emb(0)</td>
<td>2.7*</td>
<td></td>
</tr>
<tr>
<td>tetu01g11910</td>
<td>Chitinase</td>
<td>0.0022</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_152454 (3)</td>
<td>0.7</td>
<td>28.7*</td>
<td></td>
</tr>
<tr>
<td>tetu09g00900</td>
<td>Hypothetical protein</td>
<td>0.0022</td>
<td>Y</td>
<td>E</td>
<td>2</td>
<td>OG5_127143 (3)</td>
<td>2.0</td>
<td>2.8*</td>
<td></td>
</tr>
<tr>
<td>tetu10g00010</td>
<td>Fructose-bisphosphate aldolase, class-I</td>
<td>0.0018</td>
<td>N</td>
<td>C</td>
<td>*</td>
<td>OG5_136350 (8)</td>
<td>1.6</td>
<td>-1.1</td>
<td></td>
</tr>
<tr>
<td>tetu03g04470</td>
<td>Fibronectin, type III-like fold</td>
<td>0.0012</td>
<td>Y</td>
<td>E</td>
<td>*</td>
<td>OG5_136350 (3)</td>
<td>6.7</td>
<td>1.3*</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ accession number in the ORCAE genome database (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur); an asterisk (*) indicates that the *T. urticae* protein has been found in faeces of mites (Santamaría et al., 2015) while a cross (†) indicates that the *T. urticae* protein was predicted with a transmembrane domain by TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/)

$^b$ maximal $r$Top3 value across the four *T. urticae* host plant adapted lines (bean, soy, maize, tomato)

$^c$ presence (Y= yes, N= no) of a signal peptide as predicted by SignalP 4.0 (Petersen et al., 2011)

$^d$ subcellular localization prediction by WoLF PSORT (CytoSkeleot/Cytoplasm/Endoplasmic Reticulum/Extracellular/Mitochondrial) (Horton et al., 2007)

$^e$ gland expression localization determined by whole-mount *in situ* hybridization (ISH): Anterior or Dorsal podocephalic glands

$^f$ homologues (BLASTp E-value threshold 1E-5) identified in the American house dust mite *D. farinae* (*Df*)

$^g$ OrthoMCL group with the number of OrthoMCL members present in the *T. urticae* genome shown between brackets (see Table S3.7 for OrthoMCL details)
overall average expression across all the feeding stages (larva, nymph, adult) relative to expression in the embryo stage (Grbić et al., 2011a); in case of gene groups the expression ratio of the first gene is shown; emb(0): no reads could be mapped to this gene in the embryo stage; (-) T. urticae gene was not annotated at time of experiment.

Fold change in gene expression between T. urticae proterosomas and intact T. urticae females (* = Benjamini-Hochberg corrected $p \leq 0.05$); in case of gene groups the FC of the first gene is shown; (-) no reads could be mapped to this gene.
3.3.2. RNAseq analysis of *T. urticae* proterosomas supports the salivary origin of many identified proteins

RNA was extracted from 100 intact adult female *T. urticae* mites with three-fold biological replication and from 250 dissected *T. urticae* proterosomas. For all *T. urticae* genes the number of mapped RNAseq reads can be found in Table S3.8. Using the RNAseq data and the DESeq2 software, we performed a differential expression analysis (Benjamini-Hochberg adjusted $p \leq 0.05$ and $|FC| \geq 8$) (Figure S3.2, Table S3.9) between dissected *T. urticae* proterosomas and intact females: 1800 *T. urticae* genes showed a significantly higher expression in the proterosoma compared to intact females, while 1268 had a significantly lower expression. About 20% (591 genes) of the genes with a significantly different expression in the proterosoma coded for extracellular proteins, with 455 out of 591 genes having a significantly higher expression and 136 having a significantly lower expression in the proterosoma (Figure 3.2, Table S3.9). Fifty-three percent of the top 100 (ranked based on decreasing log$_2$FC values) of the genes coding for extracellular proteins with significantly higher expression in the proterosoma encoded for proteins with unknown function, while genes coding for serine proteases (eight genes), neuropeptides (six genes) or homologues of ‘salivary gland peptides’ of ticks (five genes) were also observed (Table S3.9).

About half of the genes (54%) coding for *T. urticae* putative salivary proteins detected by nano-LC-MS/MS had significantly different expression levels (Benjamini-Hochberg adjusted $p \leq 0.05$ and $|FC| \geq 8$) in the proterosoma (Figures 3.2 and S3.2, Table S3.9). The majority (88%) of these genes showed a higher expression in the proterosoma while only 6 had a lower expression. Among those genes with a higher expression, we identified actin (tetur03g09480), an aldo-keto reductase (tetur02g11340), a glutathione S-transferase (tetur01g02510), a chitinase (tetur01g11910), a protein with a trypsin inhibitor-like, cysteine rich domain (tetur40g00392), 2 beta-mannosidases (tetur16g03420, tetur28g00360), 7 serine-proteases (tetur07g00150, tetur09g03880, tetur12g03940, tetur12g03950, tetur16g03190, tetur16g03470, tetur30g01440) and 31 genes coding for proteins with a yet unknown function. Among the 6 genes with a lower expression we found genes coding for a cathepsin (tetur03g08030), a short-chain dehydrogenase (tetur28g01720), an alpha-2-macroglobulin (tetur18g03030), conserved secreted proteins with MD-2-related lipid recognition domain (tetur14g02070 and tetur14g02080), and a gene coding...
for a protein with an unknown function (tetur20g01290) (Tables 3.1 and S3.9, Figures 3.2 and S3.2).

3.3.3. Evaluation of the salivary protein repertoire with previously published datasets

Fifty of the 95 proteins (53%) identified by nano-LC-MS/MS shared homology (BLASTp with E-value threshold 1.0 E-5) with proteins of the most closely related non-phytophagous acariform mite for which proteom-

![Figure 3.2. Tetranychus urticae genes that code for extracellular proteins and are differentially expressed between proterosomas and intact females. (A) Dissected T. urticae proterosomas used for RNA extraction (see Figure 3.3 for dissection position). (B) The negative log10 of Benjamini-Hochberg adjusted p-values were plotted against the log2 FC in gene expression for all T. urticae genes (see Figure S3.2) and subsequently filtered for genes coding for extracellular proteins (predicted to be extracellular by WoLF PSORT (Horton et al., 2007) and predicted with a SP by SignalP 4.1 (Petersen et al., 2011), see Table S3.9). Differentially expressed genes (|FC| ≥8, Benjamini-Hochberg adjusted p ≤0.05) are shown as grey dots (DE), while genes that were not differentially expressed are shown as grey circles (not DE). Genes coding for extracellular proteins that were identified by nano-LC-MS/MS (‘extracellular LC-MS salivary protein genes’) are depicted as stars. Extracellular LC-MS salivary protein genes that were differentially expressed are depicted as green stars (DE/LC-MS). Extracellular LC-MS salivary protein genes that were differentially expressed and for which ISH confirmed expression in salivary glands are depicted as red stars (DE/LC-MS/ISH) while extracellular LC-MS salivary protein genes that are not differentially expressed are shown as blue stars (not DE/LC-MS).]
Proteomic composition of *T. urticae* saliva

ic resources were available, the American house dust mite *D. farinae* (TABLES 3.1 and S3.6). These were mainly proteins with a predicted function in digestion (26 out of 50), like glycoside hydrolases, beta-galactosidase/mannosidase, propylcarboxypeptidase, proteinase inhibitors and serine proteases. *Tetranychus urticae* proteins with an unknown function did not share homology with *D. farinae* proteins. A previously published dataset of the *T. urticae* transcriptome across different developmental stages (Grbić et al., 2011b) was mined for the gene expression-levels of *T. urticae* putative salivary genes. The overall average expression for all *T. urticae* genes (calculated as ‘reads per kb of transcript per million mapped reads’: RPKM) across all the feeding stages (larva, nymph, adult) relative to the overall average expression in the embryo stage was 2.21 fold higher (Grbić et al., 2011b). Seventy-five out of 95 *T. urticae* putative salivary genes had a ratio higher than this average, while 28 (30%) have an expression ratio more than 8 (TABLE 3.1). Moreover, while only three putative *T. urticae* salivary protein genes were expressed at lower levels in non-diapausing as compared to diapausing adults, we found that 26 genes were more strongly expressed in non-diapausing adults that are actively feeding (Bryon et al., 2013; TABLES 3.1 and S3.6). Finally, 15 putative *T. urticae* salivary proteins were also detected in a proteome analysis of *T. urticae* faeces (TABLE 3.1; Santamaría et al., 2015). These included cathepsins (tetur09g04400, tetur25g00650 and tetur12g01860), actins (tetur03g09480, tetur09g05350), a lipocalin (tetur04g05980), beta-galactosidase (tetur07g07380), fructose-biphosphate aldolase (tetur07g03440), glutathione S-transferases (tetur01g02510, tetur03g07920), vitellogenins (tetur39g00810, tetur43g00010), alpha2-macro-globulin (tetur18g03030) and two proteins with unknown function (tetur07g00160 and tetur29g01360).

3.3.4. ISH of salivary protein genes combined with morphological analysis distinguishes anterior and dorsal podocephalic glands

The spatial expression pattern of 15 putative *T. urticae* salivary protein genes was evaluated using whole-mount *in situ* hybridization (ISH). These genes were selected out of the nano-LC-MS/MS dataset based on the possession of a SP for secretion and a relatively high expression in the proterosoma (TABLE 3.1). Furthermore, we endeavored to include genes from as many different gene families as possible. To ensure that the observed ISH
signal is linked to the genuine morphological structure, *T. urticae* internal anatomy was investigated by means of X-ray sub-micron computed tomography. This imaging technique resulted in high-resolution three-dimensional pictures, providing additional insights into spider mite internal anatomy from those reported previously (Alberti and Crooker, 1985; Blauvelt, 1945; Mothes and Seitz, 1981), and allowed more accurate interpretation of the ISH signal observed by bright field and fluorescence confocal microscopy. Of the 15 genes, 14 were expressed specifically in either the anterior or dorsal podocephalic glands (FIGURE 3.3). None of the genes we tested showed expression in both the anterior and dorsal glands, and staining was not observed when using sense control probes. We did not detect gene expression in a third pair of glands annotated previously as being salivary, and known as the coxal organs (Mothes and Seitz, 1981). *Tetur13g00600* was also tested, yet no clear specific staining was visible in any gland or tissue.

### 3.4. Discussion

#### 3.4.1. Identification of salivary gland proteins secreted by the spider mite *T. urticae*

The spider mite *T. urticae* is well known for its ability to feed on an extraordinary wide range of different plant species even though their feeding activities induce plant defenses that can negatively affect mite fitness (Glas et al., 2014; Kant et al., 2008; Li et al., 2002). However, arthropod populations have adapted to suppress these defenses, including *T. urticae* populations (Alba et al., 2015; Godinho et al., 2015; Sarmento et al., 2011; Wybouw et al., 2015), suggesting that effectors in their saliva are crucial determinants for success in colonizing host plants. To gain insight into the salivary proteome of *T. urticae*, spider mites were allowed to feed on small hemispheres filled with artificial diet. Using this setup, mites can secrete saliva into the diet, which in turn can be used for nano-LC-MS/MS analysis. By combining this approach with extensive follow-up analyses, we aimed to obtain an as accurate as possible set of proteins that are actually injected by *T. urticae* into host plants. In the past, this kind of approach was also successful to identify aphid salivary proteins (Chaudhary et al., 2015; Rao et al., 2013b; Thorpe et al., 2016a). *Tetranychus urticae* gene expression profiles are known to be highly host plant dependent (Díaz-Riquelme et al., 2016; Wybouw et al., 2015). By using spider mites reared on four different host plant species for a minimum of five generations, we maximized
Proteomic composition of *T. urticae* saliva

Please see the diagrams for a visual representation.
chances of capturing a higher variety of salivary proteins. In addition, this allowed to study the influence of the host plant on spider mite salivary protein composition (Chapter 4).

By design, our method is suitable to collect saliva from small numbers of potentially tiny arthropods. In recent studies on aphid saliva (Chaudhary et al., 2015; Rao et al., 2013a; Thorpe et al., 2016b), tens of thousands of aphids were needed to collect saliva in sufficient quantities for protein identification. Despite the smaller size of *T. urticae* (0.4–0.5 mm) compared to aphids (> 1 mm), we managed to collect sufficient amounts of saliva for nano-LC-MS/MS analysis using only about 500 mites per sample. Limited volumes of diet were sufficient for the analysis via state-of-the-art mass spectrometry, making the analysis of salivary proteomes of small herbivorous arthropods time- and resource-efficient where feeding on an artificial diet is feasible.

**Figure 3.3. Localization of expression of *Tetranychus urticae* genes coding for putative salivary proteins identified by nano-LC-MS/MS.** A schematic representation of a mid-sagittal section of a *T. urticae* female is shown on top (redrafted from (Alberti and Crooker, 1985) with the permission from Elsevier). Dorsal podocephalic glands (DPGL) are shaded green while the anterior podocephalic glands (APGL) are shaded red. *Tetranychus urticae* genes for which ISH confirmed expression in one of the podocephalic salivary glands are framed by a box shaded red (APGL) or green (DPGL). A red arrow indicates the dissection position for collection of proterosomas. (A-D) Virtual sections obtained by a sub-micron CT scan of a *T. urticae* adult female confirmed the internal morphology as described by Alberti and Crooker (1985). (A) Virtual mid-sagittal section, dashed straight lines represent the other virtual sections that were created and are shown in panels B, C and D. (B) Virtual cross section at the tracheal glands (TRGL): DPGL and APGL are indicated with a green and red dashed line, respectively. (C) Virtual frontal section at TRGL, DPGL are indicated with a dashed green line. (D) Virtual frontal section at the central nervous mass (CNM), APGL are indicated with a dashed red line. (E-H) Confocal images of whole-mount in situ hybridization of putative *T. urticae* salivary protein genes. A DIG-labeled antisense probe was used for hybridization and the signal was developed using anti-DIG-AP and FastRed as substrate. The reaction product is visible as a red signal while the spider mite body shows green autofluorescence. Signal development corresponded to the localization of the podocephalic glands as shown in panels B, C and D. (E,G) Dorsal (Z-stack maximum intensity projection) and lateral (sagittal stack) view of signal development in the DPGL (*tetru07g00160*). (F,H) Dorsal (Z-stack maximum intensity projection) and lateral (sagittal stack) view of signal development in the APGL (*tetru28g01360*, panel F, and *tetru01g00950*, panel H). Other abbreviations: SILKGL, silk glands; ES, oesophagus; OV, ovaria; ST, stylet; VE, ventriculus; EX/HI, excretory organ/hindgut; and L2, second pair of legs.
Proteomic composition of *T. urticae* saliva

Salivary proteins need to be secreted by the salivary glands into the salivary duct before they can be delivered into a host plant via the saliva. The majority of the 95 putative *T. urticae* salivary proteins identified by nano-LC-MS/MS possess a SP and were predicted to be localized extracellularly, consistent with secretion. However, the lack of a SP for 18 out of the 95 proteins does not exclude that these proteins are secreted, as secretory pathways independent of the canonical endoplasmic-reticulum-Golgi network may exist (Haegeman et al., 2012). Additionally, SPs can be missed by gene prediction programs during genome annotation. In fact, both reasons have been proposed to explain why many putative arthropod salivary proteins lack a predicted SP (e.g., Chaudhary et al., 2015).

Although detection by LC-MS is one of the most robust methods to identify secreted salivary proteins, the list in TABLE 3.1 should not be considered exhaustive. A number of factors determine detection by nano-LC-MS/MS, including the adhesion to the material of the collection device during sample preparation, the presence of trypsin cleavage sites (Vandermarliere et al., 2013), the quality of MS/MS spectra, the type of chemical or post-translational modifications that were accounted for in the database search, and the protein hydrophobicity influencing the fractionation (Sheng et al., 2006). We validated proteomic findings with additional expression datasets. Previously, transcriptome analyses of salivary glands of insects have been reported (e.g., Celorio-Mancera et al., 2011; Stafford-Banks et al., 2014a; Su et al., 2012). Current technology, however, does not allow straightforward generation of specific spider mite salivary transcriptomes as spider mite glands are merely 50 μm in diameter. We therefore isolated spider mite protosomas, the anterior body region that includes the salivary glands next to other tissues such as the nervous mass and the silk glands (Figure 3.3). In this study, the salivary glands were confirmed to be localized in the protosoma using X-ray tomography. In a comparison between transcriptomes of *T. urticae* intact females and *T. urticae* protosomas, candidate genes for salivary gland expression are expected to be more highly expressed in the latter. Our analysis revealed more than 400 genes with a significantly higher expression in the protosoma. Reassuringly, most of the identified putative *T. urticae* salivary proteins are predicted to be secreted, are encoded by genes of which the expression is higher in the body region containing the salivary glands, and are expressed specifically in feeding spider mites (TABLES 3.1 and S3.6).
3.4.2. Genes encoding putative salivary proteins are expressed in the salivary glands

Despite multiple lines of evidence for salivary origins, the proteins we identified could still originate from another source. Of particular note, the catalogue of putative salivary proteins did show some overlap (15 out of 95), with proteins identified from spider mite faecal extracts (Santamaría et al., 2015). However, with ISH we showed that for genes encoding two of these proteins (tetur07g00160 and tetur29g01360) expression is specific to the salivary glands (TABLE 3.1, FIGURE 3.3). Hence, the proteins present in the faeces probably reflect a salivary origin. This indicates that secreted salivary proteins may end up in the digestive system during feeding, either directly or indirectly after ingesting cytoplasm from cells in which saliva had been injected. Interestingly, many putative T. urticae salivary proteins shared homology with D. farinae proteins that were predicted to have a function in digestion. ISH in T. urticae showed that some of the genes that code for presumptive gut proteins, for example tetur07g00150 (a serine protease) and tetur28g00360 (a beta-mannosidase), are actually expressed in the salivary glands (TABLE 3.1). Hence, they may have a digestive function in the saliva instead of, or in addition to, a potential role in the gut. Nevertheless, some proteins we found in the artificial diet could be fecal contaminants. This is also suggested by the proterosomal transcriptome, as some of the proteins detected in the diet are not differentially expressed, while the digestive enzymes confirmed by ISH are highly expressed in proterosomas (FIGURE 3.2, TABLE S3.9). Adult female spider mites were found to deposit eggs and silk on the diet membrane and together with the mite’s cuticle these could be other potential sources of contamination, e.g., for vitellogenin (tetur39g00810, tetur43g00010 and tetur516g00020). Among the 15 genes selected for ISH, we included some coding for proteins with either high or low rTop3 values, but all were more highly expressed in the proterosoma and belonged to the most prominent proteins or protein families in the dataset (TABLE 3.1). Of the 15 selected genes, 14 were found to be expressed in either the paired anterior or paired dorsal podocephalic glands. Both of these podocephalic salivary glands are predicted to be rich in proteins as assessed by ribosomes presence (Mills, 1973) and staining for proteins with methylene blue (Mothes and Seitz, 1981). Intriguingly, our data show that the anterior and dorsal podocephalic glands are responsible for the production of a discrete subset of the salivary proteome. Such ‘division of labor’ has been reported for the different secreto-
Proteomic composition of T. urticae saliva

ry cell types in the salivary glands of aphids (Mutti et al., 2008; Pan et al., 2015; Wang et al., 2015) and has been suggested for thrips as well (Stafford-Banks et al., 2014b). The regulatory mechanism by which these glands or cell types are ‘assigned’ to produce particular proteins in spider mites, aphids or thrips is not known. A functional distinction between anterior and dorsal podocephalic glands in spider mites was already predicted by Mothes and Seitz (1981). The production of a serous secretion was attributed to dorsal glands, while the anterior glands were predicted to produce a mucous secretion (Mothes and Seitz, 1981). However, supporting evidence for these specific predictions is not provided by our study.

For one of the 15 genes selected for ISH, expression in the salivary glands could not be confirmed, nor did we observe staining in other tissues. This does not exclude that this gene encodes a salivary protein, as ISH ‘failure’ may have several causes including poor probe design or expression below the detection limit.

3.4.3. The saliva of T. urticae comprises a cocktail of different protein families

Among the different putative T. urticae salivary proteins we identified, several were carbohydrate and protein degrading enzymes including cathepsins, serine proteases, glycoside hydrolases, beta-galactosidases and beta-mannosidases. Genes coding for serine proteases and beta-mannosidases also showed a higher expression in the proterosoma (TABLES 3.1 and S3.9). These catabolic proteins could have a digestive function during the feeding process before ingestion. For example, polysaccharide-digesting enzymes present in the saliva of the hemipteran herbivore Homalodisca vitripennis have been proposed to play a role in the degradation of cell wall material, hereby facilitating stylet penetration (Backus et al., 2012). Pre-digestion of plant material has been suggested to be a property of spider mite saliva (De Lillo and Monfreda, 2004; Mothes and Seitz, 1981; Storms, 1971) but there is no empirical evidence to support this hypothesis. Recently, the spider mite consumption rate was estimated at a single mesophyll cell per 10 min (Bensoussan et al., 2016), which is much lower than reported previously (Liesering, 1960). Hence, such a low consumption rate might allow pre-digestion of plant cells with spider mite salivary enzymes. However, secreted digestive enzymes could have a function in the mite itself, e.g., in the oesophagous, after ingestion. Like caterpillar feeding, also spider mite feed-
ing induces the plant to produce defensive proteins like serine protease inhibitors (PI-I and PI-II) (Kant et al., 2004; Kant et al., 2008; Li et al., 2002; Martel et al., 2015). The production of salivary serine proteases may be a means to compensate for inactivation of gut serine proteases. The mite itself also produces salivary proteinase inhibitors as we detected several cystatins (cysteine protease inhibitors: tetur06g01060, tetur06g06630, tetur09g03620, tetur09g03650 and tetur09g03670) secreted in the diet. Cystatins can play important roles in plant-pathogen interactions. The maize pathogen *Ustilago maydis* manipulates expression of a plant cystatin via its effector protein pep1, causing inactivation of defensive Cys proteases thereby making the plant susceptible to infection (van der Linde et al., 2012). In addition, *Phytophthora infestans* produces secreted serine protease inhibitors and cystatin-like effectors (Tian et al., 2007) to target key extracellular defensive proteases of its hosts and a host plant shift of this pathogen was attributed to a single amino acid change in one of these cystatin effectors (Dong et al., 2014). This illustrates how decisive effectors can be for determining host plant compatibility of a pest and it may suggest that unrelated organisms might have evolved effector proteins with similar functions (i.e., similar plant targets).

We identified a chitinase (tetur01g11910) among the *T. urticae* salivary proteins as well. Chitinase was also found in saliva of an aphid species (*Diuraphis noxia*) feeding on wheat (*Triticum aestivum*) and it was suggested that this enzyme might inhibit secondary fungal infections at the feeding site by hydrolyzing fungal chitin-rich cell walls (Nicholson et al., 2012). We found the chitinase only in the saliva of mites-lines adapted to both bean and maize (TABLES 3.1 and S3.6, FIGURE 3.2) suggesting there might be a host plant specific role for these proteins. However, salivary chitinases are not uncommon and were found, for example, in saliva of humans (Van Steijn et al., 1999), remipedes (von Reumont et al., 2014), octopuses (Grisley and Boyle, 1990), nematodes (Gao et al., 2002; Niblack et al., 2006) and insects (Nicholson et al., 2012) and were suggested to play a defensive role against chitinous pathogens. Finally, among characterized proteins, a superoxide dismutase (SOD, tetur26g02320), which targets reactive oxygen species that are often induced in plants upon damage, was also detected. Noteworthy, SODs have also been identified in the saliva of a potato leafhopper (*Empoasca fabae*) and two aphids (*D. noxia* and *Ropalosiphum padi*) (Sharma et al., 2014).
Proteomic composition of T. urticae saliva

Intriguingly, several of the most abundant proteins in the saliva are coded by multi-membered gene families of unknown function. Most of these proteins were encoded by genes with proterosoma-enriched expression, and lacked homologues in D. farinae, a non-plant feeding mite. One possibility, although speculative, is that the proliferation of these families facilitated host-range expansion in T. urticae. Polyphagous species are proposed to harbor a larger collection of salivary proteins since they are exposed to a greater diversity of selection pressures, i.e., host plants that vary in morphology and (defensive) physiology (Vandermoten et al., 2014). Analogous gene family proliferations have been described for other proteins relevant for plant-spider mite interactions, such as detoxification enzymes (cytochrome P450s, carboxyl/choline esterases and glutathione S-transferases), transporters (Dermauw et al., 2013a,b; Van Leeuwen and Dermauw, 2016) and digestive cysteine peptidases (Santamaría et al., 2012).

Furthermore, host plant dependent expression of salivary protein genes was suggested when comparing the composition of salivary secretions of spider mites adapted to the different host plants. This observation was studied in more detail in Chapter 4, where statistically supported evidence for host plant dependent gene expression was supplied by whole-genome transcriptome analysis of the four host plant (T. urticae adapted to bean, maize, soy or tomato) adapted lines used for proteomics analysis (this chapter), and an additional T. urticae line adapted to cotton (Gossypium hirsutum).

3.4.4. Prediction of effector proteins

Confidently identifying salivary proteins as effectors is not straightforward. Arthropod effector proteins are characterized by the presence of a N-terminal SP, directing them to the secretory pathway, and their encoding genes show signatures of rapid evolution (poor sequence similarity with other genera, multiple gene copies, and high rates of non-synonymous nucleotide mutations) (Aggarwal et al., 2014; Mitchum et al., 2013; Pitino and Hogenhout, 2013; Zhao et al., 2015). Due to the highly specific function of effectors, variable expression levels depending on the host plant could also be indicative of a context-dependent function (Pan et al., 2015). The presence of R genes, which turn effectors effectively into elicitors, can differ among plant species or varieties (Kanvil et al., 2014; Stuart, 2015; Stuart et al., 2012), as does the presence of S genes (van Schie and Takken, 2014), the target of the effector. As such, the R and S gene composition of a host may
determine the specific transcriptional response of a herbivore’s secretome and thus its effector repertoire. Proteins of the OrthoMCL groups OG5_144177, Tu_MCL_25, Tu_MCL_35 and Tu_MCL_36 have characteristics attributed to (arthropod) effectors: (I) they are targeted to the saliva and are secreted during feeding, (II) they share no homology with proteins of non-phytophagous mite species, and (III) they belong to multi-gene families (>10 members) (TABLES 3.1, S3.6 and S3.7). Furthermore (IV), host-dependent production was suggested for some of the proteins belonging to these groups (studied in more detail in Chapter 4). Intriguingly, salivary proteins from the OrthoMCL groups Tu_MCL_25 and Tu_MCL_211 (tetur31g01040 and tetur01g01000, respectively, see TABLE S3.6), were shown to act as effector proteins (Villarroel et al., 2016). These proteins were identified using an in silico approach and were shown to enhance T. urticae reproductive performance when transiently expressed in N. benthamiana leaves (Villarroel et al., 2016) (Chapter 1). However, future research is needed to unravel the function of the T. urticae salivary constituents documented here. The peptide composition of saliva was determined as well (Chapter 5), while one effector protein family was studied in more detail (Chapter 5).

3.5. Conclusions

Despite the minute size of T. urticae, we managed to collect saliva for proteome analysis through an artificial diet system. Using nano-LC-MS/MS, 95 putative T. urticae salivary proteins were identified, indicating that these mites employ a complex protein cocktail in their interaction with host plants. A proterosoma specific transcriptome is presented as a second discovery pipeline, and a considerable overlap with nano-LC-MS/MS data was observed. A selection of genes coding for putative salivary proteins was confirmed to be expressed in the salivary glands by whole-mount in situ hybridizations. Several proteins belong to protein families with as yet unknown functions, with some having structural and gene family features suggestive of roles as effector proteins. The search for effectors and their possible targets is essential to our understanding of polyphagy and the evolution of the plant-mite interactions, but it is also of practical importance, as it would enable plant breeders to discover new R and S genes in order to develop mite-resistant crops.
Proteomic composition of *T. urticae* saliva

This project was supported by the European Commission (EC contract 618105) via FACCE ERA-NET Plus and FACCE-JP (Genomite, project ID 137), the Fund for Scientific Research Flanders (FWO) (grant G009312N to L.T. and T.V.L. and grant G053815N to L.T., T.V.L. and W.D.), The Special Research Fund (DOZA 01J131711 to W.J.) and NSF award DEB 1457346 (to R.M.C). R.G. was supported by National Institutes of Health Genetics Training Grant T32 GM07464. C.A.V was supported by CONICYT BECAS (Chile) and M.R.K. by NWO (The Netherlands) (STW-VIDI 13492). N.W. was supported by a Marie Sklodowska-Curie Action (MSCA) Individual fellowship (658795-DOGMITE) of Horizon 2020. M.G. acknowledges funding by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-046) and Ontario Research Fund- Global Leadership in Genomics and Life Sciences GL2–01-035. W.D. and G.M. are postdoctoral fellows of the Fund for Scientific Research Flanders (FWO). Some of the computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, the Hercules Foundation and the Flemish Government – department EWI.

3.6. References


Chapter 3


Blauvelt, W.E. (1945). The internal morphology of the common red spider mite (Tetranychus telarius Linn.). Cornell University Agricultural Experiment Station Memoir, 1-41.


Proteomic composition of T. urticae saliva


Proteomic composition of T. urticae saliva


Proteomic composition of *T. urticae* saliva


Chapter 3


**Proteomic composition of T. urticae saliva**


**Figure S3.1. Custom built device used for producing artificial diet hemispheres.** The custom built vacuum device consisted of (A) a 96 hole well plate (1) fitting in a bottom vacuum manifold base (2) with a 40 micron porous steel (surrounded by a neoprene gasket) in between (3). (B) a vacuum pump (not displayed) is connected to the vacuum manifold base (arrow points to connection point). Plastic packaging tape was used to restrict suction to one of the 96 holes when creating individual parafilm hemispheres (not shown).
FIGURE S3.2. Volcano plot of differentially expressed genes between *Tetranychus urticae* proterosomas and intact *T. urticae* females. The negative log_{10} of Benjamini-Hochberg adjusted p-value was plotted against the log_{2}FC in gene expression for all *T. urticae* genes. Differentially expressed *T. urticae* genes (|FC| \geq 8, Benjamini-Hochberg adjusted \( p \leq 0.05 \)) are shown as grey dots (DE), while genes that were not differentially expressed are shown as grey circles (not DE). Genes coding for salivary proteins that were identified by nano-LC-MS/MS ('LC-MS salivary protein genes') are depicted as stars. LC-MS salivary protein genes that were differentially expressed are depicted as green stars (DE/LC-MS). LC-MS salivary proteins that were differentially expressed and of which ISH confirmed their expression in the salivary glands are depicted as red stars (DE/LC-MS/ISH) while LC-MS salivary proteins that are not differentially expressed are shown as blue stars (not DE/LC-MS).
**Proteomic composition of *T. urticae* saliva**

### 3.7. Supplementary Material

Supplementary material can be found at http://doi.org/10.1074/mcp.M116.058081.

**Table S(3.1)** - Detailed list of all features of the LC-MS/MS identifications at the protein level of all treatment samples [artificial diet enriched with saliva of mites, 8 samples in total, 2 technical replicates for every host plant (bean, maize, soy, tomato) adapted *T. urticae* line]. The export generated from PeptideShaker holds all required information necessary as stated by the guidelines on publication of mass spectrometry protein identifications.

**Table S(3.2)** - Detailed list of all features of the LC-MS/MS identifications at the peptide level of all treatment samples [artificial diet enriched with saliva of mites, 8 samples in total, 2 technical replicates for every host plant (bean, maize, soy, tomato) adapted *T. urticae* line]. The export generated from PeptideShaker holds all required information necessary as stated by the guidelines on publication of mass spectrometry peptide identifications.

**Table S(3.3)** - Detailed list of all features of the LC-MS/MS identifications at the protein level in reference samples (artificial diet, 2 samples in total).

**Table S(3.4)** - MS1 intensities from the Thermo raw files extracted using modest Feature Finder (https://github.com/compomics/moFF).

**Table S(3.5)** - Primers used in this study.

**Table S(3.6)** - Detailed overview of putative *T. urticae* salivary proteins identified in artificial diet using nano-LC-MS/MS.

**Table S(3.7)** - *T. urticae* proteins and their OrthoMCL groups. Groups starting with “Tu_MCL” contain *T. urticae* proteins that could not be mapped to existing OrthoMCL groups (“OG5_” groups).

**Table S(3.8)** - The number of mapped RNA reads per *T. urticae* gene for both intact *T. urticae* females and *T. urticae* proterosomas. Mapped reads were counted using HTseq (HYPERLINK \l "Anders et al., 2014).""

**Table S(3.9)** - Differentially expressed genes (Benjamini-Hochberg adjusted p-value ≤ 0.05 and |FC| >= 8) between *T. urticae* proterosomas and intact *T. urticae* females.

**Data S(3.1)** - *T. urticae* proteome, version of December 16th 2014 (FASTA-format).