The role of horizontally transferred genes in the xenobiotic adaptations of the spider mite *Tetranychus urticae*

Wybouw, N.R.

_Citation for published version (APA):_
A Link Between Host Plant Adaptation and Pesticide Resistance in the Polyphagous Spider Mite *Tetanychus urticae*


*contributed equally to this study*
Spider Mite Responses to Chemical Pressures

4.0 Abstract

Plants produce a wide range of allelochemicals to defend against herbivore attack, and generalist herbivores have evolved mechanisms to avoid, sequester or detoxify a broad spectrum of natural defense compounds. Successful arthropod pests have also developed resistance to diverse classes of pesticides and this adaptation is of critical importance to agriculture. To test if mechanisms to overcome plant defenses predispose the development of pesticide resistance, we examined adaptation of the generalist two-spotted spider mite, *Tetranychus urticae,* to host plant transfer and pesticides. *Tetranychus urticae* is an extreme polyphagous pest with more than 1100 documented hosts and has an extraordinary ability to develop pesticide resistance. When bean-reared mites from a pesticide-susceptible strain were adapted to a challenging host (tomato), transcriptional responses increased over time with ~7.5% of genes differentially expressed after five generations. While many genes with altered expression belonged to known detoxification families (like P450 monooxygenases), new gene families not previously associated with detoxification in other herbivores showed a striking response, including intradiol ring-cleaving dioxygenase genes acquired by horizontal gene transfer. Strikingly, transcriptional profiles of tomato adapted mites resembled that of multi-pesticide resistant strains, and adaptation to tomato decreased the susceptibility to unrelated pesticide classes. Our findings suggest key roles for both an expanded environmental response gene repertoire and transcriptional regulation in the life history of generalist herbivores. They also support a model whereby selection for the ability to mount a broad response to the diverse defense chemistry of plants predisposes the evolution of pesticide resistance in generalists.

Proceedings of the National Academy of Sciences of the USA 110: E113-E122 (2012)
4.1. INTRODUCTION

Plants produce a wide variety of allelochemicals including a myriad of defense compounds. These can affect herbivore fitness in subtle ways by changing behavior or in less subtle ways by causing acute toxicity. The effectiveness of plant defenses is remarkable as herbivory has evolved successfully in only about one third of all animals (Strong et al., 1984). Nevertheless, herbivores are among the most diverse terrestrial faunas (Futuyma & Agrawal, 2009). The ability to metabolize and detoxify plant chemicals is considered one of the major responses that arthropod herbivores have evolved during their coevolution with plants. Thus, the vast majority of insect herbivores are associated with no more than one or a few plant species (Bernays & Chapman, 1994), potentially reflecting the need for specialized mechanisms to cope with plant chemicals. Herbivorous specialists encounter high levels of predictable toxicants and have often evolved efficient and specialized detoxification systems (Cornell & Hawkins, 2003). A well-known example is the role of CYP6B enzymes in *Papilio* specialists that feed on plants containing toxic furanocoumarins (Berenbaum et al., 1996). These enzymes, belonging to the large CYP family, convert these compounds to non-toxic metabolites and are thought to be a key innovation allowing the ‘escape and radiate’ diversification of Papilionidae (Thompson, 1994).

Plants too can escape and radiate by producing new chemicals that are toxic to herbivores that have not yet evolved an effective detoxification response. An extension of this reasoning is that compounds that have evolved earlier and that are taxonomically widespread should be less toxic than newer compounds, and that specialist herbivores should be less affected than generalists by the toxic compounds of their host plant (Cornell & Hawkins, 2003). This is the ‘jack of all trades, master of none’ argument comparing generalist/specialist ability to cope with plant secondary chemistry (Ali & Agrawal, 2012). However, the way generalist (polyphagous) herbivores cope with the tremendous variety of chemicals in their toxic diet is not well documented in molecular terms. The original assumption was that generalists have a greater capacity to detoxify plant chemicals than specialists (Krieger et al., 1971). This has been refined to state that generalists have detoxification enzymes, in particular CYP enzymes, with broader substrate specificity (Berenbaum et al., 1992). Recently introduced chemical pesticides can be considered as a metaphor for newly evolved or encountered plant chemicals, and a parallel has often been drawn between the evolution of resistance to insecticides and the response to host plant chemicals. This view was presented by Gordon in 1961, who thought that resistance genes are alleles of com-
mon genes, ‘the normal function of which is metabolism of biochemicals present in the diet’ (Gordon, 1961). The ‘pre-adaptation hypothesis’ for insecticide resistance has been supported by surveys of the literature (Croft & Strickler, 1983; Rosenheim et al., 1996) although the comparisons drawn between herbivores and natural enemies, or between chewing and sucking herbivores may be confounded by taxonomy, thus calling for other forms of experimental and observational evidence (Rosenheim et al., 1996). It is now well accepted that herbivore exposure to different plant allelochemicals can affect the toxicity of pesticides (Ahmad, 1986; Berry et al., 1980; Brattsten et al., 1977; Castle et al., 2009; Gould et al., 1982; Kennedy, 1984; Li et al., 2000, 2004; Lindroth, 1989; Sasabe et al., 2004; Sen Zeng et al., 2007; Yang et al., 2001; Yu et al., 1979). Moreover, metabolic resistance to pesticides is known to commonly rely on the increased expression of one or more genes coding for detoxification enzymes and formal evidence that many of these detoxification enzymes can metabolize both plant chemicals and pesticides is accumulating (Feyereisen, 2012; Li et al., 2007). However, it has also been argued that the pattern of selection by plant allelochemicals and by pesticides differs (Despres et al., 2007; Li et al., 2007), so whether the polyphagous nature of many crop pests results in a pre-adaptation potential to cope with pesticides remains conjectural.

To elucidate the relationship between host plant adaptation and pesticide resistance in a systematic way, the two-spotted spider mite, *Tetranychus urticae*, is an excellent choice. *Tetranychus urticae* is among the most polyphagous herbivores known: it can feed on over 1100 different plants in more than 140 different plant families that produce a broad spectrum of chemical defenses (Grbic et al., 2011; Migeon & Dorkeld, 2015). Spider mites have been shown to rapidly adapt to new or less favorable hosts without a correlated fitness cost when compared to the ancestral host (Agrawal, 2000; Magalhães et al., 2009). Moreover, long-term adaptation on a single host does not markedly reduce genetic variation or the capability to subsequently adapt to a different host (Fry, 1989; Magalhães et al., 2007). Also, experimental evolution has shown that although induced plant responses to *T. urticae* herbivory decrease the fitness of non-adapted mites, induced plant response result in higher fitness of adapted mites, suggesting that spider mites can overcome both constitutive and induced plant defenses (Agrawal, 2000). In parallel with an exceptionally broad host range, *T. urticae* has demonstrated an unprecedented ability to develop resistance to pesticides; regardless of the chemical class, the first cases of resistance are usually reported within a few years after the introduction of a new acaricide.
Selection for resistance in *T. urticae* is accelerated by its high fecundity and very short life cycle (Van Leeuwen et al., 2010a), and potentially also its arrhenotokous haplodiploid sex-determination system (unmated females produce haploid males) (Carriere, 2003; Denholm et al., 1998).

To date, studies of resistance in *T. urticae* have focused largely on target site mutations and on classical detoxifying enzyme systems, such as cytochrome P450 monooxygenases (CYPs), carboxyl/cholinesterases (CCEs) and glutathione-S-transferases (GSTs) (Li et al., 2007; Van Leeuwen et al., 2010a). However, these studies have not been satisfactory for understanding the scope of acaricide resistance in *T. urticae*. Under field conditions, multi-resistant strains that are resistant to all commercially available acaricides are often encountered, and strikingly these strains also resist compounds with new modes of action that have never been used in the field (Khajehali et al., 2011). In this study, we have taken advantage of the high quality genome sequence of *T. urticae* (Grbic et al., 2011; Van Leeuwen et al., 2013) to construct an expression microarray that we then used to collect transcriptome data over a time course ranging from hours to generations after transfer of mites to a new, challenging host. We then related changes in gene expression after host plant change to constitutive patterns of gene expression in two strains that are highly resistant to a spectrum of pesticides. In doing so, we defined a set of genes and gene families that are of potential adaptive relevance to both situations. Remarkably, our studies suggest that the polyphagous spider mite exploits a large and shared repertoire of ‘classical’ detoxification genes as well as potential new players as a defense against plant chemicals and pesticides.

### 4.2. MATERIAL AND METHODS

#### 4.2.1. PLANT REARING

Tomato (*Solanum lycopersicum* L. cv. ‘Moneymaker’) and kidney bean (*Phaseolus vulgaris* L. cv. ‘Prelude’) were potted in black earth (Structural Professional, pH 5.0-6.5, 20% organic matter (Snebbout NV, Belgium) and allowed to grow in a growth chamber at 26 °C, 60% relative humidity (RH) and an 16:8 h Light:Dark (L:D) photoperiod. Tomato plants were used for experiments when they had at least four completely developed leaves (about 35 days old), while bean plants were used for either experiments or spider mite rearing when they had two completely developed leaves (about 14 days old).
4.2.2. MITE STRAINS

The London reference strain originates from a wild-collected *T. urticae* population from the Vineland region (Ontario, Canada) and DNA from an inbred line of this strain was used for *T. urticae* genome sequencing (Grbic et al., 2011). This strain is susceptible to commercially available acaricides (Khajehali et al., 2011). The LS-VL laboratory reference strain, originally collected in 2000 near Ghent (Belgium), has been previously described as highly susceptible to acaricides (Van Leeuwen et al., 2005). The MR-VP resistant strain was originally collected from different cultivars of bean plants in a greenhouse at the national botanical garden (Brussels, Belgium) in September 2005, where spider mite control was reported to be extremely problematic. The strain was controlled by regular foliar applications of commercial formulations of the following acaricides: tebufenpyrad, pyridaben, clofentezine, hexythiazox, bifenthrin, fenbutatin oxide, abamectine and oxamyl. Resistance to Mitochondrial Electron Transport Inhibitor (METI) acaricides is well characterised in MR-VP (Van Pottelberge et al., 2009). The Marathonas (MAR-AB) strain was isolated from a heavily sprayed rose greenhouse near Athens (Greece) in 2009. The strain is highly resistant to abamectin, bifenthrin, clofentezine, hexythiazox, fenbutatin oxide and pyridaben. All *T. urticae* strains were mass reared on potted kidney bean in a climatically controlled room at 26 °C, 60% RH and 16:8 h L:D photoperiod. The strains were offered fresh bean plants weekly.

4.2.3. HOST CHANGE EXPERIMENT

For each time point (2 h, 12 h, 80 days) a tomato plant was infested with about 150 female mites from the London strain (cultured on bean plants). Two hours (Tomato-2h), 12 h (Tomato-12h) and 80 days (about 5 generations, Tomato-5G) after infestation, 100 mites were re-collected for RNA extraction. All experiments were performed at 26 °C, 60% RH and an 16:8 h L:D photoperiod and 4 biological replicates were performed for each time point. The Tomato-2h and Tomato-12h experiments were performed during the 16 h light period.

4.2.4. TOXICITY TESTS

Toxicity bioassays on the London strain before and after adaptation to tomato were performed in a similar way as described by Van Leeuwen et al. (2005). First, adult female mites were transferred to square kidney bean foliar discs, placed on wet cotton in a Petri dish. Subsequently, 800 µl of the LC$_{90}$ (lethal
concentration killing 90% of the population) of acaricides with different modes of action (pyridaben (Sanmite 200 mg/g WP), tebufenpyrad (Pyranica 200 g/L EC): mitochondrial complex I electron transport inhibitors; milbemectin (Milbeknock 9.3 g/L EC): chloride channel activator; fenbutatin oxide (Torque 550 g/L SC): inhibitor of mitochondrial ATP synthase and bifenthrin (Talstar 80 g/L SC): sodium channel modulator) was sprayed on the mites. Mites sprayed with double distilled water were used as a control. Finally, bioassays were placed in a climatically controlled room at 26 °C, 60% RH and 16:8 h L:D photoperiod. Four replicates were performed for each strain and for each acaricide. Mortality was assessed 24 h after acaricide application and corrected for control mortality using Abbott's formula (Abbott, 1925). Mites were considered dead when they were drowned or when they did not move after prodding with a fine hair paintbrush.

4.2.5. Microarray experiments and qPCR

4.2.5.1. Microarray construction

A custom Sureprint genome wide G3 Gene Expression 8×60K microarray was designed using the Agilent eArray platform (Agilent Technologies) based on the *T. urticae* genome annotation (version from April 20, 2010). The probe design aimed for three probes of 60 nucleotides per gene with a T_m of 80 °C and parameters set to ‘best probe design’ and ‘3’ bias’. In order to also design gene-specific probes for highly similar genes (e.g. duplicated genes), coding sequences were extended with 100 bp of their predicted 3’ UTR. Where 3’ UTRs were not predicted or predicted shorter, 100 bp downstream the stop-codon were added to the coding sequence. In total 58985 probes were designed. Before the start of GeneSpring analysis (see below), probes were re-mapped (using Bowtie, version 0.12.7; Langmead et al., 2009) on the most recent genome annotation (April 18, 2011, 18455 predicted genes). According to this mapping, 87.4% from the latter were covered by at least one probe, while 81.7% were covered by at least three probes. Standard Agilent features such as spike-ins were added. We selected 182 unique probes that mapped to *T. urticae* genes expressed across four developmental stages, as identified by RNAseq experiments (Grbic et al., 2011), with different ranges of expression (based on normalised read-counts, RPKM). These probes (probe names starting with ‘Rep’) were randomly distributed in 10 to 15 copies per array, and were used to measure array reproducibility. The array design was submitted to NCBI under the GEO-platform format (GPL15756).
4.2.5.2. Target preparation, microarray hybridisation and analysis

Total RNA was extracted from 100 1-3 day old adult female mites using RNeasy mini kit (Qiagen) in 4-6 replicates for each strain (MR-VP, MAR-AB, London) and for each time point (Tomato-2h, Tomato-12h, Tomato-5G) in the host change experiment. Contaminating DNA was removed by digestion with RNase-free Turbo DNase (Ambion). After assessing quality and quantity of the extracted RNA, 100 ng of RNA was used to generate Cy3- and Cy5-labeled cRNA, using the Agilent Low Input Quick Amp Labeling Kit (version 6.5, Agilent Technologies). RNA spike-in controls (Agilent Technologies) were added to RNA samples before cRNA synthesis. The labelled cRNA was purified with the RNeasy mini kit (Qiagen), dye content (> 8.0 pmol dye/μg cRNA) and the concentration of cRNA was measured by NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). Cy3- and Cy5-labeled cRNAs were pooled and hybridized using the Gene Expression Hybridization Kit (Agilent Technologies) for 17 h in a rotating hybridization oven at 20 rpm and 65 °C. Following hybridisation experiments were performed (the number of biological replicates is given between brackets): Cy5 labelled MR-VP cRNA/Cy3 labelled London cRNA (6), Cy5 labelled MAR-AB cRNA/Cy3 London cRNA (5), Cy5 labelled Tomato-2h cRNA/ Cy3 labelled London cRNA (4), Cy5 labelled Tomato-12h cRNA/ Cy3 labelled London cRNA (4) and Cy5 labelled Tomato-5G cRNA/ Cy3 labelled London cRNA (4). After hybridization, slides were washed using the Gene Expression Wash Buffer kit (Agilent Technologies), treated with Stabilization and Drying solution (Agilent Technologies), protected by an Ozone-Barrier cover (Agilent Technologies) until scanned by an Agilent Microarray High Resolution Scanner with default settings for 8 x 60K G3 microarrays. Data were then normalized by the Agilent Feature Extraction software version 10.5 (Agilent Technologies) with default parameter settings for gene expression two-color microarrays (protocol GE2_105_Dec08) and data was transferred to GeneSpring GX 11.0 software (Agilent Technologies) for further statistical evaluation. Experiments were constructed from these microarray data with GeneSpring GX 11.0. Next, probes were flag filtered (only probes that had flag-value ‘Present’ in 50% of all replicates of each experiment were retained) and linked to the most recent annotation file (April 18, 2011) using the ‘Create New Gene-Level Experiment’-option. Genes with a Benjamini-Hochberg FDR corrected p-value < 0.05 and having a more than 2-fold change were considered as differentially expressed. All microarray data sets are accessible through the GEO Series accession number GSE39869.
4.2.5.3. Microarray validation by qPCR

Ten genes – CYP392A16 (tetur06g04520), CYP392D2 (tetur03g04990), CYP392D8 (tetur03g05070), CYP392D10 (tetur03g05110), tetur02g09840 (UDP-glycosyltransferase), tetur16g03200 (MFS transporter), tetur13g04550, tetur01g00490 (intradiol ring-cleaving dioxygenase, ID-RCD), TuGSTd14 (tetur29g00220) and tetur06g04970 (short chain reductase) – that were differentially expressed between both resistant strains (MR-VP and MAR-AB) and the susceptible strain were selected to confirm the microarray results by quantitative realtime PCR (qPCR). To further validate the biological importance of genes identified by microarray experiments, the expression levels of these 10 differentially expressed genes were also determined for a second independent susceptible strain (LS-VL). Mite selection, culture conditions, RNA isolation and DNAse treatment was performed as previously described for microarray experiments. First strand cDNA was synthesized from 2 μg of total RNA using Maxima first strand cDNA synthesis kit (Fermentas). qPCR was done on a Mx3000P real-time PCR system (Stratagene) using Maxima SYBR green qPCR Master Mix (Fermentas) with 2-3 biological replicates and three technical replicates for each gene. Gene specific primers for the 10 differentially expressed genes and two housekeeping genes (actin and ribosomal protein 49) were designed using Primer 3 software (Rozen & Skaletsky, 2000) (TABLE 4.1, primer names with 'q' suffix). Relative expression levels were calculated according to Pfaffl et al. (2001).

4.2.6. Clustering analysis

The GeneSpring GX11.0 software (Agilent Technologies) was used to perform a hierarchical clustering analysis of microarray expression data using the Pearson centered distance metric and complete linkage rule. For gene clustering, the OrthoMCL (Van Dongen, 2000) software version 2.0 with parameters (-v all -te 2 -scheme 7 -l 1.7) was applied with the species combination Drosophila melanogaster, Tribolium castaneum, Acyrthosiphon pisum, T. urticae, Caenorhabditis elegans and Homo sapiens. In this analysis, a total of 103935 sequences clustered into 13876 gene families (76810 genes in clusters; 27125 singletons). Out of these, 3069 clusters contained sequences from all six genomes. Of the protein-coding genes predicted for T. urticae, 11831 were clustered in a total of 6098 groups. OrthoMCL uses a Markov clustering algorithm on a pre-calculated sequence similarity matrix to group (putative) orthologues and paralogues (Li et al., 2003). The matrix was built based on an all-against-all
<table>
<thead>
<tr>
<th>Tetur ID</th>
<th>Gene name/family</th>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>(q)PCR-product (nt)</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tetur18g03590</td>
<td>Ribosomal protein 49</td>
<td>18g03590_q_F</td>
<td>CTTCAAGGCAGCAGCAAGC</td>
<td>105</td>
<td>62.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18g03590_q_R</td>
<td>CGGATCTGACCCCTGAACTCT</td>
<td></td>
<td>62.09</td>
</tr>
<tr>
<td>tetur03g09480</td>
<td>Actin</td>
<td>03g09480_q_F</td>
<td>GCATCCCTCCTTCTGGATGCTT</td>
<td>113</td>
<td>69.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>03g09480_q_R</td>
<td>TCTCGGACAAATTTCTGCTCAAGCA</td>
<td></td>
<td>69.43</td>
</tr>
<tr>
<td>tetur03g04990</td>
<td>CYP392D2</td>
<td>03g04990_q_F</td>
<td>TTTAATTACGGCAGAGGTTAATTG</td>
<td>152</td>
<td>60.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>03g04990_q_R</td>
<td>CACTAATGCTTCTCATGTTGCTT</td>
<td></td>
<td>60.92</td>
</tr>
<tr>
<td>tetur03g05070</td>
<td>CYP392D8</td>
<td>03g05070_q_F</td>
<td>AGACTGCTAGAAAGCGAAAT</td>
<td>100</td>
<td>59.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>03g05070_q_R</td>
<td>CTCGATTTCATGGGTTGCTT</td>
<td></td>
<td>60.07</td>
</tr>
<tr>
<td>tetur03g05110</td>
<td>CYP392D10</td>
<td>03g05110_q_F</td>
<td>ATTAGGATTGCAAGCTCACCAC</td>
<td>109</td>
<td>59.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>03g05110_q_R</td>
<td>GGAATGAAAGGATGTTGCT</td>
<td></td>
<td>58.48</td>
</tr>
<tr>
<td>tetur06g04520</td>
<td>CYP392A16</td>
<td>06g04520_q_F</td>
<td>AAAACCGAGGTCCGCAAGCA</td>
<td>117</td>
<td>59.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>06g04520_q_R</td>
<td>AACGACTTTTCACACGCCATAC</td>
<td></td>
<td>59.92</td>
</tr>
<tr>
<td>tetur06g04970</td>
<td>Short chain reductases</td>
<td>06g04970_q_F</td>
<td>TGCTGAGGCCATGTTGCC</td>
<td>146</td>
<td>61.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>06g04970_q_R</td>
<td>TGAATATGCTTCACTCCTCCG</td>
<td></td>
<td>60.06</td>
</tr>
<tr>
<td>tetur29g00220</td>
<td>TUGSTd14</td>
<td>29g00220_q_F</td>
<td>CTTTGGCAGATCTCCACGGTA</td>
<td>119</td>
<td>60.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29g00220_q_R</td>
<td>GTCAAGGATATGGTCCAGTTG</td>
<td></td>
<td>59.43</td>
</tr>
<tr>
<td>tetur02g09840</td>
<td>UDP-glycosyltransferases</td>
<td>02g09840_q_F</td>
<td>GCTTAACTACGGCAATCTCTTTCGCC</td>
<td>104</td>
<td>59.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>02g09840_q_R</td>
<td>CTGCGAGTAATTCCGTTGCAAATCTC</td>
<td></td>
<td>59.02</td>
</tr>
<tr>
<td>tetur16g03220</td>
<td>MFS</td>
<td>16g03220_q_F</td>
<td>ATATGCGTGATCAGCTGCAAAC</td>
<td>140</td>
<td>59.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16g03220_q_R</td>
<td>AAATGGAACATATGGGCAAGC</td>
<td></td>
<td>59.10</td>
</tr>
<tr>
<td>tetur01g00490</td>
<td>ID-RCD</td>
<td>01g04900_q_F</td>
<td>CGGAAAGCTCCACCACTCAGAG</td>
<td>81</td>
<td>66.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>01g04900_q_R</td>
<td>CGTTTAAAGCTCCGACCAGAGGAGAAG</td>
<td></td>
<td>65.20</td>
</tr>
<tr>
<td>tetur13g04550</td>
<td>ID-RCD</td>
<td>13g04550_q_F</td>
<td>CTCGGCAAGGCAATGTTTCAA</td>
<td>100</td>
<td>59.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13g04550_q_R</td>
<td>ACCCGTGAGAATTCCCTACCA</td>
<td></td>
<td>60.11</td>
</tr>
<tr>
<td>tetur01g00490</td>
<td>ID-RCD</td>
<td>01g04900_s_F</td>
<td>CCTTGGTCTTGGTTGATTACCG</td>
<td>712</td>
<td>59.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>01g04900_s_R</td>
<td>TGATCGTGTTGAACTGTTG</td>
<td></td>
<td>61.71</td>
</tr>
<tr>
<td>tetur04g08620</td>
<td>ID-RCD</td>
<td>04g08620_s_F</td>
<td>TCGCATCCGATGTTGCT</td>
<td>686</td>
<td>59.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>04g08620_s_R</td>
<td>CTGCTGGAGCTCAATCAGAAGC</td>
<td></td>
<td>58.79</td>
</tr>
<tr>
<td>Genie</td>
<td>ID-RCD</td>
<td>Primer 1 (5' to 3')</td>
<td>Primer 2 (5' to 3')</td>
<td>Length</td>
<td>GC Content</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>tetur06g00460</td>
<td>ID-RCD</td>
<td>06g00460_s_F GAAAGACCTGGTAAATTCGTTG</td>
<td>06g00460_s_R TCAAGGATCAATTTCCAAGATG</td>
<td>680</td>
<td>57.80</td>
</tr>
<tr>
<td>tetur07g02040</td>
<td>ID-RCD</td>
<td>07g02040_s_F GCCTGTTGATTACTTCCTTC</td>
<td>07g02040_s_R GGCACCAATTTATTTGACCTATG</td>
<td>737</td>
<td>59.32</td>
</tr>
<tr>
<td>tetur07g05940</td>
<td>ID-RCD</td>
<td>07g05940_s_F CCACTGCAGAGATCCATTG</td>
<td>07g05940_s_R GCCTGTAACCCAATTTGCTG</td>
<td>660</td>
<td>62.63</td>
</tr>
<tr>
<td>tetur12g04671</td>
<td>ID-RCD</td>
<td>12g04671_s_F GATGTAATATGTTGATGTGTC</td>
<td>12g04671_s_R TCTGAAACCATTCTTCTGTG</td>
<td>624</td>
<td>59.14</td>
</tr>
<tr>
<td>tetur19g03360</td>
<td>ID-RCD</td>
<td>19g03360_s_F CTGTATGGTAGAGCAGGATCC</td>
<td>19g03360_s_R GGATCGATACCGAATATCATGG</td>
<td>831</td>
<td>59.49</td>
</tr>
<tr>
<td>tetur20g01160</td>
<td>ID-RCD</td>
<td>20g01160_s_F GTACTGCACTGCAATATG</td>
<td>20g01160_s_R AGATGTTCATTTTGTGCTG</td>
<td>730</td>
<td>60.38</td>
</tr>
<tr>
<td>tetur20g01790</td>
<td>ID-RCD</td>
<td>20g01790_s_F TTGCCTGCTTCCTTGTGAC</td>
<td>20g01790_s_R GACACAAAGTATTTTTGTCTCG</td>
<td>812</td>
<td>58.30</td>
</tr>
<tr>
<td>tetur28g01250</td>
<td>ID-RCD</td>
<td>28g01250_s_F TTTCCTCAATTAGACGATQ</td>
<td>28g01250_s_R TGTTAAGTAGTGACTGTGGATCG</td>
<td>710</td>
<td>58.38</td>
</tr>
<tr>
<td>tetur44g00140</td>
<td>ID-RCD</td>
<td>44g00140_s_F TGAACCTGGTAAATTCGTTG</td>
<td>44g00140_s_R TCCGAGTGATTCTCCTCG</td>
<td>850</td>
<td>59.65</td>
</tr>
</tbody>
</table>
Spider Mite Responses to Chemical Pressures

BLASTp ((Altschul et al., 1990), 2.2.24+; default parameters) and filtered according to the OrthoMCL manual. Where predicted, splice variants were removed from the data set (the longest protein sequence prediction was withheld). OrthoMCL clustering results of *T. urticae* proteins can be accessed at the *T. urticae* genome portal website (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur). Modified Fisher Exact p-values (EASE Score) were calculated (Huang et al., 2009) to measure the gene-enrichment in OrthoMCL clusters of our microarray expression data.

4.2.7. SIGNAL PEPTIDE PREDICTION

The presence of a signal peptide in protein sequences was predicted with SignalP 3.0 (Bendtsen et al., 2004) using Hidden Markov Models (HMM) and Neural Networks (NN). Protein sequences were considered to have a signal peptide under the condition that both models predicted a signal peptide.

4.2.8. GENE FAMILY ANALYSIS

4.2.8.1. Intradiol ring-cleaving dioxygenases (ID-RCDs)

Gene-specific *T. urticae* ID-RCD primers (Table 4.1, primer names with ‘s’ suffix) were designed (using Primer 3 Software (Rozen & Skaltsky, 2000)) in order to amplify a 600-850 bp fragment of ID-RCDs of the closely related spider mite *Tetranychus evansi*. *Tetranychus evansi* genomic DNA was extracted using a phenol chloroform extraction method as described by Khajehali et al. (2011). PCRs were performed in 50 μl reaction volumes with 38.7 μl distilled water; 5 μl 10x PCR buffer, 2 mM MgCl₂, 0.2 mM dNTP, 0.2 μM of each primer; 2 μl of template DNA (± 100 ng/μl) and 0.3 μl of Taq polymerase (Invitrogen) and run on a Biometra Thermocycler Professional (Westburg). PCR conditions were as follows: 2 min 94 °C, 35 × (20 sec 94 °C, 55 sec 49 °C, 2 sec 72 °C) and 5 min 72 °C. All PCR products were separated by electrophoresis on a 1% agarose gel and visualized by ethidium bromide staining. Subsequently, PCR products were purified with the Cycle Pure Kit (Omega Bio-Tek) and cloned into the pJET1.2/blunt vector (Fermentas). After heat-shock transformation of *E. coli* (DH5α) cells, plasmid DNA was obtained by miniprep (using the Plasmid Mini Kit; Omega Bio-Tek) and inserted fragments were sequenced with pJET1.2F and pJET1.2R primers by LGC Genomics. Sequences were deposited in the GenBank database (GenBank accession number JQ736355-JQ736359).
The presence of ID-RCDs in *Panonychus citri*, *Metaseiulus occidentalis*, *Varroa destructor* and *Ixodes scapularis* was determined through a tBLASTn search (Altschul et al., 1990) of *T. urticae* ID-RCD protein sequences against the published transcriptome of *P. citri* (EMBL-EBI accession number: ERP000885), the genome of *M. occidentalis* (Mocc_1.0 assembly), the genome of *V. destructor* (BRL_Vdes_1.0 assembly) and against the genome of *I. scapularis* (ASM20861v1 assembly), respectively.

All protein sequences of *T. urticae* ID-RCDs (Grbic et al., 2011) were used as queries in BLASTp-searches (cut off E-value ≤ e-10) of the NCBI non-redundant protein database. This resulted in 280 unique hits with a protein length between 176 and 500 amino acids. In the dataset obtained, four genera were overrepresented, namely *Aspergillus* (48 sequences), *Streptomyces* (36 sequences), *Rhodococcus* (12 sequences) and *Rhizobium* (9 sequences). All redundant amino acid sequences from these genera were removed for further analysis. Of the resulting 213 protein sequences, 191 were recognized by Conserved Domain Database (Marchler-Bauer et al., 2011) as members of the intradiol dioxygenase-like subgroup (cd03457) and 22 sequences as members of the intradiol dioxygenase superfamily (cl01383). Next, 19 functionally characterized (‘classical’) intradiol ring-cleaving dioxygenases (cd03459-cd03464), belonging to bacteria and fungi, were added to the protein set. Finally, 17 *T. urticae* ID-RCDs and 5 *T. evansi* homologues (see above) were incorporated into the analysis (TABLE S4.1). The amino acid sequence alignment was constructed using MUSCLE (Edgar, 2004). Model selection was done with ProtTest 2.4 (Abascal et al., 2005) and according to the Akaike information criterion (AIC) the model WAG+I+G+F was optimum for phylogenetic analysis. Finally, a maximum likelihood analysis was performed using Treefinder (Jobb et al., 2004), bootstrapping with 1000 pseudoreplicates (LR-ELW). Phylogenetic trees were visualized and edited using MEGA5 (Tamura et al., 2011) and CorelDraw X3 (Corel Inc.), respectively.

4.2.8.2. Lipocalins

Pfam domain searches of our microarray gene expression data revealed that several significantly up- and down-regulated genes contained the lipocalin signature (PF08212.7, PF00061.18). Protein sequences of these genes were used as query in BLASTp searches (Altschul et al., 1990) of the NCBI non-redundant protein database. Each BLASTp search resulted mainly in hits with the highest bitscore for apolipoprotein D proteins of mammals. Next, a reference apolipoprotein D protein sequence of *H. sapiens* (GenBank accession number: P05090) was used as
Spider Mite Responses to Chemical Pressures

query in BLASTp (cut off E-value ≤ e^{-5}) against the proteome of *T. urticae* at the ORCAE website. Finally, the resulting hits were used as query in BLASTp (cut off E-value ≤ e^{-5}) against the *T. urticae* proteome. Using this approach, we identified 67 lipocalin candidates. Pseudogenes and gene fragments were separated from putative full-length lipocalins (58, grouped via OrthoMCL into clusters 10134, 10107, 19721, 19288 and 21421). The latter were together with a selected reference dataset of lipocalin genes (Ganfornina et al., 2006; Table S4.2) aligned using the profile alignment mode of ClustalX and the alignment of Sanchez et al. (2003) as profile (Anderson et al., 2011). Tick lipocalin protein sequences were not included in our phylogenetic analysis in order to decrease the risk of long-branching artefacts (Ganfornina et al., 2006). Model selection was performed with ProtTest 2.4 (Abascal et al., 2005) and according to the AIC the model WAG+G+F was optimum for phylogenetic analysis. Finally, a maximum likelihood analysis was performed using Treefinder (Jobb et al., 2004), bootstrapping with 500 pseudoreplicates (LR-ELW). Phylogenetic trees were visualized and edited using MEGA5 (Tamura et al., 2011) and CorelDraw X3 (Corel Inc.), respectively. Secondary structures of *T. urticae* lipocalins were predicted using Jpred 3 (Cole et al., 2008) while GPI-anchor sites were predicted using PredGPI (Pierleoni et al., 2008).

4.2.8.3. Major Facilitator Superfamily (MFS) transporters

Pfam domain searches of our microarray gene expression data revealed the up- and down-regulation of genes containing the MFS signature (PF07690.11). Most differentially expressed MFS genes grouped into three OrthoMCL clusters: 10032, 10082 and 10236. To determine the MFS class of genes in these clusters, protein sequences were used as query in BLASTp-searches in the Transporter Classification DataBase (Saier et al., 2009). Transmembrane regions were predicted using TMHMM Server v. 2.0. (http://www.cbs.dtu.dk/services/TMHMM/).

4.3. Results

4.3.1. Effect of Host Plant Shift on Gene Expression

To examine genome-wide patterns of gene expression in *T. urticae*, we constructed an expression microarray (using the Agilent eArray platform, see Material and methods) with long oligonucleotide probes against all predicted genes of the London reference strain. We then used this array to examine expression changes
associated with host plant shift, as well as expression patterns in acaricide resistant strains. For the host plant change experiment, we transferred 1-3 day old females (London strain) from their common host, *P. vulgaris* (kidney bean), to a more challenging and less accepted host, *S. lycopersicum* (tomato). We used young females because this stage actively disperses with the wind to escape kin competition and over-exploitation (Bitume et al., 2011; Kennedy & Smitley, 1985), and hence are expected to encounter potentially less favorable plants, on which they must immediately feed to produce eggs for colony establishment (colonies can then persist for multiple generations). We followed transcriptional changes over the short term to understand the initial early responses, as well as after five generations on the new host. Briefly, female mites grown on beans were transferred to tomato, and transcriptional responses of mites were assessed at 2 h (Tomato-2h), 12 h (Tomato-12h) and after propagation for five consecutive generations (Tomato-5G). As assessed by the number of differentially expressed genes (\(|\log_2(FC)| \geq 1\), Benjamini-Hochberg FDR-corrected p-value < 0.05), the transcriptional response increased with time. Thirteen and 416 genes were differentially expressed after 2 and 12 h, respectively, while 1206 or about 7.5% of all predicted genes with probes on the array were differentially expressed after five generations (Figure 4.1A). There was little overlap between genes associated with the early responses (Tomato-2h and Tomato-12h) and those with changed expression after five generations (Tomato-5G) (i.e., only 8.3% of Tomato-5G was shared with Tomato-2h and Tomato-12h) (Figure 4.1A).

**Figure 4.1.** A: Venn-diagram depicting overlap of the differentially expressed genes of mites exposed to tomato at different time points (2 h, 12 h and 5 generations), relative to mites on bean. B: Venn-diagram showing overlap of the differentially expressed genes of two acaricide multi-resistant mite strains (MAR-AB and MR-VP) and mites exposed to tomato for five generations, using the susceptible London strain on bean as a common reference. Differential expression was identified by \(|\log_2(FC)| \geq 1\) and FDR-corrected p-value < 0.05. Blue: up-regulated genes; orange: down-regulated genes.
4.3.2. Effect of acaricide resistance on gene expression

To relate patterns of response between host plant adaptation and evolved pesticide resistance, we assessed gene expression patterns between two highly resistant field collected strains (MR-VP and MAR-AB) and the reference susceptible London strain (Khajehali et al., 2011). These field strains, one collected on beans and the second collected on roses (see Material and methods), are resistant when grown on bean, the host we used for assessing transcriptome variation among the 3 strains. Using the London strain as a reference, we observed differences in transcript levels (|log2(FC)| ≥ 1, FDR-corrected p-value < 0.05) for 893 and 977 genes for MR-VP and MAR-AB, respectively (FIGURE 4.1B). Earlier work has shown that mite strains can be genetically diverse (Grbic et al., 2011; Van Leeuwen et al., 2012), potentially confounding comparison of gene expression across strains (polymorphisms can affect array-hybridization). However, our long oligo array is expected to be relatively robust to SNP and small indel changes (Hughes et al., 2001); and more importantly, we validated with qPCR (FIGURE S4.1) a subset of genes predicted from the array to be differentially expressed between the London strain and a second susceptible strain (LS-VL, Van Leeuwen et al., 2005) with a different genetic background.

4.3.3. Relationships among transcriptome profiles

Although the resistant strains are genetically unrelated, there was an overlap of 415 differentially expressed genes (46.5 and 42.5% of the total number of differentially expressed genes in MR-VP and MAR-AB, respectively). Further, we found that 49.5 and 42.3% of differentially expressed genes in the resistant MR-VP and MAR-AB strains were also differentially expressed after the London strain was transferred to tomato for five generations. A scatter plot of the fold changes for the intersection of differentially expressed genes between the host transfer and resistance datasets revealed a high correlation for gene expression levels (Spearman correlation: r = 0.740, p < 0.001, FIGURE 4.2A).

Further, hierarchical clustering (Pearson centered distance metric, complete linkage rule) across all the expression data revealed that the expression signatures for the two resistant strains and for mites feeding on tomato for five generations clustered together, and not with early mite responses to tomato transfer after 2 or 12 h. The overlap between resistance and host plant change was even more striking when genes were grouped in gene (sub)families by OrthoMCL clustering (TABLE 4.2). Shared responses were largely mediated by a few gene families, and in some cases a large proportion of all family members
responded. As revealed by PFAM-domain searches of OrthoMCL clusters (48% of all genes in shared OrthoMCL clusters (TABLE 4.2) had an assigned PFAM-domain with E-value ≤ e^{-5}), some responsive families belonged to those commonly implicated in detoxification or transport of xenobiotics [e.g., CCEs, CYPs, GSTs and ABC-transporters (ABC-B/Cs)]. Among these, CYP genes stood out as being markedly differentially expressed among resistant mites and after host transfer of the susceptible London strain (FIGURE 4.2B).

Intriguingly, some of the most strongly affected gene families in both experiments had signatures that have, until now, not been commonly associated with response to xenobiotics in arthropods. To shed insights into the *T. urticae* polyphagous life history, we therefore examined the composition and the nature of transcriptional responses for such families of moderate size (10 members or more). We note, however, that genes of unknown function had some of the most striking expression changes of which the encoded products were commonly predicted to be secreted (TABLE 4.2). An example is OrthoMCL cluster 10257 for which tetur11g05420, tetur11g05450 and tetur46g00020 were up-regulated by ~700-fold upon host transfer to tomato for 5 generations.

4.3.4. INTRADIOL RING-CLEAVING DIOXYGENASES

A set of 17 genes encoding secreted proteins identified as intradiol ring-cleaving dioxygenases (ID-RCDs) were among the most prominent differentially expressed in our analysis (TABLE 4.2). These genes, belonging to the ‘intradiol dioxygenase like’ subgroup (cd03457) according to the Conserved Domain Database (Marchler-Bauer et al., 2011), were recently identified as a case of horizontal gene transfer in the genome of *T. urticae*, and have not been reported in other metazoan genomes to date (Grbic et al., 2011). More than half of the genes in this family were differentially expressed upon host plant change and in multi-resistant strains, and their expression patterns were highly correlated (FIGURES 4.2B and 4.3). ID-RCDs catalyze the oxygenolytic fission of catecholic substances, allowing bacteria and fungi to degrade aromatic rings, a crucial step in the global carbon cycle. Although bacteria usually harbor only 1 to 4 ID-RCD genes, this family has proliferated in *T. urticae* to 16 complete ID-RCDs and a pseudogene (tetur07g06560). Spider mite ID-RCDs share the conserved 2 His 2 Tyr non-heme iron (III) active site with previously described and functionally characterized ID-RCDs (such as catechol, hydroxyquinol and protocatechuate ID-RCDs) (Ferraroni et al., 2005; Grbic et al., 2011; Matera et al.,
The ID-RCD genes are distributed over 11 genomic scaffolds and all but one (tetur07g02040) were intronless. Clusters of duplicated *T. urticae* ID-RCD genes were found on several scaffolds (FIGURE 4.3). We detected by PCR five orthologous genes in the closely related species, *T. evansi*, an oligophagous specialist of Solanaceae (FIGURES 4.3 and 4.5). We also found ID-RCD sequences (E-value ≤ 2e-36) in the RNAseq data from the citrus red spider mite, *Panonychus citri* (EMBL-EBI accession number: ERP000885). However, we found no trace of their presence in the genomes of other, non plant-feeding Acari, such as *M. occidentalis* (a predatory mite), *V. destructor* (an ectoparasite of bees) or *I. scapularis* (a blood-feeding tick). This suggests that a horizontal gene transfer occurred after the split of these lineages from the Tetranychidae. Phylogenetic analysis revealed that spider mite ID-RCDs clustered with a group of fungal ID-RCDs that share a common ancestor with plant and entomopathogenic bacterial ID-RCDs (such as *Xenorhabdus* sp. and *Photorhabdus* sp. (Goodrich-Blair & Clarke, 2007)) (FIGURE 4.5). None of these fungal ID-RCDs belong to previously characterized ‘classical’ fungal and bacterial ID-RCDs known to metabolize well-characterized substrates such as catechol and protocatechuate. In contrast to the characterized cytoplasmic enzymes, this large clade of predicted secreted forms of ID-RCDs (FIGURE 4.5, TABLE S4.1) has not yet been recognized and thoroughly characterized, although we found these proteins in proteomic data on fungal secretomes (Paper et al., 2007; Yang et al., 2012).
### Table 4.2. Expression changes within gene clusters of pesticide resistance and host plant shift.

<table>
<thead>
<tr>
<th>OrthoMCL Members</th>
<th>% Sp</th>
<th>Gene family</th>
<th>PFAM domain(s)</th>
<th>MR-VP</th>
<th>MAR-AB</th>
<th>Tom-5G</th>
</tr>
</thead>
<tbody>
<tr>
<td>10134</td>
<td>25</td>
<td>Lipocalins</td>
<td>PF08212.7, PF00061.18</td>
<td>68.0*</td>
<td>68.0*</td>
<td>64.0*</td>
</tr>
<tr>
<td>10364</td>
<td>16</td>
<td>Low-density lipoprotein receptors</td>
<td>PF00057.13</td>
<td>62.5*</td>
<td>56.3*</td>
<td>68.8*</td>
</tr>
<tr>
<td>10287</td>
<td>17</td>
<td>Intradiol ring-cleaving dioxygenases</td>
<td>PF00775.16</td>
<td>52.9*</td>
<td>52.9*</td>
<td>70.6*</td>
</tr>
<tr>
<td>10012</td>
<td>52</td>
<td>P450 monoxygenases</td>
<td>PF00067.17</td>
<td>51.9*</td>
<td>65.4*</td>
<td>36.5*</td>
</tr>
<tr>
<td>11149</td>
<td>10</td>
<td>Hypothetical proteins</td>
<td>-</td>
<td>50.0*</td>
<td>30.0</td>
<td>70.0*</td>
</tr>
<tr>
<td>11121</td>
<td>10</td>
<td>Hypothetical proteins</td>
<td>-</td>
<td>10.0</td>
<td>70.0*</td>
<td>60.0*</td>
</tr>
<tr>
<td>10289</td>
<td>18</td>
<td>Hypothetical proteins</td>
<td>-</td>
<td>27.8*</td>
<td>44.4*</td>
<td>66.7*</td>
</tr>
<tr>
<td>10107</td>
<td>30</td>
<td>Lipocalins</td>
<td>PF08212.7, PF00061.18</td>
<td>43.3*</td>
<td>26.7*</td>
<td>66.7*</td>
</tr>
<tr>
<td>10176</td>
<td>24</td>
<td>Hypothetical proteins</td>
<td>-</td>
<td>50.0*</td>
<td>50.0*</td>
<td>29.2*</td>
</tr>
<tr>
<td>10422</td>
<td>15</td>
<td>Hypothetical proteins</td>
<td>-</td>
<td>33.3*</td>
<td>40.0*</td>
<td>40.0*</td>
</tr>
<tr>
<td>10254</td>
<td>19</td>
<td>Kelch-related proteins</td>
<td>PF00651.26, PF07707.10</td>
<td>52.6*</td>
<td>31.6*</td>
<td>26.3*</td>
</tr>
<tr>
<td>10475</td>
<td>14</td>
<td>Hypothetical proteins</td>
<td>-</td>
<td>21.4</td>
<td>64.3*</td>
<td>21.4</td>
</tr>
<tr>
<td>10257</td>
<td>19</td>
<td>Hypothetical proteins</td>
<td>-</td>
<td>31.6*</td>
<td>26.3*</td>
<td>47.4*</td>
</tr>
<tr>
<td>10236</td>
<td>20</td>
<td>Mfs</td>
<td>PF07690.11</td>
<td>15.0</td>
<td>40.0*</td>
<td>50.0*</td>
</tr>
<tr>
<td>10115</td>
<td>16</td>
<td>δ gsts</td>
<td>PF00043.20, PF13417.1</td>
<td>12.5</td>
<td>56.3*</td>
<td>31.3*</td>
</tr>
<tr>
<td>10041</td>
<td>29</td>
<td>Short-chain reductases</td>
<td>PF13561.1, PF00106.20</td>
<td>13.8</td>
<td>31.0*</td>
<td>51.7*</td>
</tr>
<tr>
<td>10032</td>
<td>53</td>
<td>Mfs</td>
<td>PF07690.11</td>
<td>18.9*</td>
<td>32.1*</td>
<td>45.3*</td>
</tr>
<tr>
<td>10420</td>
<td>15</td>
<td>Hypothetical proteins</td>
<td>-</td>
<td>20.0*</td>
<td>46.7*</td>
<td>26.7</td>
</tr>
<tr>
<td>10831</td>
<td>11</td>
<td>Cysteine proteases</td>
<td>PF08127.8, PF00112.18</td>
<td>18.2</td>
<td>27.3</td>
<td>45.5*</td>
</tr>
<tr>
<td>10689</td>
<td>12</td>
<td>Cuticular proteins</td>
<td>PF00379.18</td>
<td>16.7</td>
<td>41.7*</td>
<td>25.0</td>
</tr>
<tr>
<td>10040</td>
<td>51</td>
<td>NYN-domain family</td>
<td>PF01936.13</td>
<td>29.4*</td>
<td>21.6*</td>
<td>31.4*</td>
</tr>
<tr>
<td>10008</td>
<td>11</td>
<td>P450 monoxygenases</td>
<td>PF00067.17</td>
<td>9.1</td>
<td>36.4*</td>
<td>36.4*</td>
</tr>
<tr>
<td>10841</td>
<td>11</td>
<td>Hypothetical proteins</td>
<td>DUF1768</td>
<td>27.3</td>
<td>36.4*</td>
<td>18.2</td>
</tr>
<tr>
<td>10085</td>
<td>27</td>
<td>Alpha/beta hydrolase fold-family</td>
<td>PF04083.11, PF00561.15</td>
<td>29.6*</td>
<td>18.5</td>
<td>33.3*</td>
</tr>
<tr>
<td>10204</td>
<td>21</td>
<td>Hypothetical proteins</td>
<td>-</td>
<td>14.3</td>
<td>19.0</td>
<td>47.6*</td>
</tr>
<tr>
<td>10157</td>
<td>25</td>
<td>Hypothetical proteins</td>
<td>-</td>
<td>40.0*</td>
<td>12.0</td>
<td>28.0*</td>
</tr>
<tr>
<td>10158</td>
<td>24</td>
<td>Carboxyl/cholinesterases</td>
<td>PF00315.23</td>
<td>12.5</td>
<td>29.2*</td>
<td>37.5*</td>
</tr>
<tr>
<td>10074</td>
<td>40</td>
<td>Carboxyl/cholinesterases</td>
<td>PF00315.23</td>
<td>22.5*</td>
<td>30.0*</td>
<td>25.0*</td>
</tr>
</tbody>
</table>
4.3.5. Lipocalins

We also found that lipocalins, small proteins capable of binding to hydrophobic molecules, were strongly differentially expressed (68% of genes of OrthoMCL cluster 10134) between London and resistant strains, and dynamically over time to host plant change (TABLE 4.2, FIGURE S4.2). As revealed by hierarchical clustering, lipocalin expression patterns of resistant strains and mites feeding on tomato for five generations grouped together. Some of these lipocalin genes were strongly and progressively induced in mites feeding for 2 and 12 h on tomato, but were completely down-regulated after five generations, while stable induced expression was maintained for other lipocalin genes.
Spider Mite Responses to Chemical Pressures

**Figure 4.3.** Phylogenetic relationship of the *T. urticae* ID-RCD family linked to expression patterns [\(\log_2(FC)\)] in acaricide multi-resistant strains (MR-VP and MAR-AB) and after host plant shift to tomato for five generations. Genes with detected orthologues in the related spider mite *T. evansi* are depicted with an e.

**Figure 4.4.** Alignment of conserved residues in 'classical' and secreted intradiol ring-cleaving dioxygenase (ID-RCD) amino acid sequences. Classical ID-RCDs are categorized according to their substrate; CTD: catechol, HQD: hydroxyquinol and PCD: protocatechuate. The secreted ID-RCD is labelled with IDL: intradiol dioxygenase like (cd03457). Species abbreviations; Aci: *Acinetobacter* sp., R. opa: *Rhodococcus opacus*, N. sim: *Nocardia simplex*, and A. fum: *A. fumigatus*. *Tetur*07g02040, *tetur*13g04550 and *tetur*20g01790 are ID-RCD representatives of *T. urticae*. The 2 His 2 Tyr non-heme iron (III) binding sites are indicated by gray shading. Residues defined by crystallographic studies (Ferraroni et al., 2005; Matera et al., 2010; Vetting & Ohlendorf, 2000) to have an influence on substrate interaction in 'classical' ID-RCDs (CTD, HQD and PCD) are indicated by black dots. The predicted binding residues of epicatechin in the protein sequence of *A. fumigatus* are indicated in bold (Roopesh et al., 2012).
Of the 58 complete lipocalins we annotated in the *T. urticae* genome, half were concentrated on only three scaffolds (20 on scaffold 6, 8 on scaffold 1, and 5 on scaffold 31) (Table S4.3). The number of *T. urticae* lipocalins far exceeded those reported in insects (*D. melanogaster*: 4, *Apis mellifera*: 4, *Rhodnius prolixus*: 22) (Ganfornina et al., 2006) and in humans (10; Breustedt et al., 2006) but was in the same range as in ticks (Ribeiro et al., 2006). Thirty-six (62.0%) *T.*

**Figure 4.5.** Maximum likelihood unrooted tree of 17 *T. urticae* and five *T. evansi* ID-RCDs with 232 bacterial and fungal sequences. Color codes indicate the percentage of predicted secretion within each clade; yellow: not secreted, blue: < 55%, green: 55-85%, and orange: > 85% secreted. All members of the *T. urticae* clade were predicted to be secreted and formed a sister clade to fungal secreted dioxygenases, sharing a most recent common ancestor with plant and entomopathogenic Proteobacteria. The ‘classical’ biochemically characterized ID-RCDs (CTD, PCD and HQD) were not secreted and clustered together as an outgroup. The phylogenetic positions of the ID-RCD protein sequences of *Naegleria gruberi* (Protozoa), *Schistosoma mansoni* (Metazoa), *P. infestans* (oomycete) and *Haloferax volcanii* (Archaea) are highlighted by *, **, *** and ****, respectively.
urticae lipocalins were predicted to have an anti-parallel β-barrel, while 25.9% (n = 15) had only a small deviation from the canonical lipocalin secondary structure (Flower et al., 2000) (TABLE S4.3). T. urticae lipocalins did not have a GPI-anchor signal omega site and were, with the exception of tetur31g00780, predicted to have a signal peptide (TABLE S4.3).

Two main types of lipocalin gene organization were apparent. Thirty-seven genes had five exons and an intron phase pattern of 0-2-1-1 (corresponding to the arthropod lipocalin gene consensus pattern), while 18 lipocalin genes had only four exons and a 0-1-1 intron phase pattern, a gene structure also reported for a moth lipocalin (Sanchez et al., 2006; TABLE S4.3 and FIGURE S4.2C). Most T. urticae lipocalins clustered together with a previously described clade (Ganfornina et al., 2000; Wade et al., 2009) comprising vertebrate apolipoprotein D (ApoD) and crustacyanins with a high bootstrap support (83%) (FIGURE S4.2B, TABLES S4.2 and S4.3). Within this large T. urticae lipocalin clade, most lipocalins clustered according to their intron phase pattern (FIGURES S4.2B and S4.2C). However, five T. urticae lipocalins (tetur01g01500, tetur01g01510, tetur01g01520, tetur01g16584 and tetur02g09610) grouped together with Karl, a lipocalin expressed in the blood-cells of D. melanogaster, and two (tetur174g00050 and tetur07g03970) were closely related to insect biliproteins (FIGURE S4.2).

4.3.6. TRANSPORTERS OF THE MAJOR FACILITATOR SUPERFAMILY

Among the genes differentially expressed in both MR-VP, MAR-AB and Tomato-5G (TABLE 4.2), three OrthoMCL clusters (10082, 10032 and 10236) with a total of 109 genes shared the PFAM-domain PF07690.11 that characterizes the Major Facilitator Superfamily (MFS), also named the uniporter-symporter-antiporter family. Members of OrthoMCL clusters 10236 and 10082 were highly similar (E-value ≤ e^{-17} and ≤ e^{-5}, respectively) to the Anion/Cation Symporter (ACS) family and the Na^+ dependent glucose transporter family, respectively (TABLE S4.4). On the other hand, most members of cluster 10032 showed similarity (E-value ≥ e^{-10}, TABLE S4.4) to bacterial tetracycline:H^+ antiporters and their mammalian homologues, the heme-carrier proteins/thymic-folate cotransporters (Shayeghi et al., 2005). All differentially expressed members of OrthoMCL cluster 10032 were up-regulated in both MR-VP and MAR-AB, and 87.5% of the differentially expressed members were also up-regulated on tomato for five generations. Teturi11g05410 was more than 300-fold up-regulated by transfer to tomato. Moreover, 16 members of cluster 10032 were already up-regulated in mites transferred from bean to tomato for 12 h.
4.3.7. TRANSCRIPTION FACTORS

Transcription factors belonging to the nuclear receptor family, the bHLH-PAS family, as well as to the bZIP family are known to be involved in the response to stress and xenobiotics in vertebrates and insects (Misra et al., 2011; Pascussi et al., 2008). The *T. urticae* genome harbors at least 700 transcription factors (Grbic et al., 2011), and in a hierarchical clustering analysis (Pearson centered distance metric, complete linkage rule) with transcription factor expression data, Tomato-5G clustered with MR-VP and MAR-AB, and formed a sister clade to Tomato-24h and Tomato-12h. Seventeen, 20 and 27 transcription factors were differentially expressed (|log2(FC)| ≥ 1, FDR-corrected p < 0.05) in MR-VP, MAR-AB and Tomato-5G, respectively, although only four (*tetur03g03150, tetur07g01800, tetur24g02450 and tetur36g00260*) were shared between MR-VP, MAR-AB and Tomato-5G (Table S4.5).

*Tetur36g00260* belongs to the class of nuclear receptors and is one of the eight paralogues of the vertebrate xenosensors PXR and CAR found in the spider mites (Grbic et al., 2011). In *Drosophila*, the xenosensor CncC is down-regulated by Keap1, a Kelch-like protein (Misra et al., 2011). Intriguingly, we found a large proportion of genes in the OrthoMCL cluster 10254, with PFAM-domains associated with Kelch-like proteins (Table 4.2), to be differentially expressed in Tomato-5G, MR-VP and MAR-AB, with almost all (83-100%) of these genes down-regulated. Four genes were down-regulated under all three conditions, in particular *tetur24g00770* (down-regulated 34.8, 7.5 and 2.9 times in Tomato-5G, MR-VP and MAR-AB, respectively). We also searched for homologues of the *Drosophila* xenosensor CncC and identified two paralogues of this bZIP transcription factor in the *T. urticae* genome (*tetur07g06850 and tetur07g04600*, E < e^-31 using CncC of *D. melanogaster* (GenBank accession number AAN13930) as query). However, neither significantly changed transcription in our microarray analysis.

4.3.8. EFFECT OF HOST PLANT SHIFT ON ACARICIDE TOXICITY

As the expression profiles were highly correlated between tomato adapted mites and pesticide-resistant strains, the question was raised whether acaricide toxicity was affected by host-plant shift. We chose five acaricides with four different modes of action and compared their toxicity on susceptible (London) mites kept on bean versus tomato adapted mites. The adaptation to tomato was accompanied by a significant decrease in toxicity for three of the acaricides (Figure 4.6). Among these, the two METIs pyridaben and tebufenpyrad had a different response. The toxicity of pyridaben was greatly diminished, while that of tebufenpyrad was not affected.
4.4. DISCUSSION

4.4.1. TRANSCRIPTIONAL RESPONSES TO HOST PLANT SHIFT

Classical studies of mite host plant transfer have shown that fitness on new hosts can increase rapidly over a limited number of generations (Fry, 1989; Yano et al., 2001). Here, we showed that the response of mites to a new host was accompanied by similarly rapid changes in transcriptional profiles on a timescale of hours to a relatively small number of generations. Genes that responded in the short-term (12 h or less) included those in classical detoxification families, such as CYPs and CCEs. This finding is consistent with that observed for *T. urticae* unfed larvae that were placed for 12 h on either bean, tomato or *Arabidopsis thaliana* (ascertained with another method (RNA-seq) (Grbic et al., 2011)). However, compared to the short-term response, expression changes were far more dramatic at five generations. About 3-fold more genes were detected as differentially expressed, with ~7.5% of all *T. urticae* genes changed in expression. The host adaptation experiments used the genetically diverse London reference strain that segregates for several haplotypes throughout much of the genome (Van Leeuwen et al., 2012). Whether the pronounced differences in gene expression observed at five generations reflected selection on standing genetic variation, or alternatively physiological or epigenetic changes, remains to
be determined (see chapter 5 of this thesis). However, genetic variation between spider mites for characters affecting host use has been reported previously (Fry, 1989; Gotoh et al., 1993; Kant et al., 2008), and selection is observed in spider mite populations after a few generations (Fry, 1989; Yano et al., 2001).

4.4.2. HOST PLANT SHIFT AND PESTICIDE RESISTANCE

Remarkably, the transcriptome of pesticide susceptible mites grown for five generations on tomato was closer to that of two acaricide multi-resistant strains than to that of the initial response to the host plant shift. Of note, coordinated transcriptional changes were apparent for known major environmental response genes including CYPs, GSTs, and CCEs (Berenbaum, 2002). This suggests that in response to different chemical challenges, the spider mite is ‘rounding up the usual suspects’. The mechanism involved is still uncertain, but the pattern of expression for several transcription factor genes and their regulators was also similar between Tomato-5G and the resistant strains, possibly underlying the coordinated response.

In mammals, the pregnane X receptor (PXR) and constitutive androstane receptor (CAR) are major xenosensor genes (Pascussi et al., 2008). The Drosophila parologue of these nuclear receptor genes, DHR96, is also involved in the response to xenobiotics (King-Jones et al., 2006). One of the eight *T. urticae* paralogues of *DHR96*, *tetur36g00260*, was more than 2-fold up-regulated in Tomato-5G, MAR-AB, and MR-VP (Table S4.5). Further, RNAi silencing of the Drosophila *Keap1* (Kelch-like ECH associated protein) gene, a negative regulator of the Drosophila transcription factor CncC (cap ‘n’ collar isoform-C), up-regulated detoxification gene expression (Misra et al., 2011). In our experiments, the down-regulation of OrthoMCL cluster 10254 transcripts (Kelch-like proteins) might have similarly resulted in the up-regulation of *T. urticae* detoxification genes. However, there are multiple Kelch-like proteins and two CncC orthologues in *T. urticae*, and experimental studies will be required to confirm the relative roles of upstream transcriptional regulators in establishing the observed transcriptional profiles of tomato-adapted and acaricide multi-resistant mites.

4.4.3. EXTENDING THE ARSENAL OF ENVIRONMENTAL RESPONSE GENES

Despite the involvement of known detoxification gene families, a major finding was that, after host plant shift or between acaricide susceptible and resistant strains, many genes that encode proteins without homology to proteins of
known functions changed expression. In part, this may reflect the phylogenetic distance between spider mites and insects for which most arthropod functional-genetic studies have been performed (the divergence between mites and insects is more than 450 MYA; Dunlop, 2010). However, it may also reflect the recruitment of diverse genes to contribute to the polyphagous life style of _T. urticae_.

This interpretation was supported by our finding of strong transcriptional responses to changes in chemical exposure for some gene families either absent from insects or previously not recognized to play a major role in xenobiotic response. For instance, intradiol ring-cleaving dioxygenases (ID-RCDs), which have not been reported in other metazoa, may contribute prominently to the spider mite detoxification arsenal. The expansion of this family in the _T. urticae_ genome (Grbic et al., 2011), and the transcriptional regulation of many family members, in response to host plant transfer or in pesticide resistant strains, is indicative of a selective advantage after the initial horizontal gene transfer. Currently, the substrate specificity of the spider mite ID-RCDs is not known. Spider mite ID-RCDs were most closely related to a clade of secreted fungal ID-RCD enzymes within which an _Aspergillus fumigatus_ homologue modifies an array of procyandins, which are polymers of (+)-catechin and/or (-)-epicatechin (Roopesh et al., 2010). This hints that this group of ID-RCDs can metabolize more complex structures than simple catecholic substances. The _A. fumigatus_ enzyme showed common features with the spider mite enzymes: it was predicted to be secreted, had a similar distance between active site residues, and contrasted with previously biochemically characterized enzymes in bacteria and other Fungi (Figures 4.4 and 4.5). One of the _T. urticae_ enzymes (tetur20g01790), had identical residues at the predicted binding sites for epicatechin in the _A. fumigatus_ protein sequence (Tyr184, Thr229, Arg231 and His236, _A. fumigatus_ numbering) (Roopesh et al., 2012) (Figure 4.4). Interestingly, a secreted ID-RCD from the oomycete _Phytophthora infestans_ (GenBank accession number XP_002905783) clustered within the group of fungal secreted ID-RCDs (Figure 4.5). This is in agreement with Richards et al. (2006, 2011b) who showed that the closely related species _Phytophthora ramorum_ acquired its extracellular ID-RCDs from filamentous ascomycetes through horizontal gene transfer.

Other gene families are ubiquitous in their phylogenetic distribution, but their role in environmental response is striking in the spider mite. Lipocalins are small extracellular proteins and are characterized by: 1) their ability to bind to hydrophobic molecules; 2) a conserved secondary structure (an antiparallel β-barrel, with a repeated +1 topology, with an internal ligand-binding site); 3) low
sequence conservation (< 20%) and 4) a conserved exon-intron structure (Flower et al., 2000; Sanchez et al., 2006). Members of the lipocalin family are found in a wide range of species, with roles in metabolism, coloration, perception, reproduction, developmental processes and modulation of immune and inflammatory responses (Chudzinski-Tavassi et al., 2010), resulting in a very diverse nomenclature for each specific lipocalin (e.g. apolipoprotein D, crustacyanins, biliproteins and salivacalins). As lipocalins typically bind hydrophobic molecules, they may bind pesticides/allelochemicals in mites, resulting in sequestration of these toxic, generally hydrophobic compounds (see also Figure 1.2). Moreover, the feeding strategy of mites may favor sequestration in dispensable phagocytes, as suggested by Mullin and Croft (1983). Lipocalins may also protect against oxidative stress, as loss of the Drosophila homologue of ApoD, Glial Lazarillo, increases sensitivity to oxidative stress, while overexpression increases hyperoxia tolerance (Walker et al., 2006). Similar findings were also reported for ApoD in mice and in Arabidopsis (Charron et al., 2008; Ganfornina et al., 2008). In plants, the oxidative burst is one of the earliest observable aspects of the plant defense strategy against herbivores (Bi & Felton, 1995; Hildebrand et al., 1986). Whether mite lipocalins, which were up-regulated specifically in the early response to tomato exposure (2 and 12 h), are involved in resistance to the oxidative response warrants investigation.

In addition to enzymes and small secreted proteins, membrane binding proteins and transporters featured prominently in our analysis, including the MFS family, one of the largest families of membrane transporters along with ABC transporters. Our experiments did not reveal an increased expression of many ABC transporter genes, although this gene family is highly expanded in the spider mite (Dermauw et al., 2013; Grbic et al., 2011). Multiple MFS family members were differentially expressed in both Tomato-5G and resistant strains. MFS transporters are single-polypeptide carriers that work in symport/antiport (Pao et al., 1998; Reddy et al., 2012), and studies in bacteria and fungi have identified roles in the transport of toxic substances (Hillen & Berens, 1994; Kretschmer et al., 2009; Saidijam et al., 2006). For example, overexpression of the mfsM2 gene in a sensitive strain of Botrytis cinerea, a fungal plant pathogen, led to drug resistance levels similar to those of a fungicide resistant B. cinerea strain (Kretschmer et al., 2009). If spider mite MFS proteins function as efflux transporters, their up-regulation might result in a higher efflux of xenobiotics (both natural and synthetic) or their metabolites out of spider mite cells (see also Figure 1.2).
Spider Mite Responses to Chemical Pressures

A third group of non-catalytic binding proteins/transporters proteins with a high percentage of genes changing expression in Tomato-5G, MR-VP, and MAR-AB were Low-Density Lipoprotein receptor proteins (LDLRs) (OrthoMCL cluster 10364, Table 4.2). In humans, these endocytic receptors bind hydrophobic molecules (Krieger et al., 1979) and participate in a wide range of physiological processes. Some members of this family have also multi-ligand binding properties (Herz & Strickland, 2001; May et al., 2007). In our experiments, all these genes were down-regulated. If hydrophobic pesticides and allelochemicals interact with these LDLRs, a down-regulation could result, through lower receptor mediated endocytosis, in a lower uptake into spider mite cells.

The role of binding proteins/transporters in the response of insects to chemically challenging environments has been generally overlooked, and would merit closer attention (Sorensen & Dearing, 2006). Earlier work using dedicated microarrays (Daborn et al., 2002; David et al., 2005) could not, by definition, uncover the importance of these new players. However, microarray studies with restricted random sets of ESTs already pointed out that transporters were differentially expressed in lepidopteran larvae that were fed with plants that differed in their defensive phytochemical profile (Govind et al., 2010). Our study with a pangenomic array extends this early observation, and emphasizes the importance of (largely) unbiased approaches for studies to understand the basis of generalist herbivore life histories. Obtaining formal evidence that some members of these gene families actually contribute to xenobiotic tolerance when up- or down-regulated should be a priority for future research.

4.4.4. HOST PLANT SHIFT AND ACARICIDE TOLERANCE

We show here that adaptation to tomato changed not just the transcript levels of many detoxification enzymes, but resulted in a decreased acute toxicity for three out of five acaricides tested. The transcriptional remodeling we observed after tomato adaptation may thus be the proximal cause for the well-known effect of host plants on the efficacy of acaricides in *T. urticae* (Gould et al., 1982; Ibrahim, 2009; Neiswander et al., 1950; Yang et al., 2001). Host plants are also known to affect the toxicity of insecticides to insects (Ahmad, 1986; Berry et al., 1980; Brattsten et al., 1977; Castle et al., 2009; Kennedy, 1984; Li et al., 2000, 2004; Lindroth, 1989; Sasabe et al., 2004; Sen Zeng et al., 2007; Yu et al., 1979). Feeding on alternative hosts changes the activity of mite detoxification enzymes measured with some model substrates (Mullin & Croft, 1983; Yang et al., 2001), and performance of mites on tomato was negatively affected by a CYP inhibitor (Agrawal
et al., 2002). These findings indicate that the transcriptomic changes we observed may be essential for performance. The importance of active (expressed) herbivore detoxification enzymes to survival on chemically challenging hosts is also well documented in insects (e.g. Brattsten et al., 1977; Snyder & Glendinning, 1996).

4.4.5. POLYPHAGY AND PESTICIDE RESISTANCE

We found a common pattern of gene expression between mites that adapted to a new host and those constitutively resistant to diverse pesticides. Moreover, the unexpected gene families that figured so prominently in our analysis, as well as the usual suspects, formed a coherent whole from transcription factors to effector genes in xenobiotic detoxification, binding and transport. This points towards an orchestrated response rather than a random deregulation caused by the toxic effects of the plant or acaricide challenge. Many gene families we found to have strong transcriptional responses were large, reflecting a (presumed) long evolutionary history of gene duplications, sometimes in a lineage specific manner (Feyereisen, 2011; Grbic et al., 2011). In the response to plants, such patterns are expected given the long evolutionary timescale over which plant-arthropod interactions are fine tuned. In contrast, the number of detoxification genes with altered transcription in the pesticide resistant strains, and with fold-changes mirroring that of host plant transfer, is not expected from classical theory. It has long been argued that field-evolved resistance would select single genes with major effects rather than many with limited effect (McKenzie & Batterham, 1994; Roush & McKenzie, 1987). In some cases of target site and metabolic resistance, the monogenic inheritance has been established experimentally (ffrench-Constant et al., 2004; Van Leeuwen et al., 2012). Additionally, there is a substantial difference between the moderate tolerance level of the tomato-adapted mites and the high resistance level of the acaricide-resistant mites. If all common genes in the transcriptomic signature contributed directly to resistance, then Tomato-5G should be much more resistant. This difference between the high resistance level of the MR-VP and MAR-AB strains and the more moderate tolerance level of the Tomato-5G mites may be due to the presence of target site mutations in the 2 resistant strains.

We propose an explanation to these apparent paradoxes that may be relevant to the rapid development of resistance in polyphagous pests (FIGURE 4.7). While a single resistance gene with major effect may eventually be selected from rare alleles, initial survival in an environment with heterogeneous dose or distribution of the pesticide may depend on multiple alleles that confer moderate
resistance. Such multiple alleles may be present in modules and include genes controlling detoxification, binding and translocation processes thus affecting all aspects of the toxicokinetic balance. Intraspecific genetic variation in host preference is a common aspect of polyphagy (Futuyma & Peterson, 1985; Via,

**Figure 4.7.** Model for rapid evolution of pesticide resistance in phytophagous generalists versus specialists. Selection on multiple host plants in generalists is postulated to increase polymorphism in environmental responses leading to several specific subsets of alleles (different subsets of alleles are indicated by different colors in the generalist panel). The initial stages of selection by a pesticide mimics a host plant shift, resulting in rapid selection of the best adapted subset of environmental response alleles. For generalists, these preadaptations provide a larger initial population from which a rare (high) resistance allele can be selected, thus accelerating the development of agriculturally significant resistance levels when compared to specialists. After both types of selection (host shift, pesticide), the transcriptome signature is similar, because it is drawn from a similar subset of genotypes. Figure is adapted and modified from McKenzie & Batterham (1994).
1990), and has been shown in the spider mite (Fry, 1989; Gould, 1979). To the mite, encountering a plant treated with an acaricide may be akin to encountering a new host plant. There would be rapid selection for a genotype carrying a set of genes whose expression would best buffer against the chemical signature of the new hostile environment. That genetic variation in environmental response can come in groups of connected genes has been recently documented. Phenotypic variation in the transcriptome profile of 40 inbred *Drosophila* strains was shown to consist of groups of interconnected genes. This formed hundreds of ‘modules’ of ecologically relevant correlated genetic variation (Ayroles et al., 2009). In a polyphagous pest such as the spider mite, such ‘modules’ of co-regulated genes may provide an explanation for the common transcriptional patterns of MR-VP, MAR-AB and Tomato-5G. In this sense, when polyphagy is seen as genetic polymorphism in the response to different chemical environments, it may represent a pre-adaptation to xenobiotic resistance as suggested previously (Gordon, 1961; Rosenheim et al., 1996). The selection of the rare resistance allele to the acaricide would be facilitated by the initial, higher survival rate of a subset of the population harboring it.

A plant-specific transcriptional response has been observed in polyphagous Lepidoptera and in the spider mite (Celorio-Mancera et al., 2012; Celorio-Mancera et al., 2013; Govind et al., 2010; Grbic et al., 2011). Specialist herbivores, on the other hand, are characterized by a xenobiotic response that is more constitutive and more targeted towards the favored host plants (Berenbaum, 2002), and the transcriptional response to change in phytochemistry is much more restricted in a specialist than in a generalist (Govind et al., 2010). It remains to be shown that intraspecific genetic variation in the spider mite includes the differential regulation of specific subsets of genes involved in fitness on different host plants, such as detoxification, binding and transport as we predict here. Such experimental verification is as important as it is difficult to obtain. However, the rapid development of resistance in polyphagous herbivores and its relative absence in specialists such as natural enemies and predators is well known (Croft & Strickler, 1983). Our results provide an unprecedented insight into the transcriptional correlation that may link polyphagy with development of pesticide resistance. They also highlight the need to study enzymes which not just metabolize, but also bind and translocate xenobiotics as major contributors to survival in a toxic environment.
4.5. REFERENCES


Spider Mite Responses to Chemical Pressures

genase from Nocardioides simplex 3E, a key enzyme involved in polychlorinated aromatics biodegradation. *Journal of Biological Chemistry, 280*, 21144-21154.


Spider Mite Responses to Chemical Pressures


Spider Mite Responses to Chemical Pressures


**Supplementary Material**

Supplementary data can be located at www.pnas.org/cgi/doi/10.1073/pnas.1213214110

**Figure S4.1.** Microarray validation by qPCR. Validation was performed for 10 differentially expressed genes (CYP392A16 (tetur06g04520), CYP392D2 (tetur03g04990), CYP392D8 (tetur03g03070), CYP392D10 (tetur03g05110), tetur02g09840 (UDP-glycosyltransferase), tetur16g03200 (MFS transporter), tetur13g04550, tetur01g00490 (intradiol ring-cleaving dioxygenase, ID-RCD), TuGSTd14 (tetur29g00220) and tetur06g04970 (short chain reductase) for two susceptible strains (London and LS-VL) and both resistant strains (MR-VP and MAR-AB). Error bars represent the standard error of the calculated mean based on three biological replicates. Microarray expression data (MR-VP and MAR-AB, microarray) from this gene selection is shown adjacent to their qPCR expression data.

**Figure S4.2.** A: Clustering of lipocalin gene expression between conditions (MAR-AB, MR-VP and host plant change; Tomato-2h, Tomato-12h and Tomato-5G), relative to London on bean, revealed correlation of expression levels. Clustering was performed in GeneSpring GX11.0, using the hierarchical clustering algorithm with Pearson centered distance metric and complete linkage rule. The color bar with corresponding log2FC values is show at the bottom of the figure. The bar under the asterisk indicates the number and phases of introns of each Tetranychus urticae lipocalin (as explained in panel C; black: type a, grey: type b, white: neither type a or b). B: Maximum likelihood unrooted tree depicting the phylogenetic relationship of the expanded (58 genes) lipocalin family of T. urticae. Most of T. urticae lipocalins clustered with mammalian Apolipoprotein D and crustacyanins. T. urticae homologues of insect biliproteins and Karl of D. melanogaster are indicated as Tu b1 (tetur07g03790), Tu b2 (tetur17g00050) and TuK1 (tetur02g11960), TuK2 (tetur01g01510), TuK3 (tetur01g1500), TuK4 (tetur01g16584) and TuK5 (tetur01g01520), respectively. Members within two subclades of the T. urticae lipocalins had similar numbers and phases of introns (type a or b, see panel C) and are depicted by circles. The highly divergent tick lipocalin protein sequences were not included in our phylogenetic analysis in order to decrease the risk of long-branch artefacts (Ganfornina et al., 2006). C: Comparison of lipocalin gene structure consensus between arthropods (Sanchez et al., 2006) and T. urticae revealed a new T. urticae-specific gene structure (type b). Square boxes (not drawn to scale) represent exons. Numbers above boxes represent the exon size range (bp) while numbers between boxes represent intron phases.
Spider Mite Responses to Chemical Pressures

**TABLE S4.1.** GenBank accession numbers of ID-RCDs used for phylogenetic analysis.

**TABLE S4.2.** Accession numbers of lipocalins used for phylogenetic analysis.

**TABLE S4.3.** *T. urticae* lipocalin properties.

**TABLE S4.4.** Classification of *T. urticae* genes coding for MFS transporters (from OrthoMCL 10032, 10082 and 10236) determined by BLASTp in the Transporter Classification DataBase (Saier et al., 2009).

**TABLE S4.5.** Differentially expressed genes coding for transcription factors in MR-VP, MAR-AB and Tomato-5G (Tom-5G).