The salivary proteome of Tetranychus urticae

Key to its polyphagous nature?

Jonckheere, W.S.A.

Publication date
2018
Document Version
Other version
License
Other

4

A link between host plant range and the salivary protein repertoire of spider mites

Partly 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, M.R., Vanholme, B., Menschaert, G., Van Leeuwen, T. (2016)
The salivary protein repertoire of the polyphagous spider mite *Tetranychus urticae*: a quest for effectors. *Molecular & Cellular Proteomics* 15, 3594-3613

4.0. Abstract

The proteomic composition of secreted saliva from *Tetranychus urticae* lines adapted to different host plant species (bean, maize, soy and tomato) was analyzed using a custom-developed feeding assay coupled with nano-LC tandem mass spectrometry. The mite's salivary protein composition depended on the host plant the mite was adapted to (**Chapter 3**). To supplement the proteomics analysis with statistically supported data, genome-wide expression profiling was performed on *T. urticae* lines adapted to five different hosts (bean, maize, soy, tomato, and additionally, cotton). Indeed, the plant species had an effect on the expression level of several salivary protein genes. Furthermore, the genomes of two closely related spider mite species with a distinct host range – the *Ulex europaeus* specialist *T. lintearius* and the Solanaceae specialist *T. evansi* – were searched for homologs of the salivary proteins identified in *T. urticae*. For some salivary protein families the number of members and the size of the host plant repertoire of the mite was correlated. This study confirms that the salivary protein cocktail delivered in the host depends on the plant species. In addition, it provides support for the hypothesis that the salivary proteome of spider mites co-evolved with the host plant range. This would suggest polyphagous herbivores require a more complex salivary (effector) protein repertoire than monophagous species.

4.1. Introduction

The cosmopolitan spider mite *Tetranychus urticae* is able to feed on an impressive array of different plant species. With more than 1100 different hosts recorded, belonging to more than 140 different families (Migeon and Dorkeld, 2006-2017), this species can considered to be one of the most polyphagous arthropods. Distinct host plants represent different challenges for herbivores, as each is characterized by its own nutritional composition and defensive compounds. Furthermore, distinct plant species may respond differently to herbivory **(Chapter 1)**. Considering these differences amongst plant species, it does not come as a surprise that mites adjust the expression of genes coding for proteins involved in digestion and detoxification to the host plant they are feeding on (Dermauw et al., 2013; Wybouw et al., 2015; Van Leeuwen and Dermauw, 2016). As such, *T. urticae* is well equipped to deal with the diverse secondary metabolites it encounters when feeding on its various hosts (Dermauw et al., 2013; Van Leeuwen and Dermauw, 2016).

In order to breach plant defensive barriers, herbivores produce effectors which target the products of certain plant susceptibility genes (S genes). Plants have, however, evolved to recognize some of these effectors by means of resistance gene (R gene) encoded receptors, effectively turning these effectors into elicitors of plant defenses. The interaction between effector and R or S gene product is highly specific and the nature of these genes can determine plant-herbivore compatibility (Kaloshian, 2004; Harris et al., 2012; van Schie and Takken, 2014). Reminiscent of digestive enzymes and ABC-transporters, the expression level of herbivore effectors may differ between host plants. The putative aphid effector ACYPI006346, for example, has different transcript levels in aphids adapted to different plants (Pan et al., 2015). Differential expression of effectors after adaptation to a new host plant may also occur in spider mites since mite adaptation has been shown to affect the transcriptional response of the host plant (Wybouw et al., 2015).

With the salivary proteome of *T. urticae* elucidated (**Chapter 3**), and with a host-specific salivary protein composition suggested by this proteomics analysis, we additionally studied the host-dependent expression level of all identified salivary proteins using genome-wide transcriptomics analysis of *T. urticae* lines adapted to five host plants (i.e., bean, maize, soy, tomato, and additionally, cotton).

Furthermore, the occurrence of homologs of *T. urticae* salivary proteins was studied in the related spider mite species *T. lintearius* and *T. evan-*

si. These mites are specialists on *Ulex europaeus* (Fabaceae) and the Solanaceae family respectively (Ireson et al., 2003; Migeon and Dorkeld, 2006-2017; Navajas et al., 2013). A glimpse into their salivary protein repertoire can provide additional insights in the molecular machinery of salivary proteins and will allow a better understanding of the link between salivary proteins and the compatibility of the plant-mite interaction.

4.2. Material and methods

4.2.1. 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, and the genome of this strain has been sequenced (Grbić et al., 2011). 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. 'Ronaldinio', Poaceae), soy (*Glycine max* cv. 'Merlin', Fabaceae) and tomato (*Solanum lycopersicum*, cv. 'Moneymaker', Solanaceae). Three independent lines were generated for tomato (Wybouw et al., 2015) and cotton, while four independent lines were obtained for the other hosts. The mite lines were maintained in a climatically controlled environment at 26°C with 60% RH, and a photoperiod of 16:8 h light:dark. Mites were offered fresh plants as needed, and were used in experiments after 5 generations for all hosts, except tomato, where replicate lines were adapted and maintained on tomato for over 30 generations (Wybouw et al., 2015).

4.2.2. Host plant-dependent salivary protein identification in artificial diet probed by *T. urticae*

The list of *T. urticae* salivary proteins, identified in **Chapter 3** using nano-LC-MS/MS analysis of *T. urticae* probed artificial diet, was split up again over the separate host plant lines (i.e., bean, maize, soy and tomato). Host plant specificity of putative *T. urticae* salivary proteins was visualized by means of a venn diagram depicting the number of proteins that are uniquely identified in the samples of a specific host plant line and the number of proteins that are shared by different samples (Figure 4.1.A). In addition, a heatmap visualizing the mean rTop3 factor of a selection of candidate salivary proteins (maximum rTop3 value higher than the 30th percentile of all

maximum rTop3 values) for the different investigated host plants was generated (FIGURE 4.1.B).

4.2.3. Transcriptome analysis of *T. urticae* maintained on different host plants

For each host plant adapted mite line, three (tomato and cotton) or four (bean, maize and soy) biologically replicated RNA samples were obtained. Each RNA sample was extracted from 100-150 pooled female adult mites using the RNeasy extraction kit (Qiagen) and was subsequently treated with DNase (Turbo DNA-free kit, Ambion). RNA quantity and integrity was measured using an Agilent TapeStation system. RNA samples were labeled with cyanine dyes following the Low Input Quick Amp Labeling Kit (Agilent Technologies), with 100 ng of total RNA as starting material. RNA samples from mites feeding on the reference bean host plant were labeled with cy3, while cy5-labelling was performed on all other samples. Samples were pooled per host plant transfer and hybridized to a custom-made Sureprint G3 8x60K array (Agilent Technologies, with a GEO platform number of GPL16890) following the standard procedure of the Gene Expression Hybridization Kit (Agilent Technologies). After washing procedures (Gene Expression Wash Buffer kit; Agilent Technologies), raw data was extracted from the 8x60k slides using the GE2_107_Sep09 protocol of the Agilent Feature Extraction Software. The intraspot correlation coefficient per array and the metrics from the arrayQualityMetrics package per host plant line were assessed for optimal background correction and normalization procedures (Kauffmann et al., 2009). Data was background corrected using the 'normexp'-method and normalized by loess and Aquantile (Ritchie et al., 2007). Cyanine intensities were extracted from the processed RG-object and averaged per host plant. Using the normalized MA-object, differential expression was assessed for mites on cotton, maize, soy and tomato against the corresponding ancestral mite population living on bean by an empirical Bayes approach. A heatmap vizualizing the expression levels (absolute and relative cyanine intensities) of putative salivary protein-encoding genes in mites adapted to the different hosts was generated (FIGURE 4.2). *Tetranychus urticae* gene expression data have been uploaded to the Gene Expression Omnibus with accession number GSE80337.

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 (v. 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) (FIGURE 4.3).

4.2.4. Mining of the *T. lintearius*, *T. evansi* and *T. urticae* proteomes for homologs of salivary proteins identified in *T. urticae*

The online OrthoMCL software tool (Li et al., 2003, Chen et al., 2006) was used to search the proteomes of *T. urticae*, the monophagous *T. lintearius* and the oligophagous *T. evansi* (versions of 11/08/2016) for homologs of the *T. urticae* salivary proteins identified in **Chapter 3** (TABLE 3.1). The genomes of *T. lintearius* and *T. evansi* are unpublished, yet are available to researchers from the Spider Mite Consortium via the ORCAE website (Sterck et al., 2012).

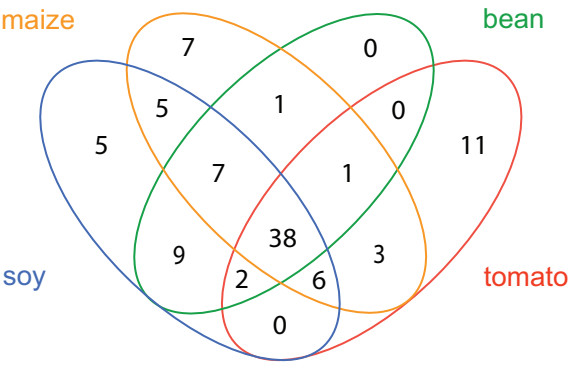
4.3. RESULTS

4.3.1. Host plant-dependent salivary protein composition

Host plant specificity of putative *T. urticae* salivary proteins (FIGURE 4.1.A), obtained by a proteomics analysis of artificial diet fed upon by mite lines

FIGURE 4.1. Overview of nano-LC-MS/MS identified putative *Tetranychus urticae* salivary proteins. (A) Venn diagram depicting overlap between putative *T. urticae* salivary proteins secreted by mites adapted to different host plants (bean, maize, soy, tomato). Only those salivary proteins with a mean PSM of at least two in at least one of the *T. urticae* host plant adapted lines were used for comparison (see TABLES 3.1 and S3.6). **(B)** Heat map of mean rTop3 values of putative *T. urticae* salivary proteins secreted by mites adapted to different host plants (bean, maize, soy, tomato). Only those salivary proteins (and 'protein inference groups') with a mean PSM of at least two in at least one of the *T. urticae* host plant adapted lines and with a maximum rTop3 value higher than the 30th percentile of maximum rTop3 values were used for comparison (see TABLES 3.1 and S3.6). The Euclidean distance metric and Ward's method were used for clustering of both rows and columns. All putative salivary proteins for which the corresponding genes were shown to be expressed in the salivary glands by ISH (FIGURE 3.3) are shaded grey.

A


maize
bean
soy
tomato
7
0
5
1
0
5
11
7
1
38
9
3
2
6
0

B


rTop3
0
0.05
0.1
0.15
tetur01g01850
tetur20g00560
tetur07g00150
tetur03g03700
tetur03g03670, tetur03g03690, tetur03g03730, tetur03g10093
tetur03g03680
tetur12g03940, tetur12g03950
tetur28g01330, tetur55g00040
tetur32g00050
tetur06g06630
tetur16g03470
tetur06g03320
tetur06g02580
tetur05g04580
tetur16g03420
tetur29g01360
tetur07g00160
tetur01g02670
tetur28g00360
tetur04g05980
tetur05g04560
tetur14g02090
tetur31g00630
tetur03g10083
tetur03g03680, tetur03g03700, tetur03g10083
tetur40g00392, tetur95g00080
tetur06g00510
tetur07g01660
tetur03g08710
tetur14g02080
tetur516g00020
tetur55g00040
tetur28g01330
tetur09g04400
tetur12g01860
tetur31g00830
tetur03g08030
tetur16g03190
tetur02g04310
tetur14g02070
tetur10g00090
tetur10g00100
tetur01g00950
tetur01g01010
tetur07g07380
tetur04g09479
tetur06g00230
tetur32g01880
tetur01g05230, tetur01g16473, tetur06g02570, tetur06g02580
tetur28g01720
tetur30g01440
tetur03g07920
tetur31g01040, tetur31g01060
tetur23g01640
tetur01g02510
tetur43g00010
tetur39g00810
tomato
maize
soy
bean

TABLE 4.1. The number of identified *Tetranychus evansi*, *T. lintearius* and *T. urticae* proteins homologous to proteins identified in artificial diet fed upon by *T. urticae*. Only proteins which could be assigned to groups already existing in the OrthoMCL database were accounted for in this list. Detailed information can be found in TABLE S4.1.

OrthoMCL	*T. evansi*	*T. lintearius*	*T. urticae*
OG5_126560	6	3	7
OG5_126583	7	6	10
OG5_126595	2	2	3
OG5_126607	12	9	21
OG5_126738	8	10	10
OG5_126942	6	8	15
OG5_127143	2	2	3
OG5_127584	8	8	8
OG5_127620	2	2	2
OG5_127800	9	10	29
OG5_128075	2	2	3
OG5_128163	3	3	4
OG5_128170	2	3	20
OG5_129300	3	1	4
OG5_129423	15	14	35
OG5_130246	1	1	1
OG5_130527	17	13	52
OG5_131746	2	3	2
OG5_132251	0	6	6
OG5_133467	8	7	15
OG5_134456	1	1	1
OG5_135950	3	5	5
OG5_136350	4	3	6
OG5_141111	3	2	2
OG5_144177	0	0	6
OG5_147492	5	4	5
OG5_149533	3	2	13
OG5_152237	0	1	2
OG5_152337	3	1	1
OG5_152454	3	3	3
OG5_158831	0	0	2
OG5_168371	0	0	5
OG5_176862	0	0	1
OG5_207753	0	0	2

adapted to different host plants (**Chapter 3**) was further illustrated by means of a heat map depicting the mean rTop3 factor of a selection of candidate salivary proteins (maximum rTop3 value higher than the 30th percentile of all maximum rTop3 values) for the different investigated host plants

(FIGURE 4.1.B). The most apparent case of host-specific salivary proteins was tetur55g00040/tetur28g01330 (WTSP1, belonging to OrthoMCL group TuMCL_36), proteins with an unknown function, which were uniquely and abundantly identified from saliva of the tomato-adapted mites. Based on the clustering analysis, the salivary proteomic repertoire of *T. urticae* feeding on bean plants seemed most similar to the repertoire of soy-adapted mites (both plant species belong to the Fabaceae). However, since only one biological replicate (with two technical replicates) was analyzed with Nano-LC-MS/MS for each host plant specific diet sample, one should be careful to draw conclusions solely based on this comparative proteomics analysis.

4.3.2. Host plant-dependent expression of *T. urticae* salivary protein genes

To additionally validate the host-specific findings based on individual biological replicates, as well as to compare in further detail the expression level of genes coding for *T. urticae* putative salivary proteins across host plant species, we performed a genome-wide expression analysis of the mite lines that were used in the proteomic experiments (i.e., *T. urticae* adapted to maize, soy or tomato relative to the reference line on bean) using an Agilent gene expression microarray. Additionally, a line adapted to cotton during five generations was also investigated. We determined the absolute expression levels using normalized cyanine 3 (cy3) intensity values but also calculated the expression levels relative to the mites feeding on bean as a benchmark (FIGURE 4.2.A and B). We confirmed what we detected previously when comparing the rTop3 values of the proteomics data across different host plant lines (FIGURE 4.1): the salivary composition of *T. urticae* is host plant dependent. For example, in the tomato-adapted mite lines several salivary protein genes were highly expressed relative to bean-adapted mites: *tetur32g00050* (Tu_MCL_35), *tetur28g01330* and *tetur55g00040* (Tu_MCL_36) and *tetur31g00830* (Tu_MCL_25). Alternatively, *tetur03g03670*, *tetur03g10093*, *tetur03g03700*, *tetur03g03680*, *tetur03g03690* and *tetur03g03730* (OrthoMCL cluster OG5_144177; SHOT family, **Chapter 5**) were expressed at much lower levels in mites feeding from all host plants except for soy, compared to bean. The absolute expression level of the latter six genes was very high after feeding on bean or soy (cy3 intensity levels were in the top 1% of most highly expressed genes in mites on either bean or soy) and relatively low after feeding on the non-leguminous host plants under study (fold changes

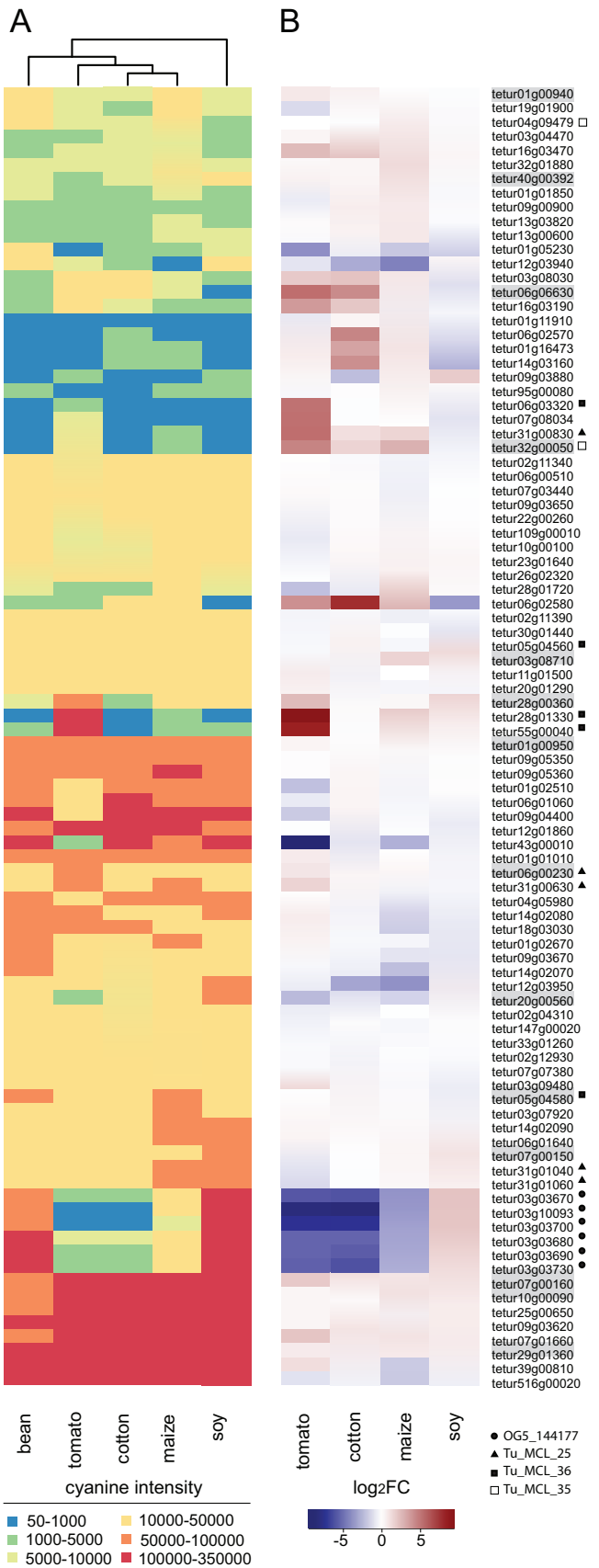

A
B
tetur01g00940
tetur19g01900
tetur04g09479
tetur03g04470
tetur16g03470
tetur32g01880
tetur40g00392
tetur01g01850
tetur09g00900
tetur13g03820
tetur13g00600
tetur01g05230
tetur12g03940
tetur03g08030
tetur06g06630
tetur16g03190
tetur01g11910
tetur06g02570
tetur01g16473
tetur14g03160
tetur09g03880
tetur95g00080
tetur06g03320
tetur07g08034
tetur31g00830
tetur32g00050
tetur02g11340
tetur06g00510
tetur07g03440
tetur09g03650
tetur22g00260
tetur109g00010
tetur10g00100
tetur23g01640
tetur26g02320
tetur28g01720
tetur06g02580
tetur02g11390
tetur30g01440
tetur05g04560
tetur03g08710
tetur11g01500
tetur20g01290
tetur28g00360
tetur28g01330
tetur55g00040
tetur01g00950
tetur09g05350
tetur09g05360
tetur01g02510
tetur06g01060
tetur09g04400
tetur12g01860
tetur43g00010
tetur01g01010
tetur06g00230
tetur31g00630
tetur04g05980
tetur14g02080
tetur18g03030
tetur01g02670
tetur09g03670
tetur14g02070
tetur12g03950
tetur20g00560
tetur02g04310
tetur147g00020
tetur33g01260
tetur02g12930
tetur07g07380
tetur03g09480
tetur05g04580
tetur03g07920
tetur14g02090
tetur06g01640
tetur07g00150
tetur31g01040
tetur31g01060
tetur03g03670
tetur03g10093
tetur03g03700
tetur03g03680
tetur03g03690
tetur03g03730
tetur07g00160
tetur10g00090
tetur25g00650
tetur09g03620
tetur07g01660
tetur29g01360
tetur39g00810
tetur516g00020
bean
tomato
cotton
maize
soy
cyanine intensity
50-1000
1000-5000
5000-10000
10000-50000
50000-100000
100000-350000
tomato
cotton
maize
soy
log2FC
-5
0
5
OG5_144177
Tu_MCL_25
Tu_MCL_36
Tu_MCL_35

Table 4.2. The number of identified spider mite specific *Tetranychus evansi*, *T. lintearius* and *T. urticae* proteins homologous to proteins identified in artificial diet fed upon by *T. urticae*. Detailed information can be found in Table S4.2.

OrthoMCL	*T. evansi*	*T. lintearius*	*T. urticae*
Tu_MCL_12	1	1	37
Tu_MCL_25	2	4	17
Tu_MCL_43	2	6	9
Tu_MCL_35	3	0	14
Tu_MCL_36	4	0	12
Tu_MCL_45	0	0	12
Tu_MCL_63	0	2	8
Tu_MCL_74	2	1	5
Tu_MCL_211	2	2	2
Tu_MCL_212	1	2	2
Tu_MCL_153	1	2	2
No group	1	3	1
No group	2	0	2
No group	1	1	1
No group	0	0	3
No group	0	0	2
No group	1	0	1
No group	1	0	1

between mites on bean and mites on tomato, maize or cotton varied between 7 and 289).

4.3.3. Mining of the *T. lintearius*, *T. evansi* and *T. urticae* genomes for homologs of salivary protein genes identified in *T. urticae*

The genomes of *T. urticae*, *T. lintearius* and *T. evansi* were searched for homologs of *T. urticae* salivary protein genes identified in **Chapter 3**

Figure 4.2. Heatmap of expression levels of putative *Tetranychus urticae* salivary protein encoding genes and their up- or downregulation in mites adapted to different host plants. (A) Heatmap of cyanine intensities of putative *T. urticae* salivary protein encoding genes. The Euclidean distance metric and Ward's method were used for clustering of both rows and columns. For 92 out of 95 putative salivary protein genes expression data was available. **(B)** Heatmap of $\log_2$FCs of putative salivary protein genes in mites adapted to soy, maize, cotton or tomato as compared to mites adapted to bean. Genes are sorted based on their order in panel A. Genes that were shown to be expressed in the salivary glands by ISH (Figure 3.3) are shaded grey. A circle, triangle, filled square and empty square indicates whether a gene belongs to OrthoMCL cluster OG5_144177, Tu_MCL_25, Tu_MCL_36 and Tu_MCL_35, respectively.

(TABLE 3.1). The OrthoMCL analysis grouped the proteins homologous to the identified putative *T. urticae* salivary proteins into 34 groups which already existed in the OrthoMCL database (http://www.orthomcl.org/orthomcl/) (TABLES 4.1 and S4.1), and in 18 groups which did not exist in the OrthoMCL database yet. Latter group names start with 'Tu_MCL' and are likely spider mite specific (TABLES 4.2 and S4.2). In what follows, I zoom in on some peculiarities.

The number of identified proteins belonging to an already existing group was, with 304 proteins, highest in *T. urticae*. *Tetranychus evansi* and *T. lintearius* had less than half of this number: 140 proteins for *T. evansi* and 135 proteins for *T. lintearius*. In 91% of these already existing OrthoMCL groups, *T. urticae* had at least as many members as *T. evansi* or *T. lintearius*. The extreme cases were OG5_127800 (Peptidase C1A, papain C-terminal), with 29 members in *T. urticae*, 9 in *T. evansi* and 10 in *T. lintearius* (TABLES 4.1 and S4.1). OG5_128170 (short-chain dehydrogenase/reductase SDR) had 20 members in *T. urticae*, yet only 2 and 3 in *T. evansi* and *T. lintearius* respectively. OG5_129423 (glycosyl hydrolase) had 35 members in *T. urticae*, 15 in *T. evansi* and 14 in *T. lintearius*. OG5_130527 (lipocalins) was, with 52 members, extremely expanded in *T. urticae*, while only 17 and 13 homologs were identified in *T. evansi* and *T. lintearius* respectively.

The number of identified proteins belonging to spider mite specific groups was, with 132 proteins, highest in *T. urticae*, while *T. evansi* and *T. lintearius* had less than one fifth of this amount: 25 proteins for *T. evansi* and 24 proteins for *T. lintearius*. In 89% of these spider mite specific OrthoMCL groups, *T. urticae* has at least as many members as *T. evansi* or *T. lintearius* (TABLES 4.2 and S4.2). The absolute expression levels of the *T. urticae* homologs of mites adapted to different host plants are shown in FIGURE 4.4.

The spider mite specific OrthoMCL group that includes most genes is **Tu_MCL_12**, which is, with 37 representatives, highly expanded in *T. urticae*. Strikingly, *T. evansi* and *T. lintearius* each possess only one homolog (tetev485g00030 and tetli56g00590 respectively). Proteins of Tu_MCL_12 lack homology with other known proteins, and their function remains unknown. Despite the expanded nature of Tu_MCL_12 in *T.* urticae, only tetur20g00560 could reliably be identified from secreted saliva. This protein may be one of the most prominent (putative) effectors in *T. urticae* saliva. Indeed, tetur20g00560 has the highest maximal rTop3 value of all putative salivary proteins identified in artificial diet fed upon by *T. urticae*

(Table 3.1). Furthermore, of all *T. urticae* homologs, only the expression of genes *tetur13g04230*, *tetur20g00050*, *tetur20g00090*, *tetur20g00540* and *tetur20g00560* was significantly higher (Benjamini-Hochberg adjusted $p \leq 0.05$ and FC ≥ 8) in the proterosoma compared to the entire body. Genes encoding other homologs were not differentially expressed (Data S3.9).

The second most gene-rich mite-specific OrthoMCL group is **Tu_MCL_25**. Again this group is expanded in *T. urticae*, where 17 representatives were found. Eight genes coding for these proteins have a significantly higher expression (Benjamini-Hochberg adjusted $p \leq 0.05$ and FC ≥ 8) in the proterosoma relative to the entire body (Data S3.9), while four proteins were detected in the artificial diet fed upon by *T. urticae*. In addition, several of the genes have a host-dependent expression level (Wybouw et al., 2015). In *T. evansi*, two homologs were identified and four could be found in *T. lintearius* (Tables 4.2 and S4.2). The Tu_MCL_25 cluster includes proteins from 'family 28' (Tu28 and Te28), which we have shown to be effectors (**Chapter 2**).

The next spider mite specific OrthoMCL group that catches the eye is **Tu_MCL_43**. This group is expanded in *T. urticae* (nine members), has two members in *T. evansi*, and is with six members relatively well represented in *T. lintearius*. Seven out of the nine homologs in *T. urticae* have a high gene expression level in the proterosoma (Benjamini-Hochberg adjusted $p \leq 0.05$ and FC ≥ 8) (Data S3.9).

Tu_MCL_35 is expanded (14 members) in *T. urticae*, yet only *tetur04g09479* and *tetur32g00050* are significantly upregulated in the proterosoma, while *tetur12g00700* is even strongly downregulated in this head region (Benjamini-Hochberg adjusted $p \leq 0.05$ and $|FC| \geq 8$). Host dependent expression levels are evident (Figure 4.4). Tu_MCL_35 has three homologs in *T. evansi* and appears to lack representatives in *T. lintearius* (Tables 4.2 and S4.2).

Some genes of the **Tu_MCL_36** group (WTSP) are highly expressed when *T. urticae* is feeding on tomato plants (Figure 4.3). Furthermore, two proteins of this group (tetur28g01330 and tetur55g00040) were uniquely and abundantly identified from saliva of *T. urticae* adapted to tomato plants. This points to a tomato- or Solanaceae-specific function for these two proteins. Intriguingly, the Solanaceae specialist *T. evansi* possesses four genes coding for Tu_MCL_36 homologs, while homologs could not be found in the gorse-specialist *T. lintearius* (Tables 4.2 and S4.2).

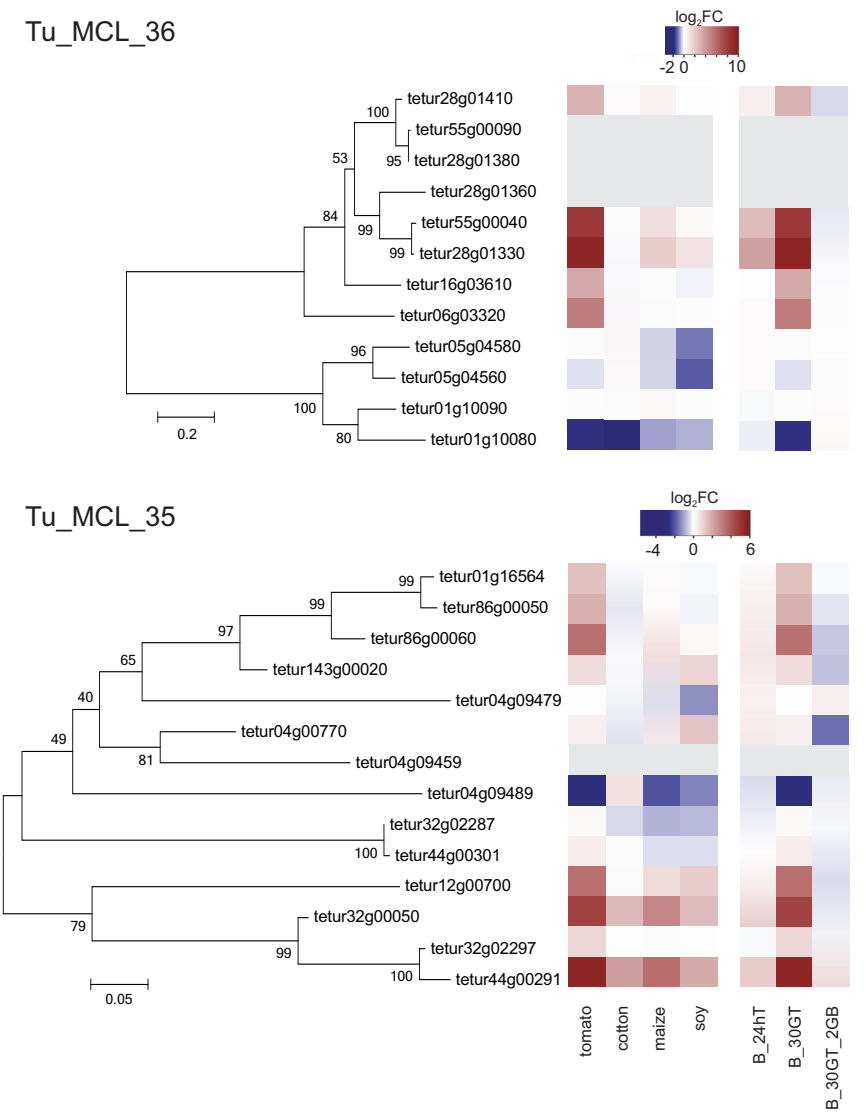
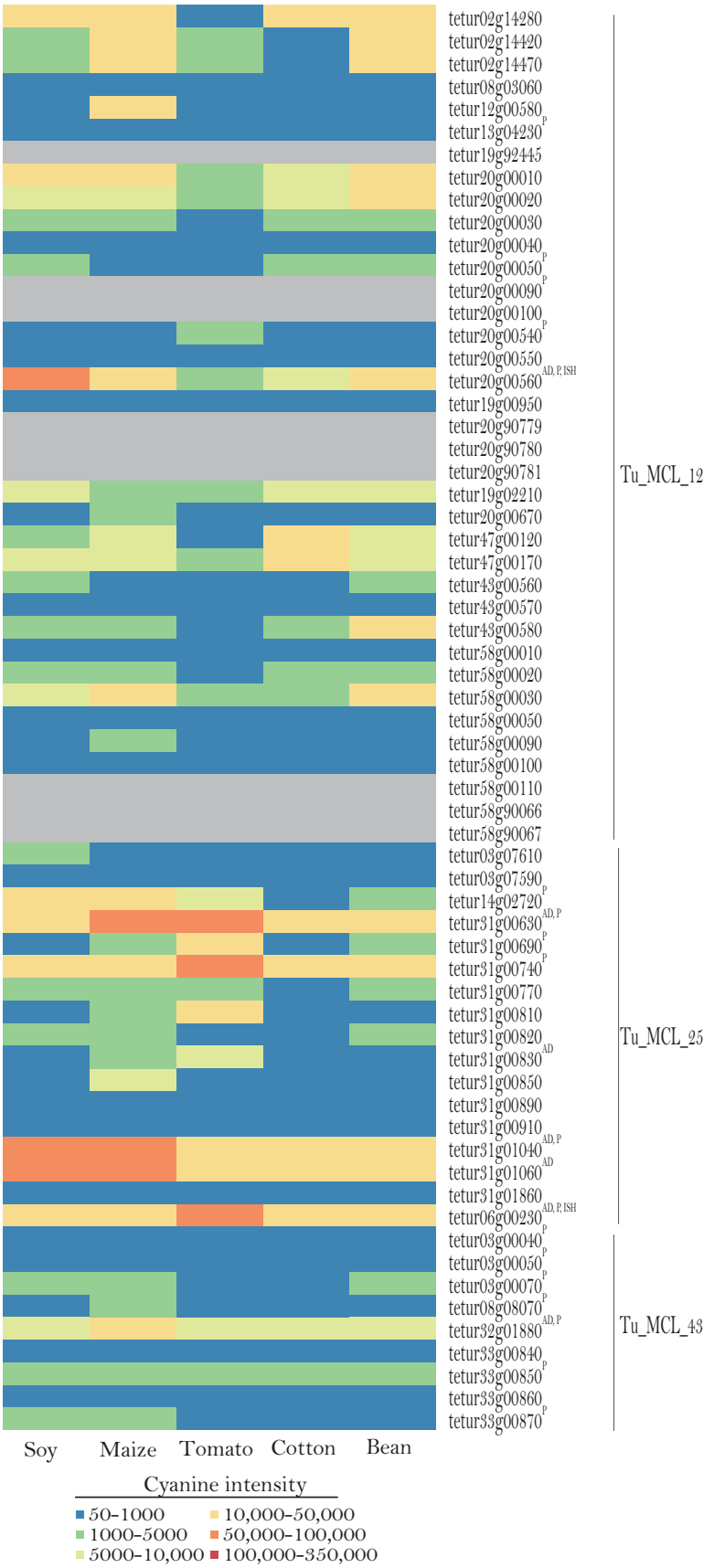


FIGURE 4.3. Expression profiles of genes coding for members of the expanded Tu_MCL_36 and Tu_MCL_35 protein families in *Tetranychus urticae*. Phylogenetic analyses of Tu_MCL_36 and Tu_MCL_35 proteins are shown next to heatmaps depicting relative gene expression of Tu_MCL_35 and Tu_MCL_36 genes in *T. urticae* subjected to different host plant regimes (log_2FCs, relative to the expression level on bean). The heatmap on the left represents log_2FCs of Tu_MCL_35 and Tu_MCL_36 genes in mites adapted to soy, maize, cotton or tomato compared to bean, while the heatmap on the right represents log_2FCs of Tu_MCL_35 and Tu_MCL_36 genes in *T. urticae* after host shifts between bean and tomato (B_24hT, mites from the London reference strain on bean transferred to tomato for 24h; B_30GT, mites from the bean strain grown on tomato for 30 generations; B_30GT_2GB, mites from the bean strain grown on tomato for 30 generations and transferred back to bean for two generations; Wybouw et al., 2015). Grey boxes indicate that for a specific gene no probes were included in the *T. urticae* microarray design, and hence expression could not be assayed.


tetur02g14280
tetur02g14420
tetur02g14470
tetur08g03060
tetur12g00580P
tetur13g04230P
tetur19g92445
tetur20g00010
tetur20g00020
tetur20g00030
tetur20g00040P
tetur20g00050P
tetur20g00090
tetur20g00100P
tetur20g00540P
tetur20g00550
tetur20g00560AD, P, ISH
tetur19g00950
tetur20g90779
tetur20g90780
tetur20g90781
tetur19g02210
tetur20g00670
tetur47g00120
tetur47g00170
tetur43g00560
tetur43g00570
tetur43g00580
tetur58g00010
tetur58g00020
tetur58g00030
tetur58g00050
tetur58g00090
tetur58g00100
tetur58g00110
tetur58g90066
tetur58g90067
Tu_MCL_12
tetur03g07610
tetur03g07590P
tetur14g02720P
tetur31g00630AD, P
tetur31g00690P
tetur31g00740P
tetur31g00770
tetur31g00810
tetur31g00820AD
tetur31g00830AD
tetur31g00850
tetur31g00890
tetur31g00910
tetur31g01040AD, P
tetur31g01060AD
tetur31g01860
tetur06g00230AD, P, ISH
Tu_MCL_25
tetur03g00040P
tetur03g00050P
tetur03g00070P
tetur08g08070P
tetur32g01880AD, P
tetur33g00840P
tetur33g00850
tetur33g00860P
tetur33g00870P
Tu_MCL_43
Soy
Maize
Tomato
Cotton
Bean
Cyanine intensity
50-1000
1000-5000
5000-10,000
10,000-50,000
50,000-100,000
100,000-350,000

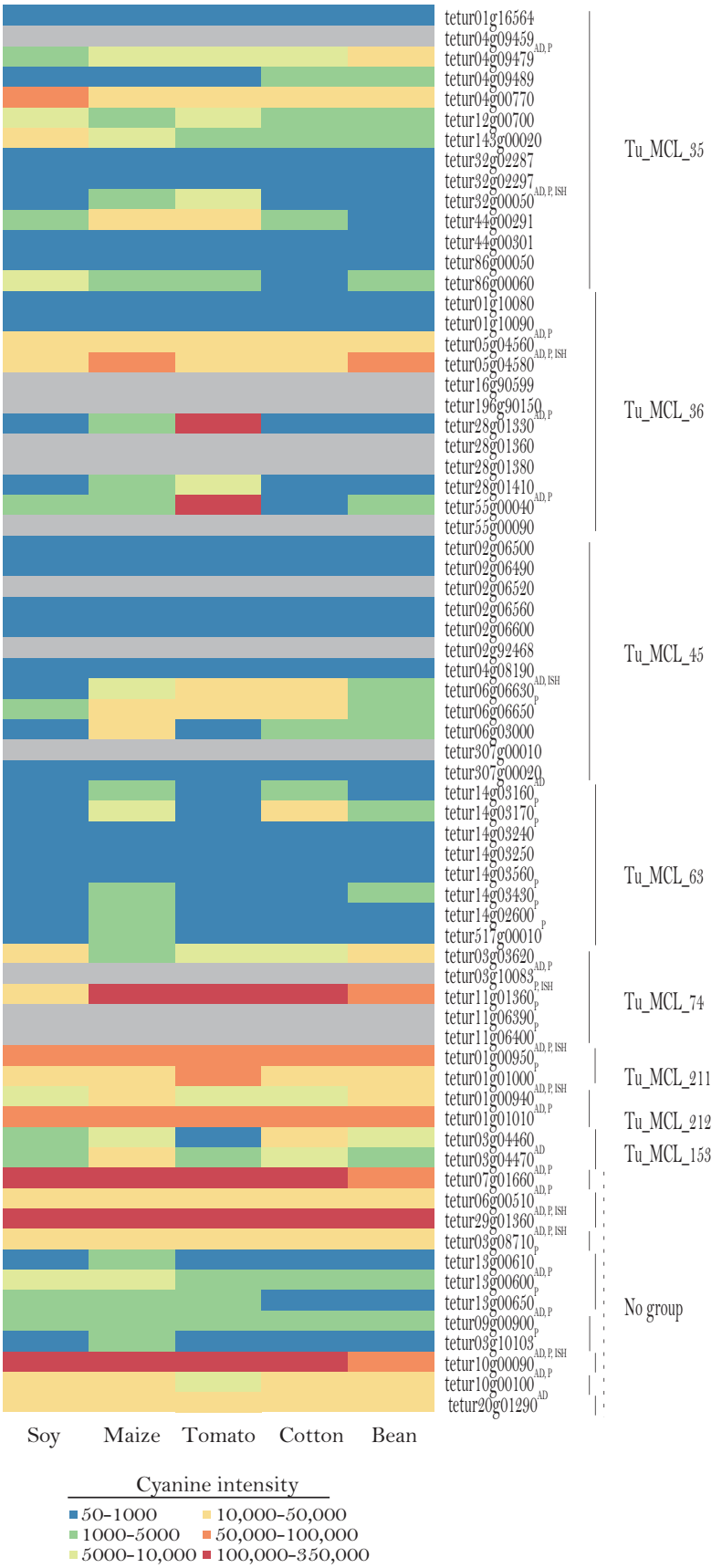

tetur01g16564
tetur04g09459 AD, P
tetur04g09479
tetur04g09489
tetur04g00770
tetur12g00700
tetur143g00020
tetur32g02287
tetur32g02297 AD, P, ISH
tetur32g00050
tetur44g00291
tetur44g00301
tetur86g00050
tetur86g00060
tetur01g10080
tetur01g10090 AD, P
tetur05g04560 AD, P, ISH
tetur05g04580
tetur16g90599
tetur196g90150 AD, P
tetur28g01330
tetur28g01360
tetur28g01380
tetur28g01410 AD, P
tetur55g00040
tetur55g00090
tetur02g06500
tetur02g06490
tetur02g06520
tetur02g06560
tetur02g06600
tetur02g92468
tetur04g08190 AD, ISH
tetur06g06630 P
tetur06g06650
tetur06g03000
tetur307g00010
tetur307g00020 AD
tetur14g03160 P
tetur14g03170 P
tetur14g03240
tetur14g03250
tetur14g03560 P
tetur14g03430 P
tetur14g02600 P
tetur517g00010
tetur03g03620 AD, P
tetur03g10083 P, ISH
tetur11g01360 P
tetur11g06390 P
tetur11g06400 AD, P, ISH
tetur01g00950 P
tetur01g01000 AD, P, ISH
tetur01g00940 AD, P
tetur01g01010
tetur03g04460 AD
tetur03g04470 AD, P
tetur07g01660 AD, P
tetur06g00510 AD, P, ISH
tetur29g01360 AD, P, ISH
tetur03g08710 P
tetur13g00610 AD, P
tetur13g00600 P
tetur13g00650 AD, P
tetur09g00900 P
tetur03g10103 AD, P, ISH
tetur10g00090 AD, P
tetur10g00100
tetur20g01290 AD
Tu_MCL_35
Tu_MCL_36
Tu_MCL_45
Tu_MCL_63
Tu_MCL_74
Tu_MCL_211
Tu_MCL_212
Tu_MCL_153
No group
Soy
Maize
Tomato
Cotton
Bean
Cyanine intensity
50-1000
1000-5000
5000-10,000
10,000-50,000
50,000-100,000
100,000-350,000

FIGURE 4.4 Heatmap of the absolute expression levels of a selection of genes of *Tetranychus urticae* adapted to different host plants. The selection consists of homologs of spider mite-specific genes encoding putative salivary proteins identified in artificial diet after feeding by *T. urticae*. The genes are grouped in 'Tu_MCL' clusters. The absolute expression level is shown as cyanine intensity, while genes for which no expression data are available are depicted as gray boxes. Superscripts supply additional info: ^{AD} protein identified in artificial diet fed upon by *T. urticae*; ^{P} expression level is significantly higher (FC≥8, adj. p≤0.05) in the proterosoma versus the entire body of *T. urticae* adapted to bean plants; ^{ISH} gene confirmed to be expressed in the salivary glands using whole-mount *in situ* hybridizations.

Tu_MCL_45 has representatives in *T. urticae*, yet not in *T. evansi* and *T. lintearius* (TABLES 4.2 and S4.2). Protein tetur06g06630 was detected in artificial diet fed upon by *T. urticae*, while its encoding gene was shown to be expressed in the salivary glands (**Chapter 3**). Only *tetur06g06650* appears to be significantly upregulated in the proterosoma (Benjamini-Hochberg adjusted p≤0.05 and FC ≥8).

Tu_MCL_63 has eight members in *T. urticae*, five of which are upregulated in the proterosoma (Benjamini-Hochberg adjusted p≤0.05 and FC ≥8). No representatives could be found in *T. evansi*, while two were present in *T. lintearius* (TABLES 4.2 and S4.2).

Proteins belonging to the spider mite-specific group **Tu_MCL_74** are homologous to proteins belonging to OG5_144177 and are further discussed in **Chapter 5 (**SHOT family**)** (TABLES 4.2 and S4.2).

The other **Tu_MCL** groups (TABLES 4.2 and S4.2, Figure 4.4) include just a few proteins each, of which several were confirmed to be produced in the salivary glands using whole-mount *in situ* hybridizations (**Chapter 3**). Most of the genes coding for these proteins are significantly upregulated in the proterosoma. Many lack homologs in *T. evansi* or *T. lintearius* (TABLES 4.2 and S4.2).

4.4. DISCUSSION

4.4.1. The salivary protein composition of *T. urticae* is host plant-dependent

A host plant-dependent *T. urticae* salivary protein composition was evident from the proteomics data in **Chapter 3** and was studied in more detail in this chapter. To this purpose, genome wide expression analyses were per-

formed on the mite lines used for the salivary proteome study. Additionally, cotton-adapted lines were also investigated.

Secreting the full salivary protein repertoire, independent of the host plant, might come with an enhanced risk that some salivary proteins are perceived as elicitors. In addition, it is unlikely that all proteins will function optimally in all hosts as targets may differ across plants or may be absent. Therefore it would be beneficial for polyphagous herbivores to alter the composition of their saliva according to the nature of the host plant. Differences between the salivary gland transcriptomes of two populations of *Nilaparvata lugens*, a rice specialist, maintained on either a resistant or susceptible rice variety, were proposed to be related to different virulence traits of these brown planthoppers (Ji et al., 2013; Wang et al., 2015). Furthermore, drastically different salivary protein profiles were found when the western tarnished plant bug *Lygus hesperus* was fed artificial diet, cotton or pinto bean (Habibi et al., 2001). Hence, we investigated host plant specific secretion and expression of the mite's salivary genes. As suggested by the proteomics data (FIGURE 4.1), transcriptome analysis revealed that the expression of *T. urticae* salivary protein genes is strongly influenced by the host plant species on which the mites had been feeding (FIGURE 4.2). For example, proteins tetur28g01330 and tetur55g00040 were uniquely and abundantly identified from diet fed upon by tomato-adapted mites (FIGURE 4.1), while expression of the corresponding gene was also extremely high when feeding on tomato, relative to mites feeding on the other tested plants (FIGURE 4.2). Next to individual genes, we also studied OrthoMCL gene family groups. Tu_MCL_25, Tu_MCL_35 and Tu_MCL_36 (TABLES 3.1 and S3.7) all have members that are highly expressed when feeding on tomato, while the expression of other members of these groups was not influenced by the host plant (FIGURES 4.2 and 4.3). Future experiments should point out whether this is due to the fact that mites from the tomato-adapted lines had been maintained on tomato for many more generations than mites on the other hosts or whether this is due to the specific allelochemicals of tomato posing digestive or defensive challenges. Members of the OrthoMCL group OG5_144177 (SHOT family, **Chapter 5**) (TABLES 3.1 and S3.7) were, relative to mites on bean, expressed at lower levels in mites feeding from maize, cotton and tomato (FIGURE 4.2). When feeding on soy, however, these genes were expressed at slightly higher levels. A similar pattern can also be deduced from the proteomics data of mites adapted to bean and

soy versus mites adapted to maize and tomato. Bean and soy both are legume species (Fabaceae) and mites feeding on these plants probably encounter analogous plant secondary compounds that select for or induce a similar repertoire of salivary proteins. The observation that different host plant species can differentially affect expression levels of genes coding for salivary proteins – such as effectors – has been reported for aphids as well (Elzinga et al., 2014; Pan et al., 2015; Wang et al., 2015). Elzinga et al. (2014) suggested that the differential expression of salivary effector genes represents a strategy to avoid activation of defenses and to facilitate feeding.

In the scope of this discussion it is important to realize that only one *T. urticae* strain (London) was used. Although mites from this strain had been maintained on bean for many years, previous studies have shown that this population is both not fully inbred (Van Leeuwen et al., 2012) and capable of extensive transcriptional plasticity upon transfer to new hosts (Grbić et al., 2011; Dermauw et al., 2013; Wybouw et al., 2014; Zhurov et al., 2014), the latter of which has been further confirmed in this study. We believe therefore that this study has captured much of the repertoire of *T. urticae*'s biologically relevant salivary proteins. However, as marked variation of genotypes exists between *T. urticae* populations and across spider mite species (Kant et al., 2008; Alba et al., 2015), the use of additional *T. urticae* strains or different spider mite species may deliver additional salivary proteins. Identification via shotgun proteomics will, however, be less straightforward without reference genomes for these strains or species.

Due to the highly specific function of effectors, variable expression levels depending on the host plant could be indicative of such 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 (Stuart et al., 2012; Kanvil et al., 2014; Stuart, 2015), as does the presence of S genes (van Schie and Takken, 2014), the target of the effector. As such, the specific R and S gene composition of a specific host may determine the specific transcriptional response of a herbivore's secretome and thus its effector repertoire.

4.4.2. A glimpse into the salivary protein composition of *T. evansi* and *T. lintearius*

Tetranychus urticae and *T. lintearius* are closely related species which are more distantly related to *T. evansi* (Toni Gabaldon, personal communication

in Cao, 2014). Despite the close phylogenetic relationship of these three spider mites, their host ranges differ dramatically. *Tetranychus lintearius* is a specialist on *U. europaeus* (common gorse, Fabaceae), while the polyphagous *T. urticae* has been reported to feed on plants belonging to 140 different families (Ireson et al., 2003; Migeon and Dorkeld, 2006-2017). *Tetranychus evansi* is an oligophagous mite which is specialized in feeding on solanacean host plants such as tomato, although it has been recorded on other plant families as well (Migeon and Dorkeld, 2006-2017; Navajas et al., 2013). With the annotated genomes of these three species available to our research group, the occurrence of homologs of genes coding for (putative) *T. urticae* salivary proteins (**Chapter 3**, TABLE 3.1) in the genomes of *T. lintearius*, *T. evansi* and *T. urticae* could be investigated. It is important to keep in mind that proteins identified in the secreted saliva of *T. urticae* were used as a query. As such, 'new' salivary protein families, unique to *T. lintearius* and/or *T. evansi*, could not be identified, and we can only compare the number of members within the query families. With an efficient proteomics approach to identify spider mite salivary proteins reported in this PhD thesis (**Chapter 3**), it should be straightforward to determine the salivary proteome of *T. lintearius* and *T. evansi* as well, particularly because their genomes will be publically available soon. Encouragingly, preliminary proteomics experiments with *T. evansi* saliva have proven to be successful (data not shown). Not all NanoLC-MS/MS identified proteins used as search query (**Chapter 3**, TABLE 3.1) are true salivary proteins as some contamination is present. The proterosoma fold change (expression relative to the entire mite body) is an important criterion to evaluate whether these proteins are produced in the salivary glands or not. Of further notice, the proterosomal transcriptome (**Chapter 3**, DATA S3.9) was constructed using a *T. urticae* line (London) adapted to bean (*P. vulgaris*), and as such, these relative gene expression data may not be representative for genes with a strong host-dependent expression level (e.g., a gene that has a low expression level when feeding on bean will have a relatively low proterosoma to body expression ratio, which might have been high if a tomato-adapted line was used to generate the proterosoma transcriptome). Furthermore, proteins likely to be contamination – based on the proterosoma FC – were not omitted from TABLE 3.1 because these proteins, although contamination from, e.g., gut and cuticle, may still induce defenses when they end up in the plant tissue. As such, they may be relevant for other studies.

The identification of salivary protein orthologs among spider mite species is an important step in determining the protein's function during the plant-mite interaction. In aphids, for example, the functionality of *Myzus persicae* and *Acyrthosiphon pisum* orthologs of the effectors c002, PIntO1 (MP1) and PIntO2 has been evaluated (Pitino and Hogenhout, 2013). *Myzus persicae* is polyphagous and can count *Arabidopsis* amongst its host plants, while *A. pisum* specializes on colonizing fabacean hosts. Transgenic *Arabidopsis* lines producing the *M. persicae* orthologs of latter three effectors were found to increase the reproduction of *M. persicae* feeding on them. Reproduction of this aphid species did, however, not increase when they fed on *Arabidopsis* lines expressing the *A. pisum* orthologs of *C002*, *PIntO1* or *PIntO2* (Pitino and Hogenhout, 2013). This suggests that effector orthologs have evolved to enhance compatibility between the herbivore and a specific group of host plant species. This adaptation may, however, compromise the effector function in other plant species, potentially excluding them as hosts. Analogously, the host plant range of *T. urticae*, *T. lintearius* and *T. evansi* may be mirrored in the diversity of their effectors. For example, proteins encoded by the genes *tetur28g01330* and *tetur55g00040* (Tu_MCL_36) were abundantly and uniquely identified in salivary samples from tomato-adapted *T. urticae* lines. This was accompanied by a high gene expression level. The Solanaceae specialist *T. evansi* possesses four Tu_MCL_36 genes, while in the *Ulex* specialist *T. lintearius*, homologs were absent. This may be indicative for a Solanaceae-specific function of (most) Tu_MCL_36 members. However, a more detailed study is clearly needed to vouch for this hypothesis. Another example (Tu_MCL_74 & OG5_144177, 'SHOT' family) of host-dependent expression in *T. urticae*, and presence or absence of homologs in closely related spider mites species with a limited host range is studied in more detail in **Chapter 5**.

The comparative OrthoMCL study revealed that *T. urticae* possesses most proteins (436 proteins) homologous to proteins identified in artificial diet fed upon by *T. urticae*. Further, we identified 165 homologs in *T. evansi* and 159 in the closely related *T. lintearius*. As such, there is a correlation between the number of recorded host plants and the (predicted) number of putative salivary proteins. Whether this represents a causal relationship remains to be elucidated. But it would not come as a surprise that salivary protein homologs provide functional variability, allowing interaction with

components of distinct host plants. Alternatively, qualitative difference between the annotated genomes of the three mite species may be at the basis of the difference between the identified number of homologs. Indeed, the genome of *T. urticae* was sequenced using whole genome shotgun sequencing with Sanger technology (Grbić et al., 2011), while the genomes of *T. lintearius* and *T. evansi* were sequenced using Illumina technology (2012) (Cao, 2014). However, the EuGene gene prediction platform (Foissac et al., 2008) was used for genome annotation of all three species, using identical parameters. Furthermore, the size of the genomes and proteomes of all three mite species appeared to be about the same (Cao, 2014). So, despite the fact that the proteomes were not obtained by an identical procedure, a comparison seemed to be justified. In addition, a comparative analysis on the gustatory receptors of *T. urticae, T. lintearius* and *T. evansi* was based on the same genomic databases used in current study. It revealed an analogous correlation between the size of the host range and the number of these chemoreceptors, which have a function in the detection of chemical components in their environment, including those of plants (Ngoc et al., 2016).

Interestingly, genes encoding (predicted) salivary proteins belonging to the same (expanded) OrthoMCL cluster are often in close proximity of each other on the genome scaffold [e.g., *tetur20g00010-tetur20g00020-tetur20g00030-tetur20g00040-tetur20g00050* (Tu_MCL_12) and *tetli84g01040-tetli84g01050-tetli84g01060-tetli84g01070* (Tu_MCL_43)]. This may be a signature of (tandem) gene duplication followed by diversification, showing the effect of diversifying selection (Chen et al., 2010; Johnson et al., 2015; Zhao et al., 2015). The ubiquitous proliferations in salivary protein families in *T. urticae* may correlate with the polyphagous nature of this pest. One member of a family may for example be active in one host plant or group of related plants, while another member is active in another host plant (group). The few salivary protein families that are expanded in *T. evansi* and *T. lintearius* may reflect proliferation of a salivary protein family that is advantageous or even essential when feeding on the host plant group these mites have specialized to feed on. However, additional research is needed. Next to the identification of protein family expansions, an additional source of information about the ecological adaptation of the three spider mite species to their host plant range may be to look at pseudogenisation events (Ngoc et al., 2016). These pseudogenes are relics of former genes that no longer possess biological functions and can be

regarded as fossilized footprints of past gene expression. Most pseudogenisation events occur through random accumulation of mutations in genes on which the functional constraints are relaxed (Podlaha and Zhang, 2010). For example, *T. lintearius* may have lost salivary protein members since its divergence from the common ancestor with *T. urticae* because these genes are no longer needed due to the reduced host arsenal. An in-depth study of one family (SHOT) of salivary proteins (**Chapter 5**) further supports this hypothesis and we conclude that a spider mite's salivary protein repertoire is a reflection of its host plant range.

4.5. Conclusions

The saliva of *T. urticae* contains many proteins which appear to be spider mite specific and as such are likely to have a unique function. Several of these proteins are encoded by genes belonging to expanded families. In addition, the expression level of some of these putative salivary protein genes is influenced by the plant species to which the mite lines were adapted. This host dependent expression may be an indication of a host plant specific mode of action, which may be characteristic for effectors. Furthermore, *in silico* prediction of *T. evansi* and *T. lintearius* homologs of salivary proteins identified in *T. urticae* has shed some light on the salivary proteome of spider mite species with a more limited host range. It further supports our hypothesis that the host range of spider mites is linked to the salivary protein composition, and as such, expansions in *T. urticae* salivary protein gene families are likely key to *T. urticae*'s polyphagous nature.

Funding information

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.

4.6. References

Alba, J. M., B. C. J. Schimmel, J. J. Glas, L. M. S. Ataide, M. L. Pappas, C. A. Villarroel, R. C. Schuurink, M. W. Sabelis and M. R. Kant (2015). 'Spider mites suppress tomato defenses downstream of jasmonate and salicylate independently of hormonal crosstalk.' *New Phytologist* **205**(2): 828-840.

Cao, T. N. P. (2014). *Genome annotation and evolution of chemosensory receptors in spider mites*, Ghent University.

Chen, F., A. J. Mackey, C. J. Stoeckert and D. S. Roos (2006). 'OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups.' *Nucleic Acids Research* **34**(suppl 1): D363-D368.

Chen, M.-S., X. Liu, Z. Yang, H. Zhao, R. H. Shukle, J. J. Stuart and S. Hulbert (2010). 'Unusual conservation among genes encoding small secreted salivary gland proteins from a gall midge.' *BMC evolutionary biology* **10**(1): 296.

Dermauw, W., E. J. Osborne, R. M. Clark, M. Grbić, L. Tirry and T. Van Leeuwen (2013). 'A burst of ABC genes in the genome of the polyphagous spider mite *Tetranychus urticae*.' *BMC Genomics* **14**(1): 1-22.

Dermauw, W., N. Wybouw, S. Rombauts, B. Menten, J. Vontas, M. Grbić, R. M. Clark, R. Feyereisen and T. Van Leeuwen (2013). 'A link between host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus urticae*.' *Proceedings of the National Academy of Sciences of the United States of America* **110**(2): E113-E122.

Elzinga, D. A., M. De Vos and G. Jander (2014). 'Suppression of plant defenses by a *Myzus persicae* (green peach aphid) salivary effector protein.' *Molecular Plant-Microbe Interactions* **27**(7): 747-756.

Foissac, S., J. Gouzy, S. Rombauts, C. Mathé, J. Amselem, L. Sterck, Y. V. de Peer, P. Rouzé and T. Schiex (2008). 'Genome annotation in plants and fungi: EuGene as a model platform.' *Current Bioinformatics* **3**(2): 87-97.

Grbić, M., T. Van Leeuwen, R. M. Clark, S. Rombauts, P. Rouzé, V. Grbić, E. J. Osborne, W. Dermauw, P. C. T. Ngoc and F. Ortego (2011). 'The genome of *Tetranychus urticae* reveals herbivorous pest adaptations.' *Nature* **479**.

Grbić, M., T. Van Leeuwen, R. M. Clark, S. Rombauts, P. Rouzé, V. Grbić, E. J. Osborne, W. Dermauw, P. C. T. Ngoc and F. Ortego (2011). 'The genome of *Tetranychus urticae* reveals herbivorous pest adaptations.' *Nature* **479**(7374): 487-492.

Habibi, J., E. A. Backus, T. A. Coudron and S. L. Brandt (2001). 'Effect of different host substrates on hemipteran salivary protein profiles.' *Entomologia Experimentalis et Applicata* **98**(3): 369-375.

Harris, M., T. Freeman, K. Anderson, J. Harmon, J. Moore, S. Payne, O. Rohfritsch and J. Stuart (2012). 'Hessian fly Avirulence gene loss of function defeats plant resistance without compromising the larva's ability to induce a gall tissue.' *Entomologia Experimentalis et Applicata* **145**(3): 238-249.

Ireson, J., A. Gourlay, R. Kwong, R. Holloway and W. Chatterton (2003). 'Host specificity, release, and establishment of the gorse spider mite, *Tetranychus lintearius* Dufour (Acarina: Tetranychidae), for the biological control of gorse, *Ulex europaeus* L.(Fabaceae), in Australia.' *Biological Control* **26**(2): 117-127.

Ji, R., H. Yu, Q. Fu, H. Chen, W. Ye, S. Li and Y. Lou (2013). 'Comparative Transcriptome Analysis of Salivary Glands of Two Populations of Rice Brown Planthopper, *Nilaparvata lugens*, That Differ in Virulence.' *PLoS One* **8**(11): e79612.

Johnson, A., R. Shukle, M. S. Chen, S. Srivastava, S. Subramanyam, B. Schemerhorn, P. Weintraub, H. Abdel Moniem, K. Flanders and G. Buntin (2015). 'Differential expression of candidate salivary effector proteins in field collections of Hessian fly, *Mayetiola destructor*.' *Insect molecular biology* **24**(2): 191-202.

Kaloshian, I. (2004). 'Gene-for-gene disease resistance: bridging insect pest and pathogen defense.' *Journal of Chemical Ecology* **30**(12): 2419-2438.

Kant, M. R., M. W. Sabelis, M. A. Haring and R. C. Schuurink (2008). 'Intraspecific variation in a generalist herbivore accounts for differential induction and impact of host plant defences.' *Proceedings of the Royal Society of London B Biological Sciences* **275**(1633): 443-452.

Kanvil, S., G. Powell and C. Turnbull (2014). 'Pea aphid biotype performance on diverse Medicago host genotypes indicates highly specific virulence and resistance functions.' *Bulletin of Entomological Research* **104**(06): 689-701.

Kauffmann, A., R. Gentleman and W. Huber (2009). 'arrayQualityMetrics - a bioconductor package for quality assessment of microarray data.' *Bioinformatics* **25**(3): 415-416.

Li, L., C. J. Stoeckert and D. S. Roos (2003). 'OrthoMCL: identification of ortholog groups for eukaryotic genomes.' *Genome research* **13**(9): 2178-2189.

Migeon, A. and F. Dorkeld. (2006-2017). 'Spider Mites Web: a comprehensive database for the Tetranychidae.' from http://www.montpellier.inra.fr/CBGP/spmweb.

Navajas, M., G. J. De Moraes, P. Auger and A. Migeon (2013). 'Review of the invasion of *Tetranychus evansi*: biology, colonization pathways, potential expansion and prospects for biological control.' *Experimental and Applied Acarology* **59**(1-2): 43-65.

Ngoc, P. C. T., R. Greenhalgh, W. Dermauw, S. Rombauts, S. Bajda, V. Zhurov, M. Grbić, Y. Van de Peer, T. Van Leeuwen, P. Rouzé and R. M. Clark (2016). 'Complex evolutionary dynamics of massively expanded chemosensory receptor families in an extreme generalist chelicerate herbivore.' *Genome Biology and Evolution* **8**(11): 3323-3339.

Pan, Y., J. Zhu, L. Luo, L. Kang and F. Cui (2015). 'High expression of a unique aphid protein in the salivary glands of *Acyrthosiphon pisum*.' *Physiological and Molecular Plant Pathology* **92**: 175-180.

Pitino, M. and S. A. Hogenhout (2013). 'Aphid protein effectors promote aphid colonization in a plant species-specific manner.' *Molecular Plant-Microbe Interactions* **26**(1): 130-139.

Podlaha, O. and J. Zhang (2010). 'Pseudogenes and their evolution.' *eLS*.

Ritchie, M. E., J. Silver, A. Oshlack, M. Holmes, D. Diyagama, A. Holloway and G. K. Smyth (2007). 'A comparison of background correction methods for two-colour microarrays.' *Bioinformatics* **23**(20): 2700-2707.

Sterck, L., K. Billiau, T. Abeel, P. Rouze and Y. Van de Peer (2012). 'ORCAE: online resource for community annotation of eukaryotes.' *Nature methods* **9**(11): 1041-1041.

Stuart, J. (2015). 'Insect effectors and gene-for-gene interactions with host plants.' *Current Opinion in Insect Science* **9**: 56–61.

Stuart, J. J., M.-S. Chen, R. Shukle and M. O. Harris (2012). 'Gall midges (Hessian flies) as plant pathogens.' *Annual review of phytopathology* **50**: 339-357.

Van Leeuwen, T., P. Demaeght, E. J. Osborne, W. Dermauw, S. Gohlke, R. Nauen, M. Grbić, L. Tirry, H. Merzendorfer and R. M. Clark (2012). 'Population bulk segregant mapping uncovers resistance mutations and the mode of action of a chitin synthesis inhibitor in arthropods.' *Proceedings of the National Academy of Sciences of the United States of America* **109**(12): 4407-4412.

Van Leeuwen, T. and W. Dermauw (2016). 'The Molecular Evolution of Xenobiotic Metabolism and Resistance in Chelicerate Mites.' *Annual Review of Entomology* **61**: 475–498.

van Schie, C. C. and F. L. Takken (2014). 'Susceptibility genes 101: how to be a good host.' *Annual Review of Phytopathology* **52**: 551-581.

Wang, W., H. Dai, Y. Zhang, R. Chandrasekar, L. Luo, Y. Hiromasa, C. Sheng, G. Peng, S. Chen and J. M. Tomich (2015). 'Armet is an effector protein mediating aphid–plant interactions.' *The FASEB Journal* **29**(5): 2032-2045.

Wang, X., M. Zhang, F. Feng and R. He (2015). 'Differentially regulated genes in the salivary glands of brown planthopper after feeding in resistant versus susceptible rice varieties.' *Archives of Insect Biochemistry and Physiology* **89**(2): 69-86.

Wybouw, N., W. Dermauw, L. Tirry, C. Stevens, M. Grbić, R. Feyereisen and T. Van Leeuwen (2014). 'A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning.' *eLife* **3**: e02365.

Wybouw, N., V. Zhurov, C. Martel, K. A. Bruinsma, F. Hendrickx, V. Grbić and T. Van Leeuwen (2015). 'Adaptation of a polyphagous herbivore to a novel host plant extensively shapes the transcriptome of herbivore and host.' *Molecular ecology* **24**(18): 4647-4663.

Wybouw, N., V. Zhurov, C. Martel, K. A. Bruinsma, F. Hendrickx, V. Grbić and T. Van Leeuwen (2015). Data from: Adaptation of a polyphagous herbivore to a novel host plant extensively shapes the transcriptome of herbivore and host, Dryad Data Repository.

Zhao, C., L. N. Escalante, H. Chen, T. R. Benatti, J. Qu, S. Chellapilla, R. M. Waterhouse, D. Wheeler, M. N. Andersson and R. Bao (2015). 'A massive expansion of effector genes underlies gall-formation in the wheat pest *Mayetiola destructor*.' *Current Biology* **25**(5): 613-620.

Zhurov, V., M. Navarro, K. A. Bruinsma, V. Arbona, M. E. Santamaria, M. Cazaux, N. Wybouw, E. J. Osborne, C. Ens and C. Rioja (2014). 'Reciprocal responses in the interaction between Arabidopsis and the cell-content-feeding chelicerate herbivore spider mite.' *Plant physiology* **164**(1): 384-399.

Table S4.1. *Tetranychus evansi*, *T. lintearius* and *T. urticae* proteins homologous to proteins identified in artificial diet fed upon by *T. urticae*. Only proteins which could be assigned to groups already existing in the OrthoMCL database were accounted for in this list.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
OG5_126560		tetev174g00270	tetli08g00930	tetur01g09300
		tetev38g02260	tetli38g00560	tetur02g10210
		tetev50g00620	tetli78g00410	tetur03g08060[P]
		tetev54g00330		tetur05g07800
		tetev59g00640		tetur12g01140
		tetev73g00390		tetur12g02880
				tetur24g01640
OG5_126583	IPR023210 (NADP-dependent oxido-reductase domain), IPR018170 (Aldo/keto reductase, conserved site), IPR020471 (Aldo/keto reductase)	tetev275g00070	tetli105g00100	tetur02g07450
		tetev33g00180	tetli154g00030	tetur02g07550
		tetev43g00610	tetli216g00070	tetur02g07590
		tetev43g00640	tetli231g00220	tetur02g11330
		tetev43g00670	tetli23g02180	tetur02g11340[AD,P]
		tetev43g00690	tetli333g00010	tetur02g11390
		tetev89g00430		tetur02g11410
				tetur02g11530
				tetur02g90723
				tetur30g00660
OG5_126595	IPR004000 (Actin family), IPR004001 (Actin, conserved site), IPR020902 (Actin/actin-like conserved site)	tetev109g00360	tetli04g00180	tetur03g09480[AD,P]
		tetev26g00150	tetli246g00020	tetur09g05350
				tetur09g05360
OG5_126607	IPR013201 (Cathepsin propeptide inhibitor domain; I29), IPR025660 (Cysteine peptidase, histidine active site), IPR025661 (Cysteine peptidase, asparagine active site), IPR000668	tetev02g00220	tetli09g00600	tetur05g92633
		tetev28g00730	tetli33g01670	tetur06g03520
		tetev321g00030	tetli344g00080	tetur09g04400[AD]
		tetev36g00820	tetli38g01500	tetur09g04420
		tetev36g00830	tetli38g01510	tetur09g04470

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
	(Peptidase C1A, papain C-terminal), IPR000169 (Cysteine peptidase, cysteine active site)	tetev36g00840	tetli38g01520	tetur123g00050
		tetev36g00850	tetli38g01530	tetur12g01810
		tetev36g00860	tetli38g01540	tetur12g01820
		tetev36g00870	tetli38g01550	tetur12g01830
		tetev36g00880		tetur12g01840
		tetev38g02400		tetur12g01850[P]
		tetev85g00370		tetur12g01860[AD]
				tetur12g04631
				tetur16g03680
				tetur16g03770
				tetur23g00050
				tetur23g00860
				tetur23g00880
				tetur23g00890
				tetur23g01290
				tetur25g00650[AD]
OG5_126738	IPR013783 (Immunoglobulin-like fold), IPR007110 (Immunoglobulin-like domain), IPR003961 (Fibronectin type III), IPR000719 (Protein kinase domain), IPR008271 (Serine/threonine-protein kinase, active site), IPR017441 (Protein kinase, ATP binding site), IPR013098 (Immunoglobulin I-set), IPR002290 (Serine/threonine/dual specificity protein kinase, catalytic domain), IPR003599 (Immunoglobulin subtype), IPR003598	tetev04g01160	tetli260g00310	tetur01g01810
		tetev101g00190	tetli260g00320	tetur01g01830
		tetev101g00220	tetli26g02110	tetur01g01850[AD]
		tetev124g00030	tetli26g02120	tetur04g02800
		tetev124g00040	tetli26g02130	tetur04g02880
		tetev169g00060	tetli302g00020	tetur05g03810
		tetev169g00210	tetli321g00010	tetur06g00290
		tetev169g00220	tetli46g00640	tetur114g00010
			tetli478g00030	tetur30g00590
			tetli69g00060	tetur31g01290

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
	(Immunoglobulin subtype 2), IPR011009 (Protein kinase-like domain)			
OG5_126942	IPR012336 (Thioredoxin-like fold), IPR010987 (Glutathione S-transferase, C-terminal-like), IPR004045 (Glutathione S-transferase, N-terminal), IPR004046 (Glutathione S-transferase, C-terminal)	tetev125g00360 tetev125g00410 tetev25g01600 tetev25g01620 tetev52g01110 tetev88g01280	tetli118g00150 tetli118g00180 tetli118g00190 tetli118g00200 tetli16g01140 tetli16g01220 tetli26g01650 tetli29g00100	tetur01g02230 tetur01g02470 tetur01g02480 tetur01g02490 tetur01g02500 tetur01g02510^{AD,P} tetur03g07920^{AD} tetur26g01450 tetur26g01460 tetur26g01490 tetur26g01500 tetur26g02801 tetur26g02802 tetur29g00220 tetur31g01330
OG5_127143	IPR013785 (Aldolase-type TIM barrel), IPR029768 (Fructose-bisphosphate aldolase class-I active site), IPR000741 (Fructose-bisphosphate aldolase, class-I)	tetev258g00130 tetev258g00160 tetev11g00600	tetli162g00220 tetli162g00240 tetli07g00150	tetur07g03440^{AD} tetur109g00010 tetur109g00020 tetur01g16010
OG5_127584	IPR001424 (Superoxide dismutase, copper/zinc binding domain), IPR018152 (Superoxide dismutase, copper/zinc, binding site)	tetev11g00760 tetev14g00420 tetev16g00390 tetev204g00040 tetev38g02250 tetev43g00880	tetli125g00040 tetli147g00210 tetli147g00500 tetli149g01060 tetli168g00290 tetli52g00820	tetur02g04630 tetur02g11180 tetur04g09509^{P} tetur07g03760 tetur09g01550 tetur26g02320^{AD}

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
		tetev44g01030	tetli80g00540	tetur26g02520
OG5_127620	IPR013780 (Glycosyl hydrolase, family 13, all-beta), IPR002241 (Glycoside hydrolase, family 27), IPR013785 (Aldolase-type TIM barrel), IPR000111 (Glycoside hydrolase family 27/36, conserved site), IPR017853 (Glycoside hydrolase super-family)	tetev05g00340	tetli02g00890	tetur02g04310[AD]
		tetev98g00090	tetli94g00190	tetur04g00130
OG5_127800	IPR000668 (Peptidase C1A, papain C-terminal)	tetev12g02780	tetli105g00130	tetur01g05230[AD]
		tetev176g00050	tetli179g00030	tetur01g05480
		tetev183g00220	tetli179g00040	tetur01g16463
		tetev32g00870	tetli179g00350	tetur01g16473
		tetev33g00190	tetli179g00580	tetur02g11420
		tetev407g00010	tetli256g00020	tetur03g07930
		tetev488g00020	tetli256g00230	tetur03g07950
		tetev650g00030	tetli256g00240	tetur03g08010
		tetev92g00890	tetli268g00120	tetur03g08020
			tetli48g01490	tetur03g08030[AD]
				tetur03g09997
				etur06g02570
				tetur06g02580[AD]
				tetur08g05010
				tetur08g05020
				tetur08g05030
				tetur08g05310
				tetur09g00570
				tetur09g00600

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
				tetur24g00270
				tetur24g00280
				tetur28g01340
				tetur28g01390
				tetur28g01400
				tetur28g01420
				tetur28g01430
				tetur55g00100
				tetur55g00120
				tetur55g00130[P]
OG5_128075	IPR002347 (Glucose/ribitol dehydrogenase), IPR002198 (Short-chain dehydrogenase/reductase SDR), IPR020904 (Short-chain dehydrogenase/reductase, conserved site), IPR016040 (NAD(P)-binding domain)	tetev61g00450 tetev61g00560	tetli84g00100 tetli84g00140	tetur32g01960 tetur32g02180 tetur32g02327[AD]
OG5_128163	IPR001944 (Glycoside hydrolase, family 35), IPR008979 (Galactose-binding domain-like), IPR017853 (Glycoside hydrolase superfamily), IPR019801 (Glycoside hydrolase, family 35, conserved site), IPR013781 (Glycoside hydrolase, catalytic domain), IPR031330 (Glycoside hydrolase 35, catalytic domain), IPR026283 (Beta-galactosidase 1-like)	tetev10g00930 tetev15g00040 tetev178g00370	tetli108g00380 tetli24g00700 tetli31g00290	tetur07g07380[AD] tetur09g05870 tetur13g04220 tetur13g04678
OG5_128170	IPR016040 (NAD(P)-binding domain), IPR002347 (Glucose/ribitol dehydroge-	tetev38g02000 tetev92g00590	tetli04g01370 tetli287g00100	tetur06g04880 tetur06g04890

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
	nase), IPR020904 (Short-chain dehydrogenase/reductase, conserved site), IPR002198 (Short-chain dehydrogenase/reductase SDR)		tetli29g01000	tetur06g04900 tetur06g04970 tetur06g04980 tetur06g04990 tetur06g05070 tetur06g05090 tetur07g07300 tetur07g07310 tetur106g00010 tetur106g00040 tetur13g04450 tetur28g01070 tetur28g01140 tetur28g01710 tetur28g01720^{AD} tetur28g01730 tetur28g01740 tetur29g00870^{P}
OG5_129300	IPR013781 (Glycoside hydrolase, catalytic domain), IPR008979 (Galactose-binding domain-like), IPR006104 (Glycosyl hydrolases family 2, sugar binding domain), IPR017853 (Glycoside hydrolase superfamily)	tetev16g01180 tetev16g01190 tetev69g00510	tetli94g00590	tetur02g02400 tetur16g03420^{AD,P} tetur16g03430^{P} tetur28g00360^{AD,ISH,P}
OG5_129423	IPR013780 (Glycosyl hydrolase, family 13, all-beta), IPR001139 (Glycoside hydrolase, family 30), IPR013781 (Glycoside hydrolase, catalytic domain), IPR017853 (Glyco-	tetev10g00050 tetev193g00310 tetev213g00090 tetev246g00020	tetli112g00490 tetli112g00550 tetli134g00100 tetli164g00130	tetur02g07230 tetur02g07250 tetur02g07770 tetur02g07780

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
	side hydrolase superfamily)	tetev291g00040	tetli168g00350	tetur02g10940
		tetev299g00050	tetli16g01210	tetur02g12930^{AD}
		tetev299g00100	tetli186g00110	tetur02g90714
		tetev370g00030	tetli236g00110	tetur04g00790
		tetev39g00060	tetli237g00180	tetur05g06960
		tetev52g00390	tetli41g00220	tetur05g07000
		tetev59g00540	tetli41g00260	tetur06g01570
		tetev689g00010	tetli70g01270	tetur06g01580
		tetev689g00020	tetli87g00560	tetur06g01590
		tetev85g00380	tetli95g00230	tetur06g01660
		tetev89g00470		tetur06g01690
				tetur06g01700
				tetur09g06680^{P}
				tetur147g00020^{AD}
				tetur147g00030
				tetur147g00040
				tetur17g00790
				tetur17g00830
				tetur17g00850
				tetur17g00870
				tetur187g00010
				tetur187g00020
				tetur24g02260^{P}
				tetur30g02260
				tetur31g01370
				tetur32g00060
				tetur33g01260

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
				tetur33g01270
				tetur33g01290
				tetur72g00010[P]
				tetur72g00020[P]
OG5_130246	IPR029058 (Alpha/Beta hydrolase fold), IPR008758 (Peptidase S28)	tetev75g00330	tetli03g01710	tetur19g01900[AD]
OG5_130527	IPR012674 (Calycin), IPR022271 (Lipocalin, ApoD type), IPR011038 (Calycin-like), IPR000566 (Lipocalin/cytosolic fatty-acid binding domain), IPR003057 (Invertebrate colouration protein)	tetev01g00490	tetli07g01010	tetur01g01510
		tetev06g00100	tetli07g01020	tetur01g05730
		tetev124g00310	tetli102g00300	tetur01g05740
		tetev124g00320	tetli112g00400	tetur01g05750
		tetev124g00340	tetli139g00390	tetur01g05770
		tetev147g00020	tetli13g00660	tetur01g05880
		tetev14g00610	tetli145g00130	tetur04g05970
		tetev14g00630	tetli145g00300	tetur04g05980[AD]
		tetev14g00640	tetli208g00170	tetur04g06000
		tetev164g00040	tetli358g00010	tetur04g06010
		tetev164g00070	tetli41g00210	tetur05g07070
		tetev210g00200	tetli80g00390	tetur06g01640[AD]
		tetev210g00210	tetli80g00400	tetur06g02130
		tetev210g00220		tetur06g02140
		tetev299g00070		tetur06g02670
		tetev52g00260		tetur06g03020
		tetev83g00450		tetur06g03090
				tetur06g03100
				tetur06g03340
				tetur06g03350
				tetur06g03360

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
				tetur06g03370
				tetur06g03440
				tetur06g03550
				tetur06g03860[P]
				tetur06g06691
				tetur06g91377
				tetur06g91385
				tetur07g03970
				tetur09g04720
				tetur09g04730
				tetur09g04920
				tetur11g05210
				tetur11g05230
				tetur16g03410
				tetur16g03450
				tetur16g03460
				tetur174g00050
				tetur18g00900
				tetur282g00020
				tetur30g01430
				tetur31g00670
				tetur31g00680
				tetur31g00710
				tetur31g00730
				tetur31g00780
				tetur31g00800
				tetur31g00880

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
				tetur31g00900
				tetur31g00920
				tetur31g01880
				tetur37g00940[P]
OG5_131746	IPR016040 (NAD(P)-binding domain)	tetev116g00460	tetli15g00980	tetur01g12240
		tetev22g00600	tetli33g01300	tetur23g01640[AD]
			tetli93g00560	
OG5_132251	IPR001846 (von Willebrand factor, type D domain)		tetli04g00850	tetur07g05480
			tetli04g00860	tetur39g00700
			tetli07g01670	tetur39g00720
			tetli07g01700	tetur39g00810[AD]
			tetli1323g00010	tetur43g00010[AD]
			tetli571g00010	tetur516g00020[AD]
OG5_133467	IPR003172 (MD-2-related lipid-recognition domain), IPR014756 (Immunoglobulin E-set)	tetev107g00370	tetli01g01890	tetur03g01850
		tetev107g00380	tetli132g00590	tetur04g08680
		tetev107g00390	tetli16g00920	tetur07g02990
		tetev107g00400	tetli25g00680	tetur13g92931
		tetev13g01150	tetli318g00040	tetur140g00020
		tetev25g01430	tetli51g00500	tetur14g01690
		tetev52g01260	tetli85g00410	tetur14g01700
		tetev95g00110		tetur14g02060
				tetur14g02070[AD]
				tetur14g02080[AD]
				tetur14g02090[AD]
				tetur17g03530
				tetur17g03550
				tetur28g01370

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
				tetur55g00080
OG5_134456	IPR002223 (Pancreatic trypsin inhibitor Kunitz domain), IPR020901 (Proteinase inhibitor I2, Kunitz, conserved site)	tetev232g00090	tetli06g00480	tetur13g03820[AD]
OG5_135950	IPR001254 (Serine proteases, trypsin domain), IPR009003 (Peptidase S1, PA clan), IPR018114 (Peptidase S1, trypsin family, active site), IPR001314 (Peptidase S1A, chymotrypsin-type)	tetev13g01730 tetev30g00470 tetev83g00070	tetli06g01170 tetli115g00180 tetli154g00160 tetli28g00930 tetli79g00510	tetur06g06761[P] tetur07g00150[AD,ISH,P] tetur09g00280[P] tetur35g00910 tetur35g01000
OG5_136350	IPR019826 (Carboxylesterase type B, active site), IPR029058 (Alpha/Beta hydrolase fold), IPR002018 (Carboxylesterase, type B)	tetev492g00030 tetev56g01620 tetev59g00690 tetev97g00030	tetli72g00640 tetli72g00650 tetli72g00720	tetur02g09310 tetur02g14551 tetur02g15189 tetur11g01500[AD] tetur207g00020 tetur29g00950
OG5_141111	IPR001254 (Serine proteases, trypsin domain), IPR009003 (Peptidase S1, PA clan)	tetev136g00120 tetev18g00990 tetev194g00150	tetli10g00400 tetli51g00300	tetur13g00200 tetur16g03190[AD,P]
OG5_144177	'Secreted host responsive protein in Tetranychidae' (SHOT) family (Chapter 5)			tetur03g03670[AD,P] tetur03g03680[AD,P] tetur03g03690[ISH,P] tetur03g03700[AD,ISH,P] tetur03g03730[P] tetur03g10093[P]
OG5_147492	IPR000010 (Cystatin domain)	tetev08g00660 tetev08g00670	tetli144g00040 tetli145g00170	tetur06g01060[AD] tetur06g06610

Table S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
		tetev100g00570	tetli37g00030	tetur09g03620[AD]
		tetev116g00510	tetli38g00430	tetur09g03650
		tetev292g00240		tetur09g04770
OG5_149533	IPR009003 (Peptidase S1, PA clan), IPR001254 (Serine proteases, trypsin domain), IPR001314 (Peptidase S1A, chymotrypsin-type), IPR018114 (Peptidase S1, trypsin family, active site)	tetev37g01130	tetli18g00950	tetur05g08340
		tetev37g01260	tetli50g00280	tetur07g08034[AD]
		tetev49g00470		tetur35g00970
				tetur35g01050
				tetur35g01250
				tetur35g01260
				tetur35g01270
				tetur35g01310
				tetur35g01540
				tetur35g01600
				tetur35g01610
				tetur35g01630
				tetur441g00030
OG5_152237	IPR009003 (Peptidase S1, PA clan), IPR001254 (Serine proteases, trypsin domain)		tetli51g00220	tetur07g03160
				tetur16g03470[AD,P]
OG5_152337	IPR009003 (Peptidase S1, PA clan), IPR018114 (Peptidase S1, trypsin family, active site), IPR001254 (Serine proteases, trypsin domain), IPR001314 (Peptidase S1A, chymotrypsin-type)	tetev229g00140	tetli208g00160	tetur30g01440[AD]
		tetev38g00570		
		tetev83g00100		
OG5_152454	IPR002557 (Chitin binding domain), IPR029070 (Chitinase insertion domain), IPR011583 (Chitinase II), IPR001579 (Glycoside hydrolase, chitinase active site),	tetev08g01790	tetli179g00520	tetur01g11910[AD,P]
		tetev22g00350	tetli184g00230	tetur08g05470
		tetev31g00080	tetli21g00780	tetur11g03440

TABLE S4.1. Continued.

OrthoMCL	Functional description	*T. evansi* protein ID	*T. lintearius* protein ID	*T. urticae* protein ID
	IPR013781 (Glycoside hydrolase, catalytic domain), IPR017853 (Glycoside hydrolase superfamily), IPR001223 (Glycoside hydrolase family 18, catalytic domain)			
OG5_158831	IPR003172 (MD-2-related lipid-recognition domain), IPR014756 (Immunoglobulin E-set)			tetur01g02650 tetur01g02670[AD]
OG5_168371	IPR001254 (Serine proteases, trypsin domain), IPR001314 (Peptidase S1A, chymotrypsin-type), IPR018114 (Peptidase S1, trypsin family, active site), IPR009003 (Peptidase S1, PA clan)			tetur09g03880[AD,P] tetur12g03940[AD,P] tetur12g03950[P] tetur66g00060 tetur66g00070
OG5_176862	IPR002919 (Trypsin Inhibitor-like, cysteine rich domain)			tetur22g00260[AD]
OG5_207753	IPR002919 (Trypsin Inhibitor-like, cysteine rich domain)			tetur22g00290[P] tetur40g00392[AD,ISH,P]

[AD] Protein identified in artificial diet fed upon by *T. urticae* (**Chapter 3**), [ISH] Gene coding for the protein is confirmed to be expressed in the salivary glands using whole-mount *in situ* hybridizations (**Chapter 2, 3 & 5**), [P] Expression level of the gene coding for the protein is significantly higher (FC≥8, adj. p≤0.05) in the proterosoma versus the entire body of *T. urticae* adapted to bean plants (**Chapter 3**, TABLE S3.9)

Table S4.2. Spider mite-specific *Tetranychus evansi*, *T. lintearius* and *T. urticae* proteins homologous to proteins identified in artificial diet fed upon by *T. urticae*.

OrthoMCL group	*T. evansi*	*T. lintearius*	*T. urticae*
Tu_MCL_12	tetev485g00030	tetli56g00590	tetur02g14280
			tetur02g14420
			tetur02g14470
			tetur08g03060
			tetur12g00580
			tetur13g04230[P]
			tetur19g92445
			tetur20g00010
			tetur20g00020
			tetur20g00030
			tetur20g00040
			tetur20g00050[P]
			tetur20g00090[P]
			tetur20g00100
			tetur20g00540[P]
			tetur20g00550
			tetur20g00560[AD, P, ISH]
			tetur19g00950
			tetur20g90779
			tetur20g90780
			tetur20g90781
			tetur19g02210
			tetur20g00670
			tetur47g00120
			tetur47g00170
			tetur43g00560
			tetur43g00570
			tetur43g00580
			tetur58g00010
			tetur58g00020
			tetur58g00030
			tetur58g00050
			tetur58g00090
			tetur58g00100
			tetur58g00110
			tetur58g90066
			tetur58g90067
Tu_MCL_25	tetev52g01490	tetli16g00750	tetur03g07610
			tetur03g07590
			tetur14g02720[P]
			tetur31g00630[AD, P]
			tetur31g00690[P]
			tetur31g00740[P]
			tetur31g00770

Table S4.2. Continued.

OrthoMCL group	*T. evansi*	*T. lintearius*	*T. urticae*
			tetur31g00810
			tetur31g00820
			tetur31g00830[AD]
			tetur31g00850
			tetur31g00890
			tetur31g00910
			tetur31g01040[AD, P]
			tetur31g01060[AD]
			tetur31g01860
Tu_MCL_25	tetev219g00200	tetli16g00760	tetur06g00230[AD, P, ISH]
		tetli174g00060	
		tetli174g00050	
Tu_MCL_43	tetev148g00350	tetli20g03110	tetur03g00040[P]
	tetev102g00780	tetli84g01040	tetur03g00050[P]
		tetli84g01050	tetur03g00070[P]
		tetli84g01060	tetur08g08070[P]
		tetli84g01070	tetur32g01880[AD, P]
		tetli84g00020	tetur33g00840
			tetur33g00850[P]
			tetur33g00860
			tetur33g00870[P]
Tu_MCL_35	tetev446g00020		tetur01g16564
	tetev370g00040		tetur04g09459
	tetev54g00740		tetur04g09479[AD, P]
			tetur04g09489
			tetur04g00770
			tetur12g00700
			tetur143g00020
			tetur32g02287
			tetur32g02297
			tetur32g00050[AD, P, ISH]
			tetur44g00291
			tetur44g00301
			tetur86g00050
			tetur86g00060
Tu_MCL_36	tetev13g01130		tetur01g10080
	tetev187g00210		tetur01g10090
	tetev192g00290		tetur05g04560[AD, P]
	tetev317g00050		tetur05g04580[AD, P, ISH]
			tetur16g90599
			tetur196g90150
			tetur28g01330[AD, P]
			tetur28g01360
			tetur28g01380
			tetur28g01410

Table S4.2. Continued.

OrthoMCL group	*T. evansi*	*T. lintearius*	*T. urticae*
			tetur55g00040[AD, P]
			tetur55g00090
Tu_MCL_45			tetur02g06500
			tetur02g06490
			tetur02g06520
			tetur02g06560
			tetur02g06600
			tetur02g92468
			tetur04g08190
			tetur06g06630[AD, ISH]
			tetur06g06650[P]
			tetur06g03000
			tetur307g00010
			tetur307g00020
Tu_MCL_63		tetli115g00430	tetur14g03160[AD]
		tetli115g00400	tetur14g03170[P]
			tetur14g03240[P]
			tetur14g03250
			tetur14g03560
			tetur14g03430[P]
			tetur14g02600[P]
			tetur517g00010[P]
Tu_MCL_74	tetev340g00035	tetli74g00350	tetur03g03620
	tetev340g00030		tetur03g10083[AD, P]
			tetur11g01360[P, ISH]
			tetur11g06390[P]
			tetur11g06400[P]
Tu_MCL_211	tetev132g00350	tetli11g00110	tetur01g00950[AD, P, ISH]
	tetev132g00300	tetli11g00160	tetur01g01000[P]
Tu_MCL_212	tetev132g00290	tetli11g00100	tetur01g00940[AD, P, ISH]
		tetli11g00170	tetur01g01010[AD, P]
Tu_MCL_153	tetev252g00100	tetli223g00110	tetur03g04460
		tetli223g00120	tetur03g04470[AD]
No group	tetev06g00640	tetli23g01520	tetur07g01660[AD, P]
		tetli23g01540	
		tetli66g00240	
No group	tetev223g00110		tetur06g00510[AD, P]
	tetev01g02160		tetur29g01360[AD, P, ISH]
No group	tetev117g00420	tetli101g00190	tetur03g08710[AD, P, ISH]
No group			tetur13g00610[P]
			tetur13g00600[AD, P]
			tetur13g00650[P]
No group			tetur09g00900[AD, P]
			tetur03g10103[P]
No group	tetev168g00210		tetur10g00090[AD, P, ISH]

TABLE S4.2. Continued.

OrthoMCL group	*T. evansi*	*T. lintearius*	*T. urticae*
No group	tetev168g00220		tetur10g00100[AD, P]
OG5_126560	tetev64g00950		tetur20g01290[AD]

[AD] Protein identified in artificial diet fed upon by *T. urticae* (**Chapter 3**)

[P] Expression level of the gene coding for the protein is significantly higher (FC≥8, adj. $p \leq 0.05$) in the proterosoma versus the entire body of *T. urticae* adapted to bean plants (**Chapter 3**, TABLE S3.9)

[ISH] Gene coding for the protein is confirmed to be expressed in the salivary glands with whole-mount *in situ* hybridizations (**Chapter 3 & Chapter 5**)