Transcriptomic plasticity in the arthropod generalist *Tetranychus urticae* upon long-term acclimation to different host plants

Simon Snoeck[*,1], Nicky Wybouw[*,1], Thomas Van Leeuwen[*,†], Wannes Dermauw[*]

[*]Laboratory of Agrozoology, Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, 9000 Ghent, Oost-Vlaanderen, Belgium.

[†]Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, 1012 Amsterdam, Noord-Holland, the Netherlands.

[1] both authors equally contributed

Reference numbers for data available in public repositories:

GSE50162, GSE80337 and GSE68708

Running Title: Transcriptomic plasticity in an herbivore

Keywords: plant defense, host plant use, xenobiotic metabolism, single PLAT domain protein, short-chain dehydrogenase

Corresponding authors:
Prof. dr. ir. Thomas Van Leeuwen & dr. ir. Wannes Dermauw
Laboratory of Agrozoology
Department of Plants and Crops
Faculty of Bioscience Engineering
Ghent University
Coupure Links 653
9000 Ghent
Belgium
Telephone: +32(0)9 264 61 43 and +32(0)9 264 61 92
thomas.vanleeuwen@ugent.be and wannes.dermauw@ugent.be

ABSTRACT

The two-spotted spider mite *Tetranychus urticae* is an important pest with an exceptionally broad host plant range. This generalist rapidly acclimatizes and adapts to a new host, hereby overcoming nutritional challenges and a novel pallet of constitutive and induced plant defenses. Although recent studies reveal that a broad transcriptomic response upon host plant transfer is associated with a generalist life style in arthropod herbivores, it remains uncertain to what extent these transcriptional changes are general stress responses or host-specific. In the present study, we analyzed and compared the transcriptomic changes that occur in a single *T. urticae* population upon long-term transfer from *Phaseolus vulgaris* to a similar, but chemically defended, host (cyanogenic *Phaseolus lunatus)* and to multiple economically important crops (*Glycine max, Gossypium hirsutum*, *Solanum lycopersicum* and *Zea mays*). These long-term host plant transfers were associated with distinct transcriptomic responses with only a limited overlap in both specificity and directionality, suggestive of a fine-tuned transcriptional plasticity. Nonetheless, analysis at the gene family level uncovered overlapping functional processes, recruiting genes from both well-known and newly discovered detoxification families. Of note, our analyses highlighted a possible detoxification role for *Tetranychus*-specific short-chain dehydrogenases and single PLAT domain proteins, and manual genome annotation showed that both families are expanded in *T. urticae*. Our results shed new light on the molecular mechanisms underlying the remarkable adaptive potential for host plant use of generalist arthropods and set the stage for functional validation of important players in *T. urticae* detoxification of plant secondary metabolites.

INTRODUCTION

Host plant acceptance by a herbivore is influenced by the nutrient composition, with the protein/carbohydrate ratio being a critical characteristic (Mattson 1980; Behmer 2009). In addition to the nutrient composition of a potential host plant, plant defenses play a pivotal role in host plant acceptance. These defenses can either be chemical or physical (e.g. thorns, and trichomes). Chemical plant defense barriers include the production of toxic plant secondary metabolites and anti-digestive compounds as well as the attraction of enemies of the herbivore via the release of plant volatiles. Plants can also re-allocate their resources towards non-attacked tissues to minimize the negative fitness consequences of tissue loss due to herbivore feeding (Howe and Jander 2008; Kant *et al.* 2015).

Based on the timing of their production, plant secondary defense metabolites can be divided into two categories. Phytoanticipins are synthesized constitutively whereas phytoalexins are induced upon herbivore or pathogen attack via damage recognition and mediated by well-characterized plant hormone systems (Kant *et al.* 2015; Rioja *et al.* 2017; Stahl *et al.* 2018). Across the plant kingdom, a staggering diversity of phytoalexins and phytoanticipins have evolved in the co-evolutionary arms-race between plants and their attackers (Rosenthal and Berenbaum 1991; Wink 2010). Phytoanticipins with well-characterized effects on feeding herbivores include the glycoalkaloid tomatine in tomato, the benzoxazinoid DIMBOA-Glc in maize and related grasses and cyanogenic glucosides in cyanogenic plants, including cassava (Glauser *et al.* 2011; DongSub *et al.* 2014; Pičmanová *et al.* 2015). On the other hand, gossypol in cotton and isoflavonoids in legumes are examples of phytoalexins (Mccormick 1982; Dakora and Phillips 1996). Another well-studied induced defense response upon herbivore feeding is the increased production of anti-digestive compounds such as proteinase inhibitors (PIs) that

77 causes an amino acid deficiency in the attacking herbivore (Green and Ryan 1972; Hartl *et al.*
78 2010). Such increase of PIs has been observed in many plants upon attack of insects but also
79 spider mite herbivores such as *Tetranychus urticae* and *Tetranychus evansi* (Kant *et al.* 2004;
80 Martel *et al.* 2015).

81 Arthropod herbivores have developed several mechanisms to avoid, resist or suppress plant
82 defenses. Two main mechanisms are thought to allow herbivores to cope with ingested plant
83 secondary metabolites: (1) mechanisms that decrease sensitivity and (2) mechanisms that
84 decrease exposure to plant metabolites, such as sequestration and increased metabolism (Kant
85 *et al.* 2015). With the exception of a number of biological systems such as herbivore resistance
86 against plant cardenolides (Dobler *et al.* 2012), the first mechanism has only been rarely
87 documented. Probably, this is because plant secondary metabolites often have multiple or
88 unspecific modes of action, in contrast to, for example, insecticides used to control insects and
89 mites (for a review, see Feyereisen et al. 2015 and Van Leeuwen & Dermauw 2016).
90 Mechanisms of decreased exposure, on the other hand, are far better documented and in many
91 cases are mediated by genes that code for enzymes and transporters that typically belong to
92 ubiquitous multi-gene families (Després *et al.* 2007; Heckel 2014; Heidel-Fischer and Vogel
93 2015; Erb and Robert 2016). Metabolic detoxification can be categorized into three phases
94 based on the interaction with the ingested toxin. These interactions are: direct metabolism
95 (phase I), conjugation (phase II) and translocation (phase III). Enzymes that operate during
96 phase I are often cytochrome P450 monooxygenases (P450s) and carboxyl/choline esterases
97 (CCEs), whereas enzymes such as glutathione-S-transferases (GSTs) and UDP-
98 glycosyltransferases (UGTs) typically operate during phase II. Finally, transport of the toxins
99 or the phase II metabolites out of the cell or into specialized cell compartments is often
100 performed by ATP-binding cassette (ABC) transporters and solute carrier (SLC) family
101 proteins (Brattsten 1988; Després *et al.* 2007; Dermauw and Van Leeuwen 2014; Heckel 2014;

Kant *et al.* 2015). Recently it was reviewed whether the genes encoding the above-mentioned enzymes and transporters are less abundant in the genomes of specialist herbivores (those restricted to one or a few related host plants) compared to generalist herbivores (able to feed on a diverse set of host plants) and whether generalist herbivores are genetically predisposed to rapidly develop pesticide resistance; so far, conclusive evidence is yet to be found for these relationships (Rane *et al.* 2016; Dermauw *et al.* 2018; Hardy *et al.* 2018). Some studies also found that arthropod generalists exhibit a remarkably stronger transcriptional response upon host plant transfer compared to specialists and suggest that this activity is linked to the ability to cope with different host plants (Wybouw *et al.* 2015; Schweizer *et al.* 2017). More studies are however needed to establish the generality of these observations across the Arthropoda phylum.

Instead of coping with ingested plant secondary metabolites, some arthropod herbivores also evolved the ability to suppress the induced plant defenses, mostly via the secretion of molecules directly into the plant tissue (named effectors, reviewed in Kant et al. 2015 and Felton et al. 2014) and expansion of salivary genes has been suggested to be important in the adaptation processes of generalist herbivores (Jonckheere 2018; Boulain *et al.* 2018). The relative importance of plant defense manipulation and detoxification in arthropod-plant interactions remains however to be determined (Rioja *et al.* 2017; Blaazer *et al.* 2018).

The two-spotted spider mite *T. urticae* is among the most polyphagous herbivores known, with a host range covering more than 1,100 different plant species, scattered over more than 140 different plant families. Together, these plants produce a staggering number of different plant defense metabolites (Migeon, Nouguier and Dorkeld, 2018; Grbić *et al.*, 2011). It is well known that *T. urticae* populations readily, but differentially, adapt to a novel plant (Gould 1979; Fry 1989; Agrawal *et al.* 2002; Magalhães *et al.* 2007, 2009). Analysis of the *T. urticae* genome

126 revealed large, lineage-specific expansions of detoxification gene families including P450s,
127 CCEs, GSTs, UGTs and ABCs (Grbić *et al.* 2011; Dermauw *et al.* 2013a; Ahn *et al.* 2014). In
128 addition, analysis of the spider mite salivome revealed a whole array of putative effectors
129 (Jonckheere *et al.* 2016; Jonckheere 2018), of which two effectively suppress plant defenses
130 and promote mite performance (Villarroel *et al.* 2016).

131 In recent years, many of the "classical" detoxification genes (coding for P450s, CCEs, GSTs,
132 UGTs and ABCs) were shown to be differentially expressed upon transfer of mite populations
133 to different host plants, but also genes previously not known to be implicated in arthropod
134 detoxification were uncovered (Grbić *et al.* 2011; Dermauw *et al.* 2013b; Ahn *et al.* 2014;
135 Zhurov *et al.* 2014; Wybouw *et al.* 2015). These include genes coding for binding/sequestering
136 proteins such as lipocalins and transporters of the Major Facilitator Superfamily. Remarkably,
137 spider mites have also acquired novel metabolic abilities via horizontal gene transfer. The
138 horizontally transferred gene repertoire of *T. urticae* includes a family of 17 intradiol ring-
139 cleavage dioxygenases (DOGs) capable of hydrolyzing aromatic ring structures (Dermauw *et*
140 *al.* 2013b; Wybouw *et al.* 2015), but also a gene (β-cyanoalanine synthase) that was horizontally
141 transferred from bacteria and of which its encoded enzyme detoxifies hydrogen cyanide
142 (Wybouw *et al.* 2014) (see Wybouw et al. 2016 and Wybouw et al. 2018 for the general role of
143 horizontal gene transfer in the evolution of insect and mite herbivory). The majority of these
144 and other gene expression studies was however based on short-term transfer (less than or equal
145 to 24 h) of plant-feeding mites to a new host (Grbić *et al.* 2011; Dermauw *et al.* 2013b; Zhurov
146 *et al.* 2014) and only few studies have assessed mite gene expression changes upon long-term
147 acclimation (> 1 generation) or adaptation to a new host (Dermauw et al., 2013b; Wybouw et
148 al., 2014, Wybouw 2015). Moreover, studies examining expression changes upon long-term
149 acclimation in non-chelicerate arthropod herbivores are very scarce (Xie *et al.* 2014; Müller *et*
150 *al.* 2017; Mathers *et al.* 2017). In this study, we examined the transcriptomic responses of *T.*

urticae to a long-term transfer from bean to five different host plants; lima bean, soybean, cotton, tomato, and maize. We assessed the host plant specificity and overlap of these transcriptomic changes and dissected the different gene families involved, including "unexpected" families such as short-chain dehydrogenases (SDRs) and single PLAT domain proteins.

MATERIALS AND METHODS

Plants and spider mites

The ancestral reference population ('London') originates from a wild-collected *T. urticae* population from the Vineland region (Ontario, Canada) and was previously described (Grbić *et al.* 2011). The London laboratory population was maintained on potted common bean plants (*Phaseolus vulgaris* L. cv. 'Prelude') at a continuously high population density and served as the ancestral population for all host plant transfers in the current study. Lines were established on different host plants by transferring about 250 adult females to lima bean (*Phaseolus lunatus* L. cv. 8078), soybean (*Glycine max*), maize (*Zea mays* L. cv. 'Ronaldinio'), tomato (*Solanum lycopersicum* L. cv 'Moneymaker') and cotton (*Gossypium hirsutum*) (see (Wybouw *et al.* 2012, 2015; Jonckheere *et al.* 2016) for a more detailed description of the experimental set-up). Three independent lines were generated on cotton and tomato, whereas four independent lines were obtained for lima bean, maize, and soybean. All lines were mass reared on their respective host plants at 26 °C (±0.5 °C), 60% relative humidity (RH) and 16/8 h light/dark photoperiod.

RNA isolation, gene expression microarray set-up and differential gene expression analysis

Samples were collected from the soybean, cotton, and maize lines three months (approximately five generations) after transfer to the new host (Jonckheere *et al.* 2016), while the tomato and lima bean lines were collected after 18 months (approximately 30 generations) (Wybouw *et al.* 2014, 2015). Per sample, RNA was extracted from a pool of 100-120 adult females using a RNeasy minikit (Qiagen). Following DNase treatment (Turbo DN*ase*, Ambion), the concentration and integrity of the RNA samples were assessed by Nanodrop and by running 1µl on a 1% agarose gel. RNA was labelled using the Low Input Quick Amplification Kit (Agilent Technologies) following the manufacturer's instructions. RNA that was collected from mites of the ancestral London population on common bean and of mites that were transferred to a novel host were consistently dyed with cy3 and cy5, respectively. Cyanine-labelled RNA was hybridized to a custom-made gene-expression microarray (GEO Platform GPL16890, Bryon et al. 2013). Hybridization, washing and scanning protocols were identical as previously described (Dermauw *et al.* 2013b). Raw intensity data were used as input for final processing and statistical analysis in limma of the Bioconductor framework (Smyth 2004). Here, background correction was first performed by the 'normexp' method, using an offset of 50 (Ritchie *et al.* 2007). Background-corrected data were within- and between-array normalized (global loess and Aquantile, respectively) and quality was subsequently assessed using arrayQualityMetrics (Kauffmann *et al.* 2009). Prior to final differential gene expression analysis, the 55,469 probe sequences were remapped to the *T. urticae* genome annotation of August 11, 2016 (File S1) using Bowtie2-2.2.6 with default settings (Langmead and Salzberg 2012). Only the 36,589 probes that uniquely aligned to the annotated genome were incorporated in the differential gene expression analysis. A linear model was fitted to the processed data that treated the ancestral population as a common reference (cy3 channel in sample GSM1214964-GSM1214967, GSM2124774-GSM2124784 and GSM1679383-GSM1679385). Significant

differential gene-expression was identified via empirical Bayesian statistics and in reference to the ancestral population on common bean. Significant differentially expressed genes (DEGs) were identified by applying a 0.05 and 0.585 cut-off for Benjamini-Hochberg corrected p-value and absolute $\log_2$FC, respectively. The DEG set of each replicated host plant population was tested for enrichment of multigene families (OrthoMCL groups with at least 10 members) using a Chi square test. A Principal Component Analysis (PCA) was performed using the relative gene expression levels of all genes present on the array platform and the prcomp function within the R environment. *T. urticae* gene expression data are accessible at the Gene Expression Omnibus with accession numbers GSE50162, GSE80337 and GSE68708.

_k_-means clustering

The optimal cluster number for the *k*-means clustering approach was assessed using the gap statistic (method="global max", seed set at 54321, cluster number ranging from 1 to 10) (Tibshirani *et al.* 2001). The centered Pearson's correlation was used as the distance metric for *k*-means clustering. The relative transcription levels of genes that were significantly differentially expressed in any transcriptomic comparison were used as input for *k*-means clustering. Venn-diagrams were created for both the upregulated and downregulated transcripts using the VennDiagram_1.6.20 package in the R environment.

GO enrichment of differentially expressed genes

Gene Ontology (GO)-terms were assigned to *T. urticae* proteins using Blast2GO. The complete *T. urticae* proteome (19,086 sequences, version of August 11, 2016) was first used as query in a blastp search against the non-redundant protein database in NCBI (version of March 12, 2018) using the following settings "-outfmt 5 -evalue 1e-5 -word_size 3 -sshow_gis -num_alignments 20 -max_hsps_per_subject 20". The resulting blastp output was then loaded into the Blast2GO

(version 5.1) program and *T. urticae* proteins were annotated using the default parameters (Conesa *et al.* 2005). InterProScan 5 and ANNEX were used to augment the annotation of GO terms. GO terms were condensed using the generic GO Slim subset. Gene set enrichment analyses were conducted using the Bioconductor package piano (Väremo *et al.* 2013). The mite transcriptomic changes associated with the five host plant transfers (lima bean, soybean, cotton, tomato, and maize) were analyzed with the differential gene expression-associated statistics in a distinct directional gene set analysis (PAGE).

OrthoMCL clustering

OrthoMCL clustering of *T. urticae* proteins was derived from Jonckheere et al. 2018. InterProScan 5.25-64, with an E-value threshold of E^{-3}, was used to identify PFAM domains in the *T. urticae* proteome (version of August 11, 2016) and PFAM domains were assigned to each OrthoMCL group based on the presence of PFAM domains in *T. urticae* proteins contained within each group. Each OrthoMCL group was filtered for those proteins of which their corresponding genes did not had probes on the microarray. For each filtered OrthoMCL protein group (having at least 5 members), we determined the percentage of corresponding genes that was differentially expressed upon long-term transfer to a host plant using the package dplyr version 0.7.4 (Wickham and Francois 2015) within the R-framework (R Developement Core Team 2015). A two-sided Fisher's exact test in combination with the Benjamini-Hochberg procedure for multiple testing correction using all *T. urticae* genes (having uniquely mapping probes on the array; 13,943 genes in total) as a reference was employed to identify significantly enriched OrthoMCL groups (FDR < 0.05) among the DEG sets of the different *T. urticae* host plant populations.

Phylogenetic analysis of short-chain dehydrogenases

Among the significantly enriched OrthoMCL groups we identified two groups containing SDRs. The *T. urticae* proteome was mined for proteins with SDR-related PFAM domains; PF00106, PF01073, PF01370 and PF13561 (Persson and Kallberg 2013). Those *T. urticae* proteins with a SDR-related PFAM domain were used as query in a tblastn and blastp search (E-value threshold of E^{-3}) against the *T. urticae* genome (Grbić *et al.* 2011) and proteome (version of August 11, 2018) respectively. *T. urticae* SDR gene models were modified when necessary or new SDR gene models were created using Genomeview (Abeel *et al.* 2012). *H. sapiens* SDRs were derived from (Bray *et al.* 2009), while those of *Drosophila melanogaster* and *Metaseiulus occidentalis* were identified by mining their proteomes (*M. occidentalis* 1.0 (GNOMON release, (Hoy *et al.* 2016)) and *D. melanogaster* release 6.16 (FlyBase (Gramates *et al.* 2017)), respectively) for the above-mentioned SDR-related PFAM domains (see File S2 for accession numbers). Full-length *T. urticae* SDRs were aligned with those of *M. occidentalis*, *D. melanogaster*, *Homo sapiens* and *T. urticae* using the online version of MAFFT 7 with the E-INS-i iterative refinement method strategy (Katoh *et al.* 2002), 1,000 iterations and the option "reorder". The SDR alignment was trimmed using trimAl v1.4 (Capella-Gutiérrez *et al.* 2009) as SDR sequences are known to be highly divergent (Persson *et al.* 2003). A phylogenetic analysis was performed on the Cipres web portal (Miller *et al.* 2010) using RAxML v8 HPC2-XSEDE (Stamatakis 2014) with the automatic protein model assignment algorithm using maximum likelihood criterion and 1,000 bootstrap replicates. The LG+G protein model was selected as the optimal model for maximum likelihood analysis. The resulting tree was midpoint rooted, visualized using MEGA6 (Tamura *et al.* 2013) and edited in CorelDRAW Home & Student ×7.

Phylogenetic analysis of single PLAT domain proteins

OrthoMCL group Tetra_22 consisted of 20 proteins, of which three (tetur02g12320, tetur02g15207 and tetur22g02180) had a single PLAT (polycystin-1, lipoxygenase, alphatoxin) PFAM domain (PF01477) and eleven proteins belonged to the CATH/Gene3D Superfamily 2.60.60.20 (PLAT/LH2). Throughout this study, we refer to the proteins in Tetra_22 as *T. urticae* single PLAT domain proteins. *T. urticae* single PLAT domain proteins were used as query in a blastp and tblastn search (E-value threshold of E^{-3}) against the *T. urticae* proteome (version of August 11, 2018) and genome (Grbić *et al.* 2011), respectively. *T. urticae* single PLAT domain protein gene models were modified when necessary or new single PLAT domain protein gene models were created using Genomeview (Abeel *et al.* 2012). The transcriptomes of related tetranychid species, *Tetranychus evansi, Panonychus ulmi* and *Panonychus citri* (Bajda *et al.* 2015; Villarroel *et al.* 2016) were mined for single PLAT domain protein genes using tblastn (with an E-value threshold E^{-5}) and *T. urticae* single PLAT domain proteins as query. Redundant tblastn transcript hits were filtered using the cd-hit-est software (Fu *et al.* 2012) with the following settings "-c 0.95 -n 10". Those *T. evansi*, *P. ulmi* and *P. citri* tblastn hits of more than 100 amino acids long were retained for further analysis. In addition, we also mined the NCBI non-redundant protein database (version of May 1 2018) for the presence of these proteins in non-tetranychid species using blastp (with an E-value threshold E^{-5}) and *T. urticae* single PLAT domain proteins as query (see File S3 for accession numbers). Full-length *T. urticae* single PLAT domain proteins were aligned with those of *T. evansi*, *P. ulmi* and *P. citri* using the online version of MAFFT 7 with the E-INS-i iterative refinement method strategy (Katoh *et al.* 2002), 1000 iterations and the option "reorder". A phylogenetic analysis was performed on the Cipres web portal (Miller *et al.* 2010) using RAxML v8 HPC2-XSEDE (Stamatakis 2014) with the automatic protein model assignment algorithm using maximum likelihood criterion and 1,000 bootstrap replicates. The LG+G protein model was selected as the optimal model for maximum likelihood analysis. Phylogenetic trees were visualized using

MEGA6 (Tamura *et al.* 2013) and edited in CorelDRAW Home & Student ×7. The resulting tree was midpoint rooted, visualized using MEGA6 (Tamura *et al.* 2013) and edited in CorelDRAW Home & Student ×7.

Detection and analysis of SDR and PLAT clusters

A sliding window approach earlier described in Ngoc et al., 2016 was used to identify clusters of both the SDR and single PLAT domain protein genes throughout the *T. urticae* genome. A 50-kb window, incremented in 10-kb steps, was used. Only complete SDR and single PLAT domain protein genes were included in the analysis. Genes were considered as a part of each sliding window cluster if any portion of them overlapped the 50-kb window. Neighboring clusters that shared at least one gene were considered to be part of the same cluster, and were merged into a single larger cluster (as described in Thomas, 2006). The midpoints of the final clusters and the number of genes within each cluster were used for plotting.

Data availability

File S1 contains the CDS sequences of the *T. urticae* genome annotation of August 11, 2016. File S2 contains the protein sequences of the short-chain dehydrogenases of *T. urticae*, *M. occidentalis*, *D. melanogaster* and *H. sapiens* that were included in the phylogenetic analysis. File S3 contains the sequences of the full-length single PLAT domain proteins of *T. urticae*, *T. evansi*, *P. ulmi* and *P. citri* that were included in the phylogenetic analysis. Figure S1 shows the expression heatmaps of genes coding for group I and II SDRs and of group I and II single PLAT domain protein genes across the replicated *T. urticae* host plant populations (lima bean, soybean, cotton, tomato and maize). Table S1 contains the differential gene expression results of the *T. urticae* host plant populations. Table S2 shows the overlap between DEGs of the different *T. urticae* host plant populations. Table S3 shows the *k*-means clustering of the DEGs

identified in the different host plant populations of *T. urticae*. Table S4 shows the OrthoMCL grouping of the *T. urticae* proteome. Table S5 contains the OrthoMCL enrichment results of the DEG sets of each *T. urticae* host plant population. Table S6 contains the significantly enriched GO terms in the DEG sets of the different host plant populations of *T. urticae*. Table S7 contains the short-chain dehydrogenase genes annotated in the *T. urticae* genome. Table S8 contains the single PLAT domain protein genes annotated in the *T. urticae* genome. *T. urticae* gene expression data are available at the Gene Expression Omnibus with accession numbers GSE50162, GSE80337 and GSE68708. Supplemental material is available at Figshare: https://figshare.com/s/b99d2ba31a466a997998

RESULTS

Effect of long-term acclimation to different host plants on the *T. urticae* transcriptome

Using a whole-genome gene expression microarray, we measured significant gene expression changes in *T. urticae* adult females upon long-term transfer from common bean to either lima bean, soybean, cotton, tomato or maize ($\log_2 FC \geq 0.585$ and Benjamini-Hochberg corrected p-value < 0.05). A PCA plot revealed that 35.5 and 20.9% of the total gene expression variation across host plant lines could be explained by PC1 and PC2, respectively (Figure 1A). Individual lines clustered by host plant on both PC1 and PC2, with PC1 clearly separating the three tomato lines from the other host plant lines. The lima bean and cotton lines clustered along PC1. Our statistical analysis showed that the host plant transfer from bean to tomato resulted in the highest number of DEGs, 1,982 DEGs in total, of which 864 were upregulated and 1,118 downregulated (Table 1). On the other hand, acclimation to lima bean resulted in the lowest amount of DEGs, 410 in total, containing 307 upregulated and 103 downregulated genes. Long-term transfer to

soybean, cotton, and maize resulted in 789, 842 and 1,111 DEGs, respectively (Table 1, Table S1). In terms of amplitude, the replicated transfers to tomato and cotton plants resulted in the highest up- and downregulated DEGs. The DEG set of each replicated host plant population was enriched in multigene families (OrthoMCL groups ≥ 10 members), with 226/410, 304/789, 292/842, 479/1982, 408/1111 of the DEGs of lima bean, soybean, cotton, tomato, and maize line belonging to multigene families, respectively (Chi-square test p-values less than E^{-30} for each DEG set). As shown in Figure 1B, the majority of DEGs was not shared between the different host plant populations, with only nine upregulated genes and four downregulated genes in common for all transfers. These common upregulated DEGs coded for an intradiol ring-cleavage dioxygenase (*tetur28g01250*), a short-chain dehydrogenase (*tetur32g01960*), two Major Facilitator Superfamily proteins (*tetur03g04330* and *tetur11g05100*), a serine protease homologue (*tetur16g03330*), a CCAAT/enhance binding protein alpha (*tetur06g04210*), a LIM-domain (PF00412) protein (*tetur06g00950*) and two hypothetical proteins (*tetur23g01600*, *tetur22g00690*). The common downregulated DEGs coded for a small secreted protein from family A (*tetur22g02750*), a viral nucleoprotein (*tetur22g01100*, which was acquired through horizontal gene transfer (Wybouw et al. 2018)), and two hypothetical proteins (*tetur01g09880* and *tetur13g01730*). Fifty-four genes were upregulated for four out of five transfers, while only 57 were downregulated (Table S2). Of particular note, the tomato transfer resulted in the highest number of up- and downregulated genes that were not shared with the response of any other host plant population, and therefore appeared to be the most specific response (Table 1 and Figure 1).

***k*-means clustering of transcriptomic responses to long-term host plant transfer**

To get more insight into the global transcriptomic patterns, we performed a *k*-means clustering of the mite transcriptomic responses to the long-term host plant transfers using eight clusters

(cluster number identified using the gap-statistic, Figure 2). The identity of the DEGs in each of the eight groups is listed in Table S3. Three global transcriptomic patterns became apparent when focusing on these groups. Cluster 6 and 5, with a total of 850 DEGs, appeared to reflect a general response and did not exhibit any host plant specificity. Genes of clusters 1, 3, 7, and 8, with a total of 769 DEGs, were differentially up- and downregulated upon feeding to the different hosts of this study, hereby creating zig-zag patterns. Finally, clusters 2 and 4 appeared to reflect a host plant specific transcriptional response. Cluster 2 was assembled of DEGs (n=95) that were mainly specifically upregulated after long-term feeding on cyanogenic lima bean. This included *tetur10g01570,* a horizontally transferred gene of bacterial origin that codes for a functionally active β-cyanoalanine synthase that is able to detoxify cyanide, the main defense compound of lima bean (Wybouw *et al.* 2014). Cluster 4 consisted of the largest number of DEGs (n=704) and largely reflected a tomato-specific transcriptional response (Figure 2, Table S3).

Gene-set enrichment analysis

To look at gene family-wide patterns, we grouped *T. urticae* genes into OrthoMCL groups (Table S4, Figure 3), determined the percentage of DEGs for each OrthoMCL group for each replicated host plant population (Table S5) and subsequently performed an OrthoMCL enrichment analysis. Ten OrthoMCL groups were significantly enriched (FDR < 0.05) in all host plant populations: DOGs (OG5_134812), lipocalins (OG5_130527), cysteine proteases, papains (OG5_127800, OG5_126607), single PLAT domain proteins (Tetra_22), CCEs (OG5_126875), MFS proteins (OG5_138329), PAN domain proteins (Tetra_5) and hypothetical proteins (Tetra_9 (Small Secreted Protein Family A) and Tetra_24). A number of these gene groups (DOGs, lipocalins, CCEs, MFS, PAN-domain proteins, Tetra_9 (cluster 10066 in Dermauw *et al.*, 2013b) and Tetra_24 (cluster 10257 in Dermauw *et al.*, 2013b) were

previously significantly enriched in DEG lists of both mite resistant strains and a tomato acclimatized (5 generations) mite line, while cysteine proteases and single PLAT domain proteins (cluster 10204 in Dermauw *et al.*, 2013b) were only enriched in the tomato acclimatized mite line (Dermauw *et al.*, 2013b). Among the remaining significantly enriched OrthoMCL groups we identified ten *T. urticae* specific gene clusters, including Tetra_19, Tetra_38, Tetra_54, Tetra_62, Tetra_73, Tetra_74, Tetra_85, Tetra_112, Tetra_112, Tetra_116 and Tetra_195). Of particular note, members of OrthoMCL groups Tetra_19 (referred to as Tu_MCL_12 in Jonckheere et al. 2016), Tetra_54 (referred to as Tu_MCL_25 in Jonckheere et al. 2016), Tetra_62 (referred to as Tu_MCL_35 in Jonckheere et al. 2016, Small Secreted Protein Family F) were previously identified in the *T. urticae* salivary proteome and shown to be expressed in the salivary glands (Jonckheere *et al.* 2016). In addition, members of Tetra_54 were also shown to be constitutively upregulated in tomato-adapted mites (Wybouw *et al.* 2015). The replicated maize population had the highest number of significantly enriched OrthoMCLs (n=33), followed by the soybean (n=29), lima bean (n=28), cotton (n=24) and tomato (n=22) populations. Four significantly enriched OrthoMCLs were unique for the tomato-fed mites, including genes coding for BTB and C-terminal Kelch related proteins (OG5_184484), while three, one, one and three significantly enriched OrthoMCL groups were unique for the lima bean, soybean, cotton, and maize populations, respectively. As a next step in our functional characterization of the mite transcriptomic responses, we complemented our OrthoMCL analysis with GO enrichment analyses (Table S6). For the DEG sets upon the lima bean, soybean, cotton, and maize long-term transfer, only a few significantly enriched GO terms could be identified, as shown in Table S6. The highest number of significantly enriched GO terms (n=15) was found for the DEG list upon the long-term transfer to tomato, ranging from "perceiving signals" (GO:0007165) over transcription factor activity (GO:0003677 and GO:0003700) to "transmembrane transport" (GO:0055085).

Phylogenetic analysis of *T. urticae* SDR and PLAT gene family

Among the OrthoMCL groups that were significantly enriched we identified two gene families that have not yet been associated with mite xenobiotic response to host transfer: short-chain dehydrogenases (OG5_126860 and OG5_128170, having PFAM domain PF00106 and/or PF13561) and single PLAT domain proteins (PFAM domain PF01477). Both families were annotated within the *T. urticae* Sanger-sequenced genome assembly and their phylogenetic relatedness was investigated using a maximum-likelihood phylogenetic approach. Eighty-eight full-length SDR genes and 24 SDR gene fragments/pseudogenes were annotated in the *T. urticae* genome (Table S7). Full-length *T. urticae* SDR proteins were, together with those of *M. occidentalis*, *D. melanogaster* and *H. sapiens,* used in a maximum-likelihood analysis. We identified clear 1:1:1:1 orthology between five SDRs of each arthropod species and human SDRs (HSD17B4, HSD17B8, KDSR, TSTA3, WWOX and DHRSX), verifying the validity of our phylogenetic approach. Furthermore, we identified several *T. urticae* specific expansions. Twenty-Five *T. urticae* SDRs (either belonging to OrthoMCl group OG5_128170 or OG5_136892) clustered with high bootstrap support with 5 SDRs of both *M. occidentalis* and *D. melanogaster* (group I and tetur08g02060 in Figure 4), while ten *T. urticae* SDRs (OG5_126860) clustered with high bootstrap support with a *Drosophila* SDR (FBtr0071183) (group II in Figure 4). The latter *Drosophila* SDR, named sniffer, is a carbonyl reductase and has been shown to prevent oxidative stress-induced neurodegeneration (Martin *et al.* 2011). Remarkably, OG5_128170 was significantly enriched in the DEG sets of the bean, soybean, and maize populations whereas OG5_126860 was significantly enriched in DEGs of the tomato and maize populations (Figure 3). Finally, we also identified two smaller *T. urticae* SDR expansions, one with five SDRs in *T. urticae* (belonging to OrthoMCL group OG5_127561) compared to one in *M. occidentalis* (Mo_rna15492) and one with five SDRs in *T. urticae*

(belonging to OrthoMCL group OG5_131031) compared to one in both *M. occidentalis* (Mo_rna12331)*, D. melanogaster* (FBtr0074654) and *H. sapiens* (DCXR). Genes encoding *T. urticae* SDRs seem to be dispersed across the genome with 61.4% of them being singletons. However, most of the genes within two *T. urticae* SDR specific expansions (Group I and II (Figure 4, panel B) were found in clusters of scaffolds 6, 12 and 28. Within the SDR gene clusters on scaffolds 6 and 12, genes were not only found in a head-to-tail orientation but in both orientations. Moreover, the largest clusters (on scaffold 6 and scaffold 12) are rich in transposable elements (TE) sequences (see e.g. tetur12g00570 at ORCAE (Sterck *et al.* 2012)), suggestive of multiple duplication events. However, features of genomic distribution will become clearer once a chromosome-wide assembly of the *T. urticae* genome will be available.

We also investigated *T. urticae* single PLAT domain proteins into more detail. Twenty-one single PLAT domain protein genes were found in the *T. urticae* genome (20 were considered as full-length genes and one as a pseudogene), and four single PLAT domain gene fragments were identified. Next, we also identified single PLAT domain protein genes in the transcriptomes of other less polyphagous tetranychid mites such as *T. evansi* (n=8), *P. ulmi* (n=6) and *P. citri* (n=10) (Table S8). A blastp search against the non-redundant protein database in NCBI, revealed that tetranychid single PLAT domain proteins do not show sequence similarity with proteins of non-tetranychid eukaryotic species. A literature search, however, revealed that single PLAT domain proteins do occur in dicot and monocot plant species (Hyun *et al.* 2015), but these do not show sequence similarity with those of *T. urticae*. Nevertheless, both *T. urticae* and plant single PLAT domain proteins do share the same protein secondary structure, as they both have the PLAT domain, a β-sandwich composed of two sheets of four strands each (Bateman and Sandford 1999). Finally, we performed a maximum likelihood phylogenetic analysis using tetranychid single PLAT domain proteins

(Figure 5). We identified two clear expansions of single PLAT domain proteins in *T. urticae*, with one expansion consisting of six single PLAT domain proteins in *T. urticae* (group I) compared to one in *P. citri* and *T. evansi* and one expansion consisting of seven single PLAT domain proteins in *T. urticae* (group II) compared to one in *P. citri* and *T. evansi*. Interestingly, tetur11g05720 and tetur11g05730 of group I showed the strongest expression changes (log_2FC between -7 and 3) of all *T. urticae* single PLAT domain genes upon long-term transfer to any of the host plant lines (Table S1, Figure S1). Only 20% of the genes encoding single PLAT domain proteins in the *T. urticae* genome are singletons. The remaining single PLAT domain proteins (n=16) are found in clusters on scaffolds 6, 11, 15 and 22 (Figure 5, panel B). Similar as for the SDR genome distribution, single PLAT domain protein gene clusters contained genes in both orientations and were rich in TE sequences (see e.g. tetur11g05730 at the ORCAE database (Sterck *et al.* 2012)), suggestive of multiple duplication events.

DISCUSSION

Arthropod herbivores are important crop pests, and the last decade has seen an unprecedented increase in our understanding of the evolutionary mechanisms associated with resistance development to insecticides and acaricides used for their control. Given the wealth of knowledge on the molecular genetic mechanisms of pesticide resistance in mites and insects (Li *et al.* 2007; Feyereisen *et al.* 2015; Van Leeuwen and Dermauw 2016), it is surprising that mechanisms that allow broad plant use have remained elusive. In the last few years, an increasing number of transcriptomic studies have revealed that short-term exposure or within-generation transfer to novel hosts in polyphagous arthropod herbivores is associated with large transcriptional responses (e.g. Govind *et al.* 2010; Grbić *et al.* 2011; Dermauw *et al.* 2013b; de

499 la Paz Celorio-Mancera *et al.* 2013; Vogel *et al.* 2014; Zhurov *et al.* 2014; Roy *et al.* 2016).
500 Fewer studies have addressed changes in gene expression upon long-term exposure and
501 adaptation, especially in comparison to an ancestral genetic background (feeding on original
502 host). For spider mites, Wybouw et al. 2015 revealed that the number of DEGs and the extent
503 of transcriptional change increases over time and generations, and based on the functional
504 prediction of the DEGs upon short- and long-term exposure, it was postulated that these
505 transcriptional responses are adaptive, enabling the herbivore to survive a shift in dietary
506 nutrients and toxins. However, the few studies addressing these important evolutionary
507 processes looked at the transfer to a single or very few new hosts (Dermauw *et al.* 2013b;
508 Wybouw *et al.* 2014, 2015; Xie *et al.* 2014; Müller *et al.* 2017; Mathers *et al.* 2017). Therefore,
509 we have addressed in the current study to what extent the long-term transcriptional responses
510 are host plant specific, using spider mites as a model. In addition, we investigated which multi-
511 gene families were associated with the different host plant transfers.

512 Mites were transferred from their ancestral host (common bean) to lima bean, soybean, cotton,
513 tomato, and maize. These plant species were selected as many of these are economically
514 important crops on which spider mites are reported as pests (Van Leeuwen *et al.* 2014). In
515 addition, some of the metabolites that are produced by this selection of plants have been well-
516 characterized as plant allelochemicals with a defensive role against attacking herbivores,
517 including the cyanogenic glucosides of lima bean, the tomato alkaloid tomatine (DongSub *et*
518 *al.* 2014), coumestrol in soybean leaves (Yuk *et al.* 2011), the terpenoid gossypol produced by
519 cotton (Mccormick 1982) and the benzoxazinoid DIMBOA-Glc in maize (Glauser *et al.* 2011).
520 Remarkably, although it is suggested that generalists have a less fine-tuned, host-specific
521 regulation of gene expression compared to a specialist (Voelckel and Baldwin 2004; Govind *et*
522 *al.* 2010; Dermauw *et al.* 2013b), the majority of *T. urticae* DEGs were not shared between the
523 different host plant populations in our study. Furthermore, the number of DEGs upon the

524 different host plant transfers could also be related to the phylogenetic distance of the novel host
525 plant to the ancestral host (common bean, Fabaceae) (Table 1). The host specificity of the mite
526 transcriptomic response was also reflected in our *k*-means clustering analysis, where the
527 majority of the DEGs were present in clusters that show a host-specific pattern (Figure 2). Such
528 transcriptomic specificity was also observed recently in the oligophagous mustard leaf beetle
529 (*Phaedon cochleariae*) (Müller et al. 2017). When this beetle is transferred from its original
530 host *Brassica rapa* to *Nasturtium officinale* and *Sinapis alba* for 26 generations, transcriptomic
531 analysis shows that most of the response is host plant specific, even though the two new hosts
532 share the same classes of defensive metabolites as the ancestral host (glucosinolates, phenolics,
533 and terpenoids – in different compositions).

534 Wybouw et al. 2015 showed that approximately half of the complete transcriptional response
535 of *T. urticae* after a long-term exposure to tomato is genetically determined and thus evolves
536 upon tomato adaptation. The genetic changes in tomato-adapted mites affect both constitutive
537 transcription and within-generation transcriptional plasticity. Here, we did not investigate
538 whether the long-term transfer resulted in adaptation and therefore cannot distinguish between
539 genetic adaptation, environmental induction and an interaction between these factors as the
540 cause of the transcriptomic changes. Nevertheless, as the ancestral population was genetically
541 diverse and the PCA plot did not show any signs of genetic drift, a substantial part of the
542 observed responses probably resulted from genetic adaptation. In corroboration, *T. urticae*
543 populations have shown great adaptive potential to a diverse set of novel hosts in addition to
544 tomato (Gould 1979; Fry 1989; Magalhães *et al.* 2007, 2009; Wybouw *et al.* 2012).

545 Although the overall response was very specific on the gene level, there was much less
546 specificity on the gene family level, which does suggest the presence of common mechanisms
547 of acclimation and adaptation. Indeed, the set of DEGs of each host plant population was

548 significantly enriched for genes from multigene families (OrthoMCL groups ≥ 10 members)
549 and many of the multigene families that were significantly enriched, were previously shown to
550 respond to xenobiotic pressure. These families were comprised of P450s and CCEs, involved
551 in detoxification, cysteine proteases, involved in digestion, and previously unknown players in
552 xenobiotic detoxification such as DOGs, lipocalins and MFS proteins (Dermauw *et al.* 2013b;
553 Santamaría *et al.* 2015). The importance of the metabolic processes associated with their
554 activities was also partially reflected in the GO enrichment analysis, where GO terms "peptidase
555 activity", "transferase activity" and "transmembrane transport" were enriched in the DEG sets
556 of the host plant populations (Table S6). In addition to overall metabolic processes, these
557 transcriptomic changes upon acclimation to different host plants also provide a first link
558 between differential expression patterns of specific genes and known defense compounds of
559 each host plant. Gossypol, for example, is a well-known phytoanticipin in cotton and it has been
560 shown that UGT-glycosylation and P450-oxygenation of gossypol are important for gossypol
561 detoxification (Mao *et al.* 2007; Krempl *et al.* 2016). Interestingly, a *CYP* and *UGT* gene were
562 the most highly upregulated when feeding on cotton (log_2FC of 7.6 and 2.5 for *tetur07g06410*
563 (*CYP392A1*) and *tetur04g02350* (*UGT203A2*), respectively, see Table S1). Similarly, a *UGT*
564 (*tetur05g05020* (*UGT201B7*)) and *GST* (*tetur05g05270* (*TuGSTd15*)) were highly upregulated
565 in the maize population (log_2FC of 3.6 and 3.7, respectively) while downregulated or not
566 differentially expressed in all other host plants, and might thus be involved in the detoxification
567 of benzoxazinoids, phytochemicals that are widespread in grasses (Loayza-Muro *et al.* 2000;
568 Wouters *et al.* 2016).

569 Next to the overall implication of gene families known to be involved in arthropod xenobiotic
570 metabolism, our analyses also revealed the prominent presence of a number of gene families
571 that have only been marginally associated with arthropod detoxification (Figure 3 and Table
572 S5). For example, OrthoMCL analysis revealed that SDRs were significantly enriched in the

573 DEG sets of the *T. urticae* host plant populations (Table S1). The SDR superfamily is one of
574 the largest and most highly divergent protein superfamilies found in all domains of life
575 (Kallberg *et al.* 2010). SDR enzymes are 250-300 amino acids long (see InterPro domain
576 IPR020904) and are NAD(P)(H)-dependent oxidoreductases with low pairwise sequence
577 identities. They contain at least 2 domains, a structurally conserved N-terminal region which
578 binds NAD(H) or NADP(H) as a co-factor and a structurally variable C-terminal region that
579 binds the substrate and contains the amino acids involved in catalysis (Bray *et al.* 2009). In
580 contrast to the P450 superfamily, functional insights on the SDR superfamily are very scarce
581 (Škarydová and Wsól 2012). Carbonyl-reducing enzymes (CDRs) from the SDR superfamily
582 are known to be involved in the biosynthesis/metabolism of endogenous signaling molecules
583 like steroid hormones and retinoids, but are as well involved in the detoxification of endobiotics
584 and xenobiotics (Hoffmann and Maser 2007; Oppermann 2007; Škarydová and Wsól 2012). In
585 humans, SDRs have been shown to play a central role in phase I metabolism by converting
586 aldehydes or ketones into the corresponding alcohols, thereby reducing the overall chemical
587 activity of their substrates (Škarydová and Wsól 2012; Ebert *et al.* 2016). In insects, the best
588 characterized SDRs are alcohol dehydrogenases (Zhang *et al.* 2004; Mayoral *et al.* 2013;
589 Figueroa-Teran *et al.* 2016). However, only few studies report upon the possible role of SDRs
590 in arthropod-plant interactions. SDR genes are overexpressed in the Asian longhorn beetle
591 *Anoplophora chinensis* upon dietary changes (Mason *et al.* 2016) and are present in the saliva
592 of aphids, white flies and thrips (Su *et al.* 2012; Stafford-Banks *et al.* 2014). Reduction of
593 quinone by a carbonyl reductases in the luna moth *Actias luna*, is presumably the best known
594 example of an SDR that is involved in detoxification of a plant allelochemical (Lindroth 1991).
595 *Actias luna* larvae are able to feed on plants of the Juglandaceae family, which contain juglone,
596 a compound toxic to a variety of insects. Feeding larvae exhibited high carbonyl reductase and
597 glutathione transferase activity, and these activities have been linked to the metabolism of

598 juglone and related quinones in the plant family of the Jungladaceae (Lindroth 1989). Since
599 SDRs have only been marginally described in both the context of host plant transfer as well as
600 xenobiotic metabolism in mites, we have provided a survey of the SDR superfamily in *T.*
601 *urticae* and identified eighty-eight full length SDRs in the genome of *T. urticae*, including
602 several apparent species-specific expansions, which increased the diversity of the SDR
603 repertoire. One of the expansions clustered together with a *Drosophila* SDR, named sniffer, a
604 carbonyl reductase involved in the prevention of oxidative stress-induced neurodegeneration
605 (Martin *et al.* 2011). The production of reactive-oxygen species is an essential part of the plant
606 response towards herbivore attack, including those of spider mites (Santamaria et al. 2018).
607 Several *T. urticae* SDRs that clustered with *Drosophila* sniffer were differentially expressed
608 upon acclimation of *T. urticae* to different host plants (Figure 3, Figure S1, Table S1) and,
609 hence, might play a protective role during spider mite feeding.

610 Next to the SDR gene family, the presence of a remarkable set of proteins containing a single
611 PLAT domain was also evident from the OrthoMCL enrichment analysis (Figure 3). Proteins
612 with a PLAT domain are ubiquitously present across eukaryotic species (see species distribution
613 of PF01477 at https://pfam.xfam.org/) and PLAT domains are for example present in pancreatic
614 triglyceride lipases (cd01755 at Conserved Domain Database (CDD)). However, short single
615 PLAT domain proteins (less than 200 amino acids) are to our knowledge only present in plants
616 (see EOG09360P3N at OrthoDB v9.1 and cd01754 at the Conserved Domain Database for
617 phylogenetic distribution of these plant PLAT proteins) and apparently tetranychid mites (this
618 study). There is virtually nothing known about the possible role of these proteins in plants.
619 Hyun et al. 2015,2014, showed that a single PLAT domain protein of *Arabidopsis* (PLAT1,
620 AT4G39730*)* is involved in abiotic stress tolerance while in *Capsicum annuum* a single PLAT
621 domain protein, named CaTin1, interferes with the redox balance of plants, leading to an altered
622 response to ethylene and biotic/abiotic stress (Shin *et al.* 2004). Coker et al. 2005, on the other

623 hand, showed that a single PLAT domain protein gene (*FIT-6*) is upregulated upon fire damage.
624 In *T. urticae,* several single PLAT domain protein genes were among the DEGs with the
625 strongest transcriptional response upon long-term host transfer, with a single PLAT gene
626 (*tetur11g05730*) being more than 100-fold lower expressed upon long-term cotton feeding
627 while being about 10-fold overexpressed in the maize population. Although one must be
628 cautious when comparing genomic and transcriptomic data (e.g. recent duplications and lowly
629 expressed genes might be missed in transcriptomic data), a phylogenetic analysis using
630 tetranychid single PLAT domain protein sequences derived from genomic (*T. urticae*) and
631 transcriptomic data *(T. evansi*, *P. ulmi*, and *P. citri)* showed that single PLAT domain protein
632 genes were expanded in the polyphagous *T. urticae* compared to oligophagous tetranychid
633 species (Figure 5). Overall, it can be speculated that single PLAT domain proteins are involved
634 in the stress response of *T. urticae* and that their expansion might have contributed to the
635 polyphagous nature of this species.

636 In summary, we investigated long-term acclimation to five novel host plants in the spider mite
637 *T. urticae*. Using different analytical tools, we uncovered that responses were specific on the
638 individual gene level, but that similar gene families and metabolic processes were involved in
639 host plant use. A number of surprising new gene families have entered the stage, such as genes
640 encoding single PLAT domain proteins and short-chain dehydrogenases. Our data set identified
641 specific enzymes that likely underlie resistance to specific plant allelochemicals and now await
642 *in vitro* functional validation by recombinant expression in model systems like insect cells or
643 *E. coli* and/or *in vivo* functional validation by reverse and forward genetic approaches, once
644 they become available as robust tools for spider mite research.

Acknowledgments

The authors want to thank Pedro Hernandez Crespo for the initial annotation and analysis of the *T. urticae* single PLAT domain protein gene family. We also thank Wim Reubens and Wim Jonckheere for their help in establishing the *T. urticae* soybean, cotton, and maize lines and Robert Greenhalgh for running the InterProScan analysis and providing the script for drafting panel B in both Figure 4 and Figure 5.

Author Contributions Statement

TVL and WD designed experiment. WD, SS and NW conducted experiments. Analysis and interpretation of the results was done by SS, WD, NW and TVL. The manuscript was written by SS, WD, NW and TVL. All authors reviewed the manuscript.

Conflict of Interest Statement

The authors state that they have no conflict of interest.

Funding disclosure

NW and WD are postdoctoral fellows of the Research Foundation Flanders (FWO). This project was supported by NWO (the Netherlands) under the Joint Programming Initiative on Agriculture, Food Security and Climate Change (FACCEJPI) European Research Area Networks (ERA-NET) Plus framework (GENOMITE, project ID 137), the Research Foundation Flanders (FWO, Belgium) (grant G009312N to TVL and grant G053815N to TVL and WD).

668 **References**

669 Abeel, T., T. Van Parys, Y. Saeys, J. Galagan, and Y. Van De Peer, 2012 GenomeView: A
670 next-generation genome browser. Nucleic Acids Res. 40: e12.

671 Agrawal, A. A., F. Vala, and M. W. Sabelis, 2002 Induction of preference and performance
672 after acclimation to novel hosts in a phytophagous spider mite: adaptive plasticity? Am.
673 Nat. 159: 553–565.

674 Ahn, S. J., W. Dermauw, N. Wybouw, D. G. Heckel, and T. Van Leeuwen, 2014 Bacterial
675 origin of a diverse family of UDP-glycosyltransferase genes in the *Tetranychus urticae*
676 genome. Insect Biochem. Mol. Biol. 50: 43–57.

677 Bajda, S., W. Dermauw, R. Greenhalgh, R. Nauen, L. Tirry *et al.*, 2015 Transcriptome
678 profiling of a spirodiclofen susceptible and resistant strain of the European red mite
679 *Panonychus ulmi* using strand-specific RNA-seq. BMC Genomics 16: 974.

680 Bateman, A., and R. Sandford, 1999 The PLAT domain: a new piece in the PKD1 puzzle.
681 Curr. Biol. 9: R588-90.

682 Behmer, S. T., 2009 Insect herbivore nutrient regulation. Annu. Rev. Entomol. 54: 165–187.

683 Blaazer, C. J. H., E. A. Villacis-Perez, R. Chafi, T. Van Leeuwen, M. R. Kant *et al.*, 2018
684 Why do herbivorous mites suppress plant defenses? Front. Plant Sci. 9: 1057.

685 Boulain, H., F. Legeai, E. Guy, S. Morlière, N. E. Douglas *et al.*, 2018 Fast evolution and
686 lineage-specific gene family expansions of aphid salivary effectors driven by interactions
687 with host-plants. Genome Biol. Evol. 10: 1554–1572.

688 Brattsten, L. B., 1988 Enzymic adaptations in leaf-feeding insects to host-plant
689 allelochemicals. J. Chem. Ecol. 14: 1919–1939.

690 Bray, J., B. Marsden, and U. Oppermann, 2009 The human short-chain
691 dehydrogenase/reductase (SDR) superfamily: A bioinformatics summary. Chem. Biol.

692 Interact. 178: 99–109.
693 Bryon, A., N. Wybouw, W. Dermauw, L. Tirry, and T. Van Leeuwen, 2013 Genome wide
694 gene-expression analysis of facultative reproductive diapause in the two-spotted spider
695 mite *Tetranychus urticae*. BMC Genomics 14: 815.
696 Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón, 2009 trimAl: A tool for
697 automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:
698 1972–1973.
699 Coker, J. S., A. Vian, and E. Davies, 2005 Identification, accumulation, and functional
700 prediction of novel tomato transcripts systemically upregulated after fire damage.
701 Physiol. Plant. 124: 311–322.
702 Conesa, A., S. Götz, J. M. García-Gómez, J. Terol, M. Talón *et al.*, 2005 Blast2GO: A
703 universal tool for annotation, visualization and analysis in functional genomics research.
704 Bioinformatics 21: 3674–3676.
705 Dakora, F. D., and D. A. Phillips, 1996 Diverse functions of isoflavonoids in legumes
706 transcend anti-microbial definitions of phytoalexins. Physiol. Mol. Plant Pathol. 49: 1–
707 20.
708 De la Paz _Celorio-Mancera, M., C. W. Wheat, H. Vogel, L. Söderlind, N. Janz *et al.*, 2013
709 Mechanisms of macroevolution: polyphagous plasticity in butterfly larvae revealed by
710 RNA-Seq. Mol. Ecol. 22: 4884–4895.
711 Dermauw, W., and T. Van Leeuwen, 2014 The ABC gene family in arthropods: Comparative
712 genomics and role in insecticide transport and resistance. Insect Biochem. Mol. Biol. 45:
713 89–110.
714 Dermauw, W., E. J. Osborne, R. M. Clark, M. Grbić, L. Tirry *et al.*, 2013a A burst of ABC
715 genes in the genome of the polyphagous spider mite *Tetranychus urticae*. BMC
716 Genomics 14: 317.

717 Dermauw, W., A. Pym, C. Bass, T. Van Leeuwen, and R. Feyereisen, 2018 Does host plant
718 adaptation lead to pesticide resistance in generalist herbivores? Curr. Opin. Insect Sci.
719 26: 25–33.
720 Dermauw, W., N. Wybouw, S. Rombauts, B. Menten, J. Vontas *et al.*, 2013b A link between
721 host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus*
722 *urticae*. Proc. Natl. Acad. Sci. U. S. A. 110: E113-22.
723 Després, L., J.-P. David, and C. Gallet, 2007 The evolutionary ecology of insect resistance to
724 plant chemicals. Trends Ecol. Evol. 22: 298–307.
725 Dobler, S., S. Dalla, V. Wagschal, and A. A. Agrawal, 2012 Community-wide convergent
726 evolution in insect adaptation to toxic cardenolides by substitutions in the Na,K-ATPase.
727 Proc. Natl. Acad. Sci. 109: 13040–13045.
728 DongSub, K., N. HaeYoung, Y. Kwack, and C. ChangHoo, 2014 Secondary metabolite
729 profiling in various parts of tomato plants. Korean J. Hortic. Sci. & Technol. 32:
730 252–260.
731 Ebert, B., M. Kisiela, and E. Maser, 2016 Transcriptional regulation of human and murine
732 short-chain dehydrogenase/reductases (SDRs) – an in silico approach. Drug Metab. Rev.
733 48: 183–217.
734 Erb, M., and C. A. M. Robert, 2016 Sequestration of plant secondary metabolites by insect
735 herbivores: Molecular mechanisms and ecological consequences. Curr. Opin. Insect Sci.
736 14: 8–11.
737 Felton, G. W., S. H. Chung, M. G. E. Hernandez, J. Louis, M. Peiffer *et al.*, 2014 Herbivore
738 oral secretions are the first line of protection against plant-induced defences, pp. 37–76
739 in *Annual Plant Reviews*, John Wiley & Sons, Ltd, Chichester, UK.
740 Feyereisen, R., W. Dermauw, and T. Van Leeuwen, 2015 Genotype to phenotype, the
741 molecular and physiological dimensions of resistance in arthropods. Pestic. Biochem.

742 Physiol. 121: 61–77.
743 Figueroa-Teran, R., H. Pak, G. J. Blomquist, and C. Tittiger, 2016 High substrate specificity
744 of ipsdienol dehydrogenase (IDOLDH), a short-chain dehydrogenase from *Ips pini* bark
745 beetles. J. Biochem. 160: 141–151.
746 Fry, J. D., 1989 Evolutionary adaptation to host plants in a laboratory population of the
747 phytophagous mite *Tetranychus urticae* Koch. Oecologia 81: 559–565.
748 Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li, 2012 CD-HIT: Accelerated for clustering the next-
749 generation sequencing data. Bioinformatics 28: 3150–3152.
750 Glauser, G., G. Marti, N. Villard, G. a. Doyen, J. L. Wolfender *et al.*, 2011 Induction and
751 detoxification of maize 1,4-benzoxazin-3-ones by insect herbivores. Plant J. 68: 901–
752 911.
753 Gould, F., 1979 Rapid host range evolution in a population of the phytophagous mite
754 *Tetranychus urticae* Koch. Evolution (N. Y). 33: 791–802.
755 Govind, G., O. Mittapalli, T. Griebel, S. Allmann, S. Böcker *et al.*, 2010 Unbiased
756 transcriptional comparisons of generalist and specialist herbivores feeding on
757 progressively defenseless *Nicotiana attenuata* plants. PLoS One 5: e8735.
758 Gramates, L. S., S. J. Marygold, G. Dos Santos, J. M. Urbano, G. Antonazzo *et al.*, 2017
759 FlyBase at 25: Looking to the future. Nucleic Acids Res. 45: D663–D671.
760 Grbić, M., T. Van Leeuwen, R. M. Clark, S. Rombauts, P. Rouzé *et al.*, 2011 The genome of
761 *Tetranychus urticae* reveals herbivorous pest adaptations. Nature 479: 487–92.
762 Green, T. R., and C. A. Ryan, 1972 Wound-induced proteinase inhibitor in plant leaves: a
763 possible defense mechanism against insects. Science (80-.). 175: 776–777.
764 Hardy, N. B., D. A. Peterson, L. Ross, and J. A. Rosenheim, 2018 Does a plant-eating insect's
765 diet govern the evolution of insecticide resistance? Comparative tests of the pre-

766 adaptation hypothesis. Evol. Appl. 11: 739–747.
767 Hartl, M., A. P. Giri, H. Kaur, and I. T. Baldwin, 2010 Serine protease inhibitors specifically
768 defend *Solanum nigrum* against generalist herbivores but do not influence plant growth
769 and development. Plant Cell 22: 4158–4175.
770 Heckel, D. G., 2014 Insect detoxification and sequestration strategies, pp. 77–114 in *Annual*
771 *Plant Reviews: Insect-Plant Interactions*,.
772 Heidel-Fischer, H. M., and H. Vogel, 2015 Molecular mechanisms of insect adaptation to
773 plant secondary compounds. Curr. Opin. Insect Sci. 8: 8–14.
774 Hoffmann, F., and E. Maser, 2007 Carbonyl reductases and pluripotent hydroxysteroid
775 dehydrogenases of the short-chain dehydrogenase/reductase superfamily. Drug Metab.
776 Rev. 39: 87–144.
777 Howe, G., and G. Jander, 2008 Plant immunity to insect herbivores. Annu. Rev. Plant Biol.
778 59: 41–66.
779 Hoy, M. A., R. M. Waterhouse, K. Wu, A. S. Estep, P. Ioannidis *et al.*, 2016 Genome
780 sequencing of the phytoseiid predatory mite *Metaseiulus occidentalis* reveals completely
781 atomized Hox genes and superdynamic intron evolution. Genome Biol. Evol. 8: 1762–
782 1775.
783 Hyun, T. K., A. Albacete, E. van der Graaff, S. H. Eom, D. K. Großkinsky *et al.*, 2015 The
784 Arabidopsis PLAT domain protein1 promotes abiotic stress tolerance and growth in
785 tobacco. Transgenic Res. 24: 651–653.
786 Hyun, T. K., E. Van Der Graaff, A. Albacete, S. H. Eom, D. K. Großkinsky *et al.*, 2014 The
787 Arabidopsis PLAT domain protein1 is critically involved in abiotic stress tolerance.
788 PLoS One 9: e112946.
789 Jonckheere, W., 2018 The salivary proteome of *Tetranychus urticae*: Key to its polyphagous

790 nature? University of Amsterdam, 258 p.
791 Jonckheere, W., W. Dermauw, M. Khalighi, N. Pavlidi, W. Reubens *et al.*, 2018 A gene
792 family coding for salivary proteins (SHOT) of the polyphagous spider mite *Tetranychus*
793 *urticae* exhibits fast host-dependent transcriptional plasticity. Mol. Plant-Microbe
794 Interact. 31: 112–124.
795 Jonckheere, W., W. Dermauw, V. Zhurov, N. Wybouw, J. Van den Bulcke *et al.*, 2016 The
796 salivary protein repertoire of the polyphagous spider mite *Tetranychus urticae*: a quest
797 for effectors. Mol. Cell. Proteomics 15: 3594–3613.
798 Kallberg, Y., U. Oppermann, and B. Persson, 2010 Classification of the short-chain
799 dehydrogenase/reductase superfamily using hidden Markov models. FEBS J. 277: 2375–
800 2386.
801 Kant, M. R., K. Ament, M. W. Sabelis, M. A. Haring, and R. C. Schuurink, 2004 Differential
802 timing of spider mite-induced direct and indirect defenses in tomato plants. Plant
803 Physiol. 135: 483–495.
804 Kant, M. R., W. Jonckheere, B. Knegt, F. Lemos, J. Liu *et al.*, 2015 Mechanisms and
805 ecological consequences of plant defence induction and suppression in herbivore
806 communities. Ann. Bot. 115: 1015–1051.
807 Katoh, K., K. Misawa, K. Kuma, and T. Miyata, 2002 MAFFT: a novel method for rapid
808 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:
809 3059–3066.
810 Kauffmann, A., R. Gentleman, and W. Huber, 2009 arrayQualityMetrics - A bioconductor
811 package for quality assessment of microarray data. Bioinformatics 25: 415–416.
812 Krempl, C., T. Sporer, M. Reichelt, S.-J. Ahn, H. Heidel-Fischer *et al.*, 2016 Potential
813 detoxification of gossypol by UDP-glycosyltransferases in the two Heliothine moth
814 species *Helicoverpa armigera* and *Heliothis virescens*. Insect Biochem. Mol. Biol. 71:

815 49–57.
816 Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat.
817 Methods 9: 357–359.
818 Li, X., M. a Schuler, and M. R. Berenbaum, 2007 Molecular mechanisms of metabolic
819 resistance to synthetic and natural xenobiotics. Annu. Rev. Entomol. 52: 231–253.
820 Lindroth, R. L., 1989 Chemical ecology of the luna moth - Effects of host plant on
821 detoxification enzyme activity. J. Chem. Ecol. 15: 2019–2029.
822 Lindroth, R. L., 1991 Differential toxicity of plant allelochemicals to insects: roles of
823 enzymatic detoxication systems, pp. 1–34 in *Insect-plant interactions*, CRC Press.
824 Loayza-Muro, R., C. C. Figueroa, and H. M. Niemeyer, 2000 Effect of two wheat cultivars
825 differing in hydroxamic acid concentration on detoxification metabolism in the aphid
826 Sitobion avenae. J. Chem. Ecol. 26: 2725–2736.
827 Magalhães, S., E. Blanchet, M. Egas, and I. Olivieri, 2009 Are adaptation costs necessary to
828 build up a local adaptation pattern? BMC Evol. Biol. 9: 182.
829 Magalhães, S., J. Fayard, A. Janssen, D. Carbonell, and I. Olivieri, 2007 Adaptation in a
830 spider mite population after long-term evolution on a single host plant. J. Evol. Biol. 20:
831 2016–2027.
832 Mao, Y. B., W. J. Cai, J. W. Wang, G. J. Hong, X. Y. Tao *et al.*, 2007 Silencing a cotton
833 bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance
834 of gossypol. Nat. Biotechnol. 25: 1307–1313.
835 Martel, C., V. Zhurov, M. Navarro, M. Martinez, M. Cazaux *et al.*, 2015 Tomato whole
836 genome transcriptional response to *Tetranychus urticae* identifies divergence of spider
837 mite-induced responses between Tomato and Arabidopsis. Mol Plant-Microbe Interact
838 28: 343–361.

839 Martin, H. J., M. Ziemba, M. Kisiela, J. A. Botella, S. Schneuwly *et al.*, 2011 The Drosophila
840 carbonyl reductase sniffer is an efficient 4-oxonon-2-enal (4ONE) reductase. Chem.
841 Biol. Interact. 193: 48–54.
842 Mason, C. J., E. D. Scully, S. M. Geib, and K. Hoover, 2016 Contrasting diets reveal
843 metabolic plasticity in the tree-killing beetle, *Anoplophora glabripennis* (Cerambycidae:
844 Lamiinae). Sci. Rep. 6: 33813.
845 Mathers, T. C., Y. Chen, G. Kaithakottil, F. Legeai, S. T. Mugford *et al.*, 2017 Rapid
846 transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid
847 to colonise diverse plant species. Genome Biol. 18: 27.
848 Mattson, W. J., 1980 Herbivory in relation to plant nitrogen content. Annu. Rev. Ecol. Syst.
849 11: 119–161.
850 Mayoral, J. G., K. T. Leonard, M. Nouzova, F. G. Noriega, L. A. Defelipe *et al.*, 2013
851 Functional analysis of a mosquito short-chain dehydrogenase cluster. Arch. Insect
852 Biochem. Physiol. 82: 96–115.
853 Mccormick, J. P., 1982 Secondary metabolites of *Gossypium* - a biogenetic analysis. Acs
854 Symp. Ser. 189: 275–300.
855 Migeon, A., E. Nouguier, and F. Dorkeld, 2018 Spider Mites Web: A comprehensive database
856 for the Tetranychidae, pp. 557–560 in *Trends in Acarology*,.
857 Miller, M. A., W. Pfeiffer, and T. Schwartz, 2010 Creating the CIPRES science gateway for
858 inference of large phylogenetic trees, pp. 1–8 in *2010 Gateway Computing Environments*
859 *Workshop, GCE 2010*,.
860 Müller, C., H. Vogel, and D. G. Heckel, 2017 Transcriptional responses to short-term and
861 long-term host plant experience and parasite load in an oligophagous beetle. Mol. Ecol.
862 26: 6370–6383.

Ngoc, P. C. T., R. Greenhalgh, W. Dermauw, S. Rombauts, S. Bajda *et al.*, 2016 Complex evolutionary dynamics of massively expanded chemosensory receptor families in an extreme generalist chelicerate herbivore. Genome Biol. Evol. 8: 3323–3339.

Oppermann, U., 2007 Carbonyl reductases: The complex relationships of mammalian carbonyl- and quinone-reducing enzymes and their role in physiology. Annu. Rev. Pharmacol. Toxicol. 47: 293–322.

Persson, B., and Y. Kallberg, 2013 Classification and nomenclature of the superfamily of short-chain dehydrogenases/reductases (SDRs), pp. 111–115 in *Chemico-Biological Interactions*,.

Persson, B., Y. Kallberg, U. Oppermann, and H. Jörnvall, 2003 Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs), pp. 271–278 in *Chemico-Biological Interactions*,.

Pičmanová, M., E. H. Neilson, M. S. Motawia, C. E. Olsen, N. Agerbirk *et al.*, 2015 A recycling pathway for cyanogenic glycosides evidenced by the comparative metabolic profiling in three cyanogenic plant species. Biochem. J. 469: 375–389.

R Developement Core Team, 2015 R: a language and environment for statistical computing. R Found. Stat. Comput. 1: 409.

Rane, R. V, T. K. Walsch, S. L. Pearce, L. S. Jermiin, K. H. Gordon *et al.*, 2016 Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores? Curr. Opin. Insect Sci. 13: 70–76.

Rioja, C., V. Zhurov, K. Bruinsma, M. Grbic, and V. Grbic, 2017 Plant-herbivore interactions: a case of an extreme generalist, the two-spotted spider mite, *Tetranychus urticae*. Mol. Plant-Microbe Interact. 30: 935–945.

Ritchie, M. E., J. Silver, A. Oshlack, M. Holmes, D. Diyagama *et al.*, 2007 A comparison of background correction methods for two-colour microarrays. Bioinformatics 23: 2700–

888 2707.
889 Rosenthal, G. A., and M. (May) Berenbaum, 1991 *Herbivores: their interactions with*
890 *secondary plant metabolites.* Elsevier Science.
891 Roy, A., W. B. Walker, H. Vogel, S. Chattington, M. C. Larsson *et al.*, 2016 Diet dependent
892 metabolic responses in three generalist insect herbivores *Spodoptera* spp. Insect
893 Biochem. Mol. Biol. 71: 91–105.
894 Santamaría, M. E., J. González-Cabrera, M. Martínez, V. Grbic, P. Castañera *et al.*, 2015
895 Digestive proteases in bodies and faeces of the two-spotted spider mite, Tetranychus
896 urticae. J. Insect Physiol. 78: 69–77.
897 Schweizer, F., H. Heidel-Fischer, H. Vogel, and P. Reymond, 2017 *Arabidopsis*
898 glucosinolates trigger a contrasting transcriptomic response in a generalist and a
899 specialist herbivore. Insect Biochem. Mol. Biol. 135: 561–573.
900 Shin, R., J. M. An, C. J. Park, Y. J. Kim, S. Joo *et al.*, 2004 *Capsicum annuum* tobacco
901 mosaic virus-induced clone 1 expression perturbation alters the plant's response to
902 ethylene and interferes with the redox homeostasis. Plant Physiol 135: 561–573.
903 Škarydová, L., and V. Wsól, 2012 Human microsomal carbonyl reducing enzymes in the
904 metabolism of xenobiotics: well-known and promising members of the SDR
905 superfamily. Drug Metab. Rev. 44: 173–191.
906 Smyth, G. K., 2004 Linear models and empirical bayes methods for assessing differential
907 expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3: 1–25.
908 Stafford-Banks, C. A., D. Rotenberg, B. R. Johnson, A. E. Whitfield, and D. E. Ullman, 2014
909 Analysis of the salivary gland transcriptome of *frankliniella occidentalis*. PLoS One 9:
910 e94447.
911 Stahl, E., O. Hilfiker, and P. Reymond, 2018 Plant–arthropod interactions: who is the winner?

912 Plant J. 93: 703–728.
913 Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of
914 large phylogenies. Bioinformatics 30: 1312–3.
915 Sterck, L., K. Billiau, T. Abeel, P. Rouzé, and Y. Van de Peer, 2012 ORCAE: online resource
916 for community annotation of eukaryotes. Nat. Methods 9: 1041–1041.
917 Su, Y. L., J. M. Li, M. Li, J. B. Luan, X. D. Ye *et al.*, 2012 Transcriptomic analysis of the
918 salivary glands of an invasive whitefly. PLoS One 7: e39303.
919 Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar, 2013 MEGA6: Molecular
920 evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30: 2725–2729.
921 Thomas, J. H., 2006 Analysis of homologous gene clusters in *Caenorhabditis elegans* reveals
922 striking regional cluster domains. Genetics 172: 127–143.
923 Tibshirani, R., G. Walther, and T. Hastie, 2001 Estimating the number of clusters in a data set
924 via the gap statistic. J. R. Stat. Soc. Ser. B, Stat. Methodol. 63: 411–423.
925 Van Leeuwen, T., and W. Dermauw, 2016 The molecular evolution of xenobiotic metabolism
926 and resistance in chelicerate mites. Annu. Rev. Entomol. 61: 475–498.
927 Van Leeuwen, T., L. Tirry, A. Yamamoto, R. Nauen, and W. Dermauw, 2014 The economic
928 importance of acaricides in the control of phytophagous mites and an update on recent
929 acaricide mode of action research. Pestic. Biochem. Physiol. 121: 12–21.
930 Väremo, L., J. Nielsen, and I. Nookaew, 2013 Enriching the gene set analysis of genome-wide
931 data by incorporating directionality of gene expression and combining statistical
932 hypotheses and methods. Nucleic Acids Res. 41: 4378–4391.
933 Villarroel, C. A., W. Jonckheere, J. M. Alba, J. J. Glas, W. Dermauw *et al.*, 2016 Salivary
934 proteins of spider mites suppress defenses in *Nicotiana benthamiana* and promote mite
935 reproduction. Plant J. 86: 119–131.

936 Voelckel, C., and I. T. Baldwin, 2004 Generalist and specialist lepidopteran larvae elicit
937 different transcriptional responses in *Nicotiana attenuata*, which correlate with larval
938 FAC profiles. Ecol. Lett. 7: 770–775.
939 Vogel, H., R. O. Musser, and M. de la Paz Celorio-Mancera, 2014 Transcriptome responses in
940 herbivorous insects towards host plant and toxin feeding, pp. 197–234 in *Annual Plant*
941 *Reviews: Insect-Plant Interactions*,.
942 Wickham, H., and R. Francois, 2015 dplyr: a grammar of data manipulation.
943 Wink, M., 2010 Annual plant reviews, biochemistry of plant secondary metabolism. Annu.
944 Plant Rev. 40: 483.
945 Wouters, F. C., B. Blanchette, J. Gershenzon, and D. G. Vassão, 2016 Plant defense and
946 herbivore counter-defense: benzoxazinoids and insect herbivores. Phytochem. Rev. 15:
947 1127–1151.
948 Wybouw, N., V. Balabanidou, D. J. Ballhorn, W. Dermauw, M. Grbić *et al.*, 2012 A
949 horizontally transferred cyanase gene in the spider mite *Tetranychus urticae* is involved
950 in cyanate metabolism and is differentially expressed upon host plant change. Insect
951 Biochem. Mol. Biol. 42: 881–889.
952 Wybouw, N., W. Dermauw, L. Tirry, C. Stevens, M. Grbić *et al.*, 2014 A gene horizontally
953 transferred from bacteria protects arthropods from host plant cyanide poisoning. Elife 3:
954 e02365.
955 Wybouw, N., T. Van Leeuwen, and W. Dermauw, 2018 A massive incorporation of microbial
956 genes into the genome of *Tetranychus urticae*, a polyphagous arthropod herbivore. Insect
957 Mol. Biol. 27: 333–351.
958 Wybouw, N., Y. Pauchet, D. G. Heckel, and T. Van Leeuwen, 2016 Horizontal gene transfer
959 contributes to the evolution of arthropod herbivory. Genome Biol. Evol. 8: 1785–1801.

960 Wybouw, N., V. Zhurov, C. Martel, K. A. Bruinsma, F. Hendrickx *et al.*, 2015 Adaptation of
961 a polyphagous herbivore to a novel host plant extensively shapes the transcriptome of
962 herbivore and host. Mol. Ecol. 24: 4647–4663.
963 Xie, W., Q. Wu, S. Wang, X. Jiao, L. Guo *et al.*, 2014 Transcriptome analysis of host-
964 associated differentiation in *Bemisia tabaci* (Hemiptera: Aleyrodidae). Front. Physiol.
965 10: 487.
966 Yuk, H. J., J. H. Lee, M. J. Curtis-Long, J. W. Lee, Y. S. Kim *et al.*, 2011 The most abundant
967 polyphenol of soy leaves, coumestrol, displays potent α-glucosidase inhibitory activity.
968 Food Chem. 126: 1057–1063.
969 Zhang, J., A. M. Dean, F. Brunet, and M. Long, 2004 Evolving protein functional diversity in
970 new genes of *Drosophila*. Proc. Natl. Acad. Sci. 101: 16246–16250.
971 Zhurov, V., M. Navarro, K. A. Bruinsma, V. Arbona, M. E. Santamaria *et al.*, 2014
972 Reciprocal responses in the interaction between *Arabidopsis* and the cell-content-feeding
973 chelicerate herbivore spider mite. Plant Physiol. 164: 384–399.
974

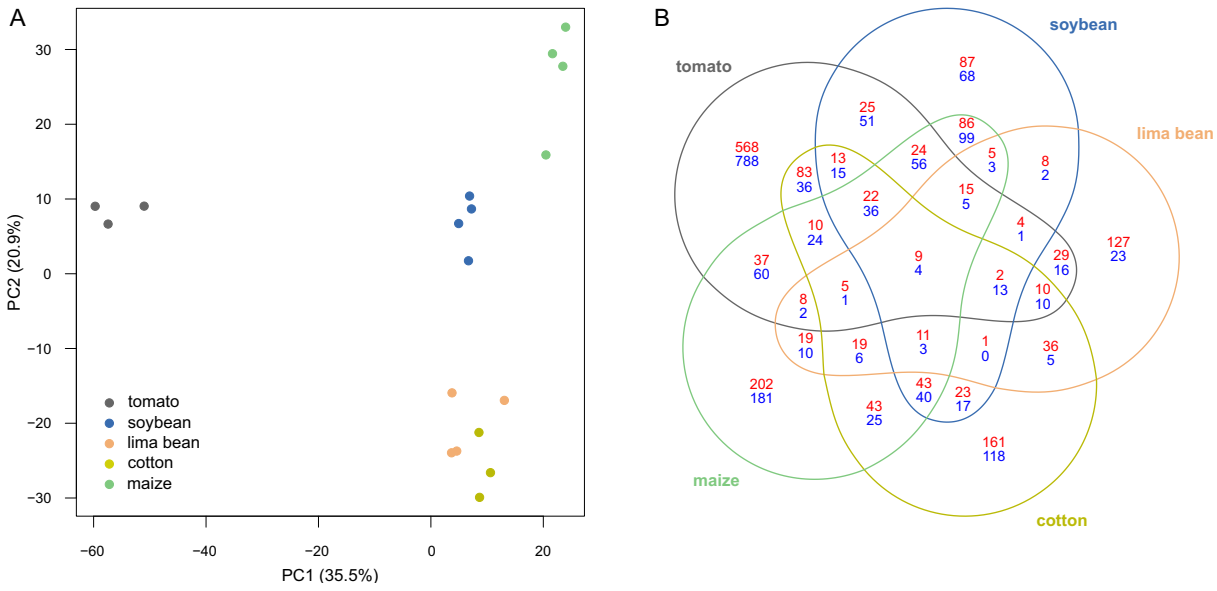

Figure 1 - Principal component analysis (PCA) and differential gene expression of the different host plant populations of *T. urticae*

(A) PCA plot of the relative gene expression levels in *T. urticae* populations after long-term transfer (≥ five generations) from common bean to different host plants: lima bean, soybean, cotton, tomato, and maize. (B) Venn-diagram depicting the overlap among the DEG sets of the populations after long-term transfer (≥ five generations) from common bean to different host plants. Red numbers: upregulated genes, blue numbers: downregulated genes.

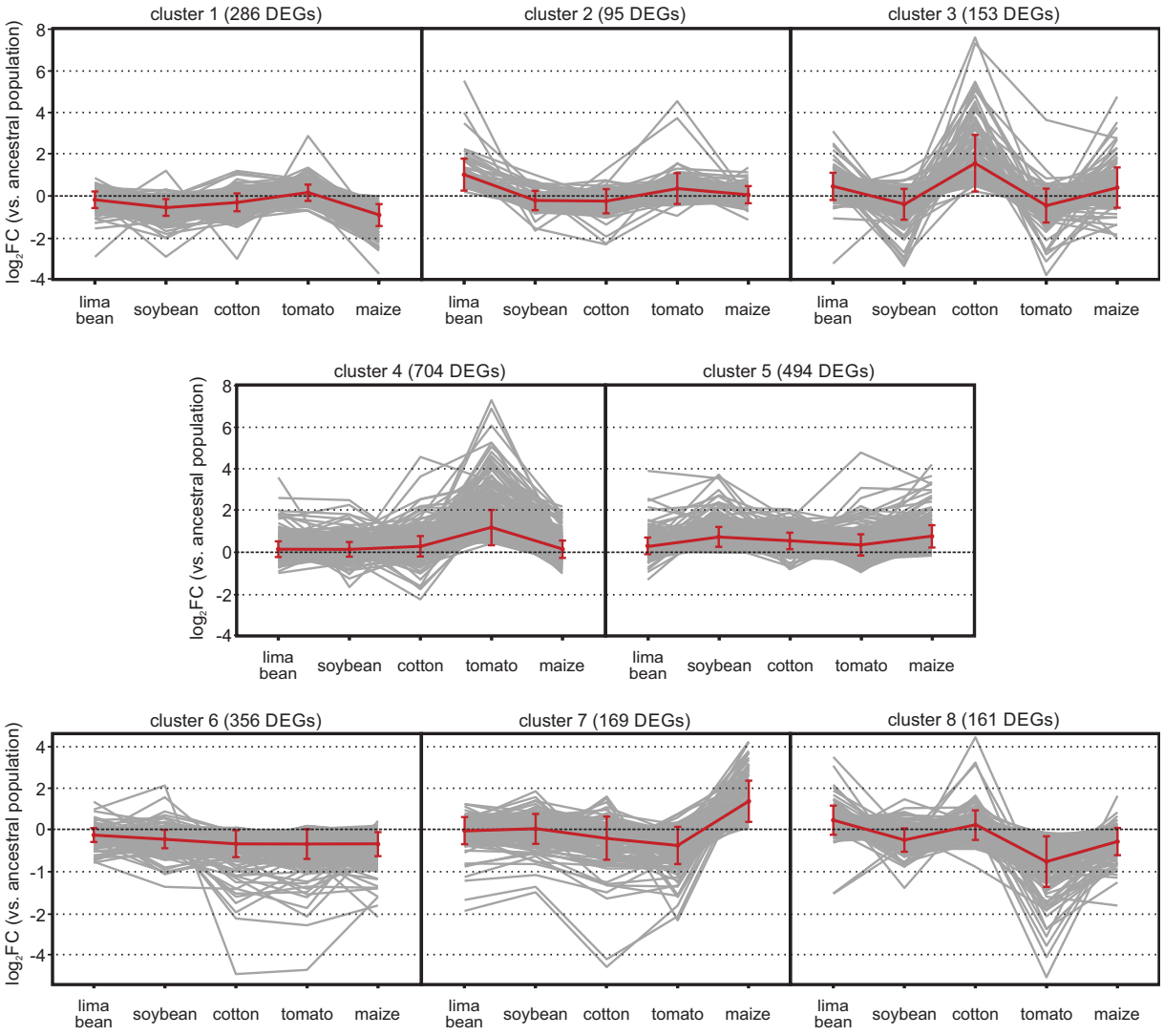

Figure 2 - *k*-means clustering of the *T. urticae* DEGs upon the different long-term host plant transfers

Mite transcriptomic responses to the long-term host plant transfers were categorized into eight clusters using centered Pearson's correlation as the distance metric. Clusters were arranged according to the magnitude of log_2FC of the DEGs. Red lines connect the averages of log_2FC of the different host plant populations within each cluster, with error bars representing the standard deviation.

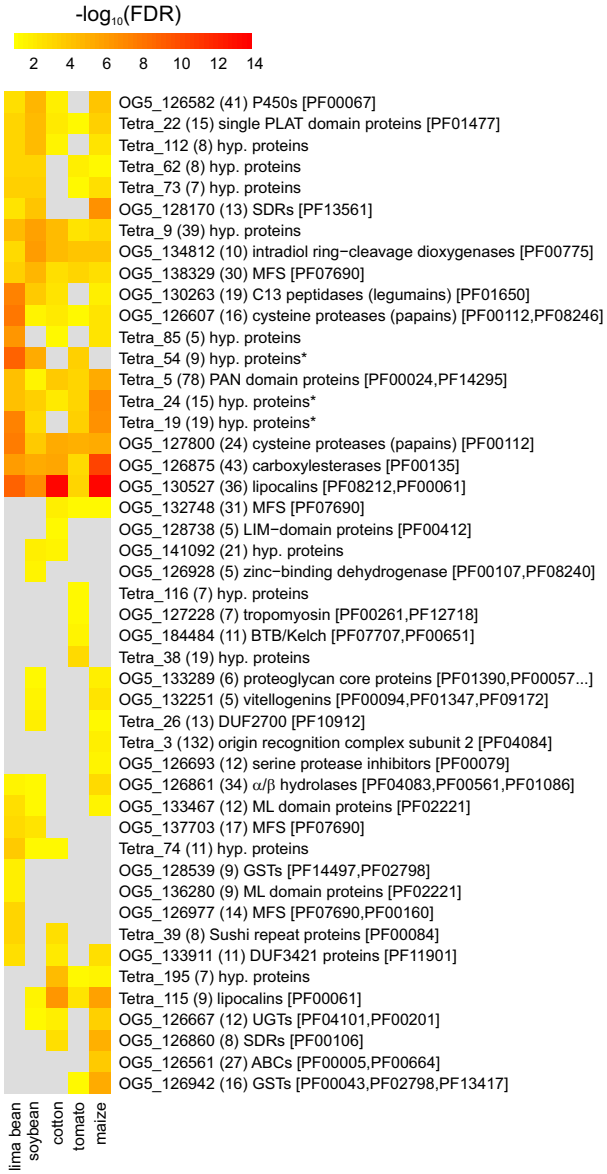

Figure 3 - OrthoMCL enrichment analysis of the DEGs identified in the different host plant populations of *T. urticae*

Heatmap showing the Benjamini-Hochberg corrected p-value (FDR) significance of OrthoMCL groups among DEGs of each host plant population. A grey colored cell indicates that the OrthoMCL group was not significantly enriched (FDR ≥ 0.05) for a certain host plant population. The number between parentheses represents the total number of *T. urticae* genes in an OrthoMCL group (corrected for those genes that have probes on the array), while PFAM accessions associated with any of the genes in a certain OrthoMCL group are shown between square brackets. An asterisk indicates that members of these hypothetical protein OrthoMCL groups were found in the salivary proteome of *T. urticae* (Jonckheere *et al.* 2016).

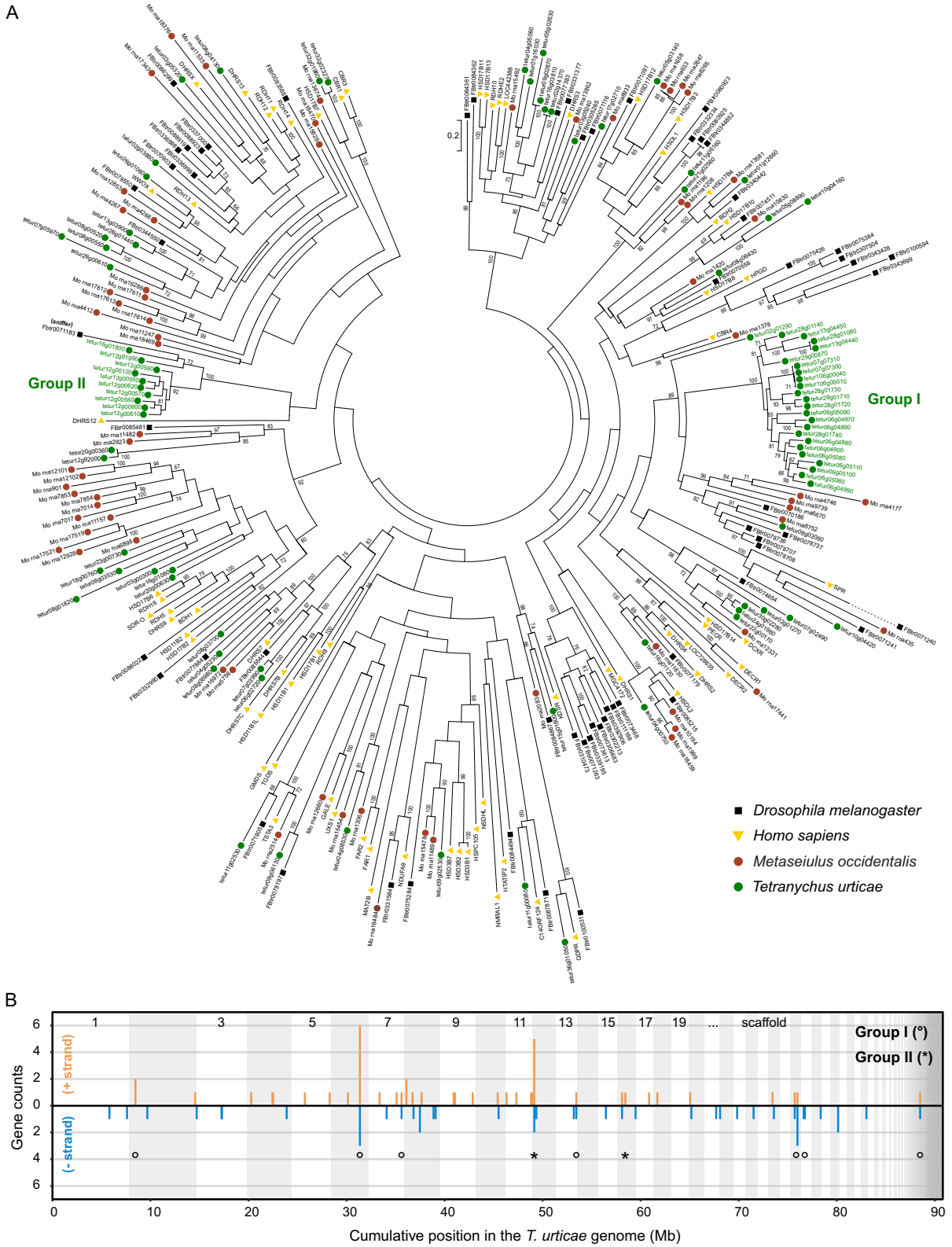

Figure 4 - Maximum likelihood phylogenetic analysis and genomic distribution of *T. urticae* SDRs

(A) Maximum likelihood phylogenetic analysis of the SDRs of *Homo sapiens*, *Drosophila melanogaster*, *Metaseiulus occidentalis* and *Tetranychus urticae*. Only bootstrapping values higher than 65 are shown. The scale bar represents 0.2 amino acid substitutions per site. *T. urticae* SDR expansions containing members of OrthoMCL groups that were significantly enriched among one of the DEG sets of the host plant populations are indicated by green font and labeled as Group I (°) and Group II (*). Branches that were shortened for figure clarity are shown as dashed lines. Information and accession numbers of the used SDRs can be found in Table S7 and File S2. (B) Genomic distribution of *T. urticae* SDRs is shown with lengths of vertical line segments corresponding to counts in a gene cluster; gene counts for the forward (+, orange) and reverse (−, blue) strand orientations. Clusters of SDRs were calculated such that, for a given gene, its count contributes to only one vertical line segment. Only intact SDRs were included in the analysis. Genes of the expansions of Group I and II (see panel A) are marked with their respective symbol. The genome was concatenated from largest to smallest scaffolds for display, alternating scaffolds are indicated by shading.

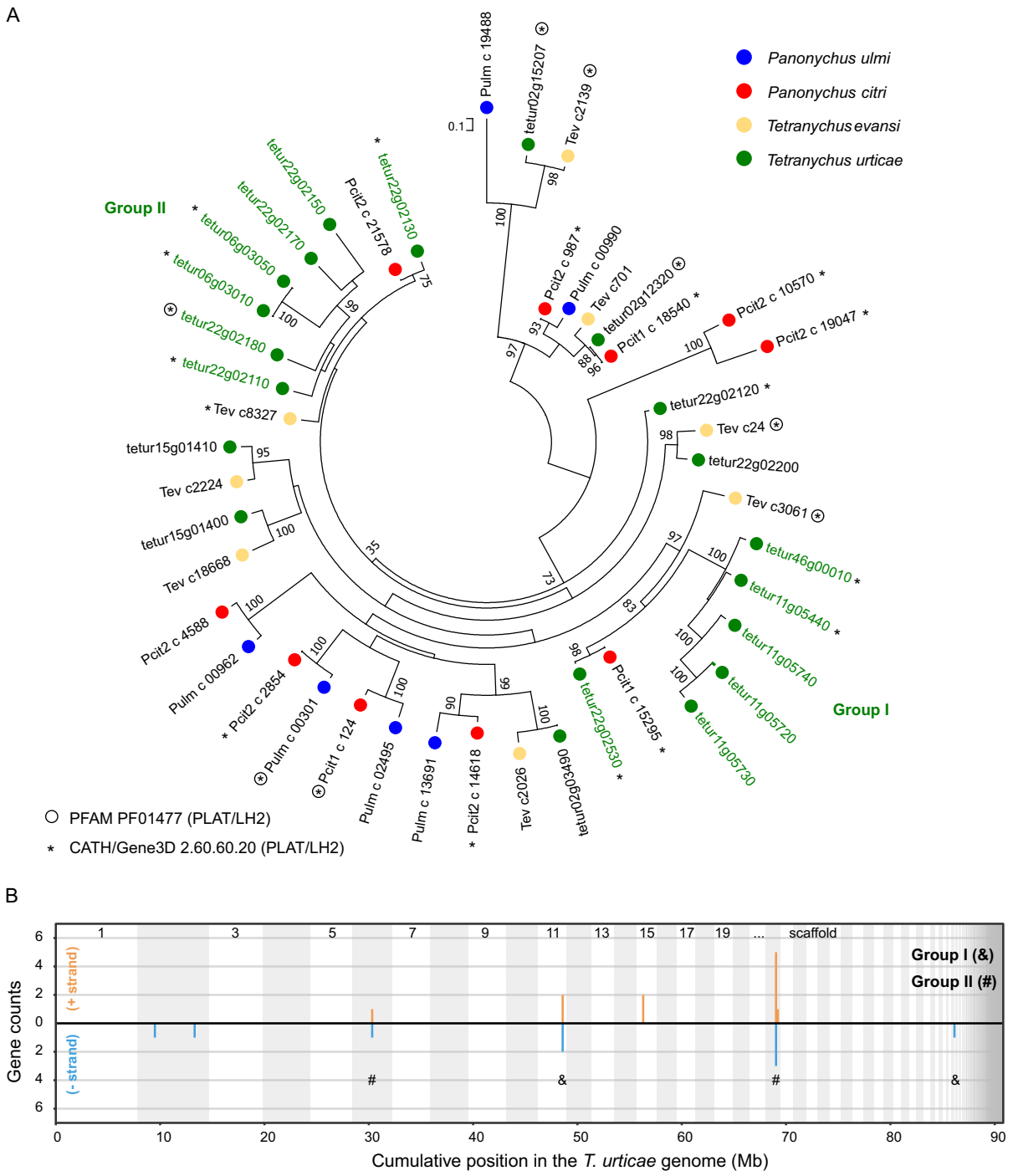

Figure 5 - Maximum likelihood phylogenetic analysis of tetranychid single PLAT domain proteins

(A) Maximum likelihood phylogenetic analysis of the single PLAT domain protein genes of *Panonychus ulmi*, *Panonychus citri*, *Tetranychus evansi* and *Teteranychus urticae*. Only bootstrapping values higher than 65 are shown. The scale bar represents 0.1 amino acid substitutions per site. *T. urticae* single PLAT domain protein expansions are indicated by green font and labeled Group I (&) and Group II (#). Information and accession numbers of the tetranychid single PLAT domain proteins can be found in Table S8 and File S3. Those proteins with the PFAM PLAT domain (PF01477) or belonging to the CATH/Gene3D PLAT/LH2 Superfamily (2.60.60.20) are indicated with a circle and an asterisk, respectively (B) Genomic distribution of *Tetranychus urticae* single PLAT domain protein genes is shown with lengths of vertical line segments corresponding to counts in a gene cluster; gene counts for the forward (+, orange) and reverse (−, blue) strand orientations. Clusters of single PLAT domain protein genes were calculated such that a given gene its count contributes to only one vertical line segment. Only intact single PLAT domain protein genes were included in the analysis. Genes of the expansions of Group I and II (see panel A) are marked with their respective symbol. The genome was concatenated from largest to smallest scaffolds for display, alternating scaffolds are indicated by shading.

Table 1 - Differentially expressed genes in different host plant populations of *T. urticae* (lima bean, soybean, cotton, tomato and maize) compared to an ancestral population on common bean

	total number of DEGs	upregulated DEGs	downregulated DEGs	specific upregulated DEGs* (%)	specific downregulated DEGs* (%)
lima bean	410	307	103	127 (41)	23 (22)
soybean	789	377	412	87 (23)	68 (17)
cotton	842	490	352	161 (33)	118 (34)
tomato	1,982	864	1,118	568 (66)	788 (70)
maize	1,111	557	554	202 (36)	181 (33)

*DEGs specific for a given host plant population