Supporting Information

Dermauw et al. 10.1073/pnas.1213214110

SI Materials and Methods

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 nt per gene with a T_m of 80 °C and parameters set to "best probe design" and "3′ bias". 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 to be shorter, 100 bp downstream the stop codon were added to the coding sequence. In total 58,985 probes were designed. Before the start of GeneSpring analysis (see below), probes were remapped [using Bowtie, version 0.12.7 (1)] on the most recent genome annotation (April 18, 2011; 18,455 predicted genes). According to this mapping, 87.4% from the latter were covered by at least one probe, whereas 81.7% were covered by at least three probes. Standard Agilent features such as spike-ins were added (IS-62976-8-V2_60kby8_GX_EQC_20100210). We selected 182 unique probes that mapped to *T. urticae* genes expressed across four developmental stages, as identified by RNAseq experiments (2), with different ranges of expression [based on normalized read counts (rpkm)]. These probes (probe names starting with "Rep") were randomly distributed in 10–15 copies per array and were used to measure array reproducibility. The array design was submitted to the National Center for Biotechnology Information (NCBI) under the Gene Expression Omnibus (GEO)-platform format (GPL15756).

Target Preparation, Microarray Hybridization, and Analysis. Total RNA was extracted from one hundred 1- to 3-d-old adult female mites, using the RNeasy mini kit (Qiagen), in four to six replicates for each strain (MR-VP, MAR-AB, and London) or for each time point (Tomato-2h, Tomato-12h, and Tomato-5G) in the host change experiment. Contaminating DNA was removed by digestion with RNase-free Turbo DNase (Ambion). The quality and quantity of the RNA were assessed by a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and by running an aliquot on a 1% agarose gel. One hundred nanograms 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 labeled cRNA was purified with the RNeasy mini kit (Qiagen), and dye content (>8.0 pmol dye per microgram cRNA) and the concentration of cRNA were measured by a 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. The following hybridization experiments were performed (the number of biological replicates is given in brackets): Cy5-labeled MR-VP cRNA/Cy3-labeled London cRNA [6], Cy5-labeled MAR-AB cRNA/Cy3-labeled London cRNA [5], Cy5-labeled Tomato-2h cRNA/Cy3-labeled London cRNA [4], Cy5-labeled Tomato-12h cRNA/Cy3-labeled London cRNA [4], and Cy5-labeled Tomato-5G cRNA/Cy3-labeled 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), and protected by an ozone-barrier cover (Agilent Technologies) until scanned by an Agilent Microarray High-Resolution Scanner with default settings for 8 × 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 were 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 that showed a more than twofold change were selected for a *t*-test *P* value (false discovery rate <5% as assessed with the Benjamini–Hochberg correction for multiple testing). All our microarray datasets are accessible through the GEO series accession no. GSE39869.

Microarray Validation by qPCR. Microarray validation by quantitative real-time PCR (qPCR) was performed for 10 differentially expressed genes [*CYP392A16* (*tetur06g04520*), *CYP392D2* (*tetur03g04990*), *CYP392D8* (*tetur03g05070*), *CYP392D10* (*tetur03g05110*), *tetur02g09840* (glycosyltransferase), *tetur16g03200* [major facilitator superfamily (MFS) transporter], *tetur13g04550*, *tetur01g00490* [intradiol ring-cleavage dioxygenase (ID-RCD)], *TuGSTd14* (*tetur29g00220*), and *tetur06g04970* (short chain reductase)] of the susceptible London strain and both resistant strains (MR-VP and MAR-AB). 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 were performed as previously described for microarray experiments. First-strand cDNA was synthesized from 2 μg of total RNA, using the Maxima first-strand cDNA synthesis kit (Fermentas). Real-time PCR was done on an Mx3000p real-time PCR system (Stratagene), using Maxima SYBR green qPCR Master Mix (Fermentas) with two to three biological replicates and three technical replicates for each gene. Gene-specific primers for the 10 differentially expressed genes and 2 housekeeping genes (actin, ribosomal protein 49) were designed using Primer 3 software (3) (Table S6, primer names with "q" suffix). Relative levels of expression were calculated according to Pfaffl et al. (4).

OrthoMCL Clustering. For gene clustering, the OrthoMCL (5) software version 2.0 with parameters (-v all -te 2 -scheme 7 -I 1.7) was applied with the species combination *Drosophila melanogaster*, *Tribolium castaneum*, *Acyrthosiphon pisum*, *Tetranychus urticae*, *Caenorhabditis elegans*, and *Homo sapiens.* In this analysis, a total of 103,935 sequences clustered into 13,876 gene families (76,810 genes in clusters; 27,125 singletons). Of these, 3,069 clusters contained sequences from all six genomes. Of the protein-coding genes predicted for *T. urticae*, 11,831 were clustered in a total of 6,098 groups. OrthoMCL uses a Markov clustering algorithm on a precalculated sequence similarity matrix to group (putative) orthologs and paralogs (6). The matrix was built on the basis of an all-against-all BLASTp (7) (2.2.24+; default parameters) and filtered according to the OrthoMCL manual. Where predicted, splice variants were removed from the dataset (the longest protein sequence prediction was withheld). OrthoMCL clustering results of *T. urticae* proteins can be accessed at the BOGAS website (http://bioinformatics.psb.ugent.be/webtools/bogas/).

Signal Peptide Prediction. The presence of a signal peptide in all protein sequences investigated in this study was predicted with SignalP 3.0 (8), using hidden Markov models (HMMs) and neural networks (NNs). Protein sequences were considered to have a signal peptide under the condition that both models predicted a signal peptide.

Gene Family Analysis. ***ID-RCDs.*** Gene-specific *T. urticae* ID-RCD primers (Table S6, primers with "s" suffix) were designed [using Primer 3 software (3)] to amplify a 600- to 850-bp fragment of ID-RCDs of the closely related spider mite *T. evansi*. *T. evansi* genomic DNA was extracted using a phenol chloroform extraction method as described by Khajehali et al. (9). PCRs were performed in 50-μL reaction volumes with 38.7 μL distilled water; 5 μL 10× PCR buffer, 2 mM $MgCl_2$, 0.2 mM dNTP, and 0.2 μM of each primer; and 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 at 94 °C, 35 × (20 s at 94 °C, 55 s at 49 °C, 2 min at 72 °C), and 5 min at 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 *Escherichia 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 nos. JQ736355–JQ736359).

The presence of ID-RCDs in *Panonychus citri*, *Metaseiulus occidentalis*, *Varroa destructor*, and *Ixodes scapularis* was determined through a tBLASTn search (7) of *T. urticae* ID-RCD protein sequences against the published transcriptome dataset of *P. citri* [European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL-EBI) accession no. ERP000885], the genome of *M. occidentalis* (Mocc 1.0 assembly), the genome of *V. destructor* (BRL_Vdes_1.0 assembly), and the genome of *I. scapularis* (ASM20861v1 assembly).

All protein sequences of *T. urticae* ID-RCDs (2) were used as queries in BLASTp (cutoff *E*-value ≤ e^{-10}) searches of the NCBI nonredundant protein database. This resulted in 280 unique hits with a protein length between 176 and 500 aa. 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 the Conserved Domain Database (10) 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 dioxygenases (cd03459–cd03464), belonging to bacteria and fungi, were added to the dataset. Finally, 17 *T. urticae* ID-RCDs and five *T. evansi* homologs (see above) were incorporated into the analysis (see Table S1 for GenBank accession numbers). The amino acid sequence alignment was constructed using MUSCLE (11). Model selection was done with ProtTest 2.4 (12) and according to the Akaike information criterion the model WAG+I+G+F was optimum for phylogenetic analysis. Finally, a maximum-likelihood analysis was performed using Treefinder (13), bootstrapping with 1,000 pseudoreplicates (LR-ELW). Phylogenetic trees were visualized and edited using MEGA5 (14) and CorelDraw X3 (Corel), respectively.

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 queries in BLASTp searches (7) of the NCBI nonredundant 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 *Homo sapiens* (GenBank accession no. P05090) was used as a query in BLASTp (cutoff *E*-value ≤ e^{-5}) against the proteome of *T. urticae* at the BOGAS website (http://bioinformatics.psb.ugent.be/webtools/bogas/). Finally, the resulting hits were used as a query in BLASTp (cut off *E*-value ≤ e^{-5}) against the proteome of *T. urticae*. 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; Table S2). The latter were together with a selected reference dataset of lipocalin genes (15) (see Table S3 for GenBank accession numbers) aligned using the profile alignment mode of Clustal X and the alignment of Sanchez et al. (16) as a profile (17). Tick lipocalin protein sequences were not included in our phylogenetic analysis to decrease the risk of long-branching artifacts (15). Model selection was performed with ProtTest 2.4 (12) and according to the Akaike information criterion the model WAG+G+F was optimum for phylogenetic analysis. Finally, a maximum-likelihood analysis was performed using Treefinder (13), bootstrapping with 500 pseudoreplicates (LR-ELW). Phylogenetic trees were visualized and edited using MEGA5 (14) and CorelDraw X3 (Corel), respectively. Secondary structures of *T. urticae* lipocalins were predicted using Jpred 3 (18) whereas GPI-anchor sites were predicted using PredGPI (19).

MFS transporters. Pfam domain searches of our microarray 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 a query in BLASTp in the Transporter Classification DataBase (20) (Table S4). Transmembrane regions were predicted using TMHMM Server v. 2.0. (http://www.cbs.dtu.dk/services/TMHMM/).

1. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
2. Grbić M, et al. (2011) The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479(7374):487–492.
3. Rozen S, Skaletsky HJ (2000) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, eds Krawetz S, Misener S (Humana Press, Totowa, NJ), pp 365–386.
4. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29(9):e45.
5. Van Dongen S (2000) Graph clustering by flow simulation. PhD thesis (Univ of Utrecht, Utrecht, The Netherlands).
6. Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13(9):2178–2189.
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
8. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340(4):783–795.
9. Khajehali J, Van Nieuwenhuyse P, Demaeght P, Tirry L, Van Leeuwen T (2011) Acaricide resistance and resistance mechanisms in *Tetranychus urticae* populations from rose greenhouses in the Netherlands. *Pest Manag Sci* 67(11):1424–1433.
10. Marchler-Bauer A, et al. (2011) CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39(Database issue):D225–D229.
11. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
12. Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104–2105.
13. Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18.
14. Tamura K, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739.
15. Ganfornina MD, Kayser H, Sanchez D (2006) *Lipocalins*, eds Åkerström B, Borregaard N, Flower DR, Salier JP (Landes Bioscience, Austin, TX), pp 49–74.

16. Sánchez D, Ganfornina MD, Gutiérrez G, Marín A (2003) Exon-intron structure and evolution of the Lipocalin gene family. *Mol Biol Evol* 20(5):775–783.
17. Anderson CL, Strope CL, Moriyama EN (2011) SuiteMSA: Visual tools for multiple sequence alignment comparison and molecular sequence simulation. *BMC Bioinformatics* 12:184.
18. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36(Web Server issue):W197–W201.
19. Pierleoni A, Martelli PL, Casadio R (2008) PredGPI: A GPI-anchor predictor. *BMC Bioinformatics* 9:392.
20. Saier MH, Jr., Yen MR, Noto K, Tamang DG, Elkan C (2009) The Transporter Classification Database: Recent advances. *Nucleic Acids Res* 37(Database issue): D274–D278.

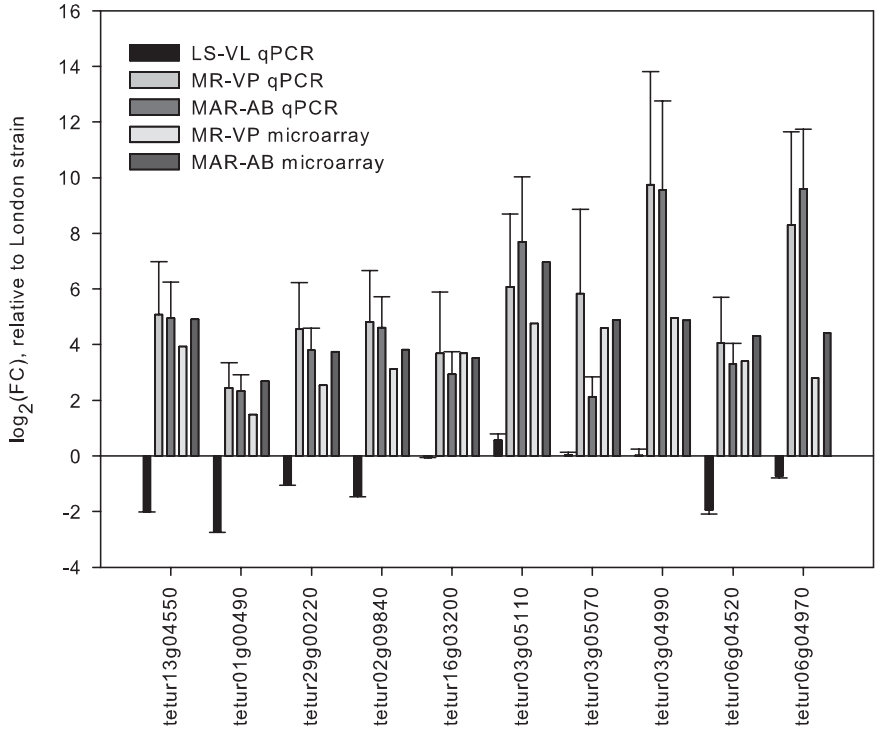

Fig. S1. Microarray validation by quantitative real-time PCR. Validation was performed for 10 differentially expressed genes [*CYP392A16* (*tetur06g04520*), *CYP392D2* (*tetur03g04990*), *CYP392D8* (*tetur03g05070*), *CYP392D10* (*tetur03g05110*), *tetur02g09840* (glycosyltransferase), *tetur16g03200* (MFS transporter), *tetur13g04550*, *tetur01g00490* (ID-RCDs), *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 SE of the calculated mean based on three biological replicates. Microarray expression data (MR-VP and MAR-AB, microarray) from this selection of genes are shown next to their qPCR expression data.

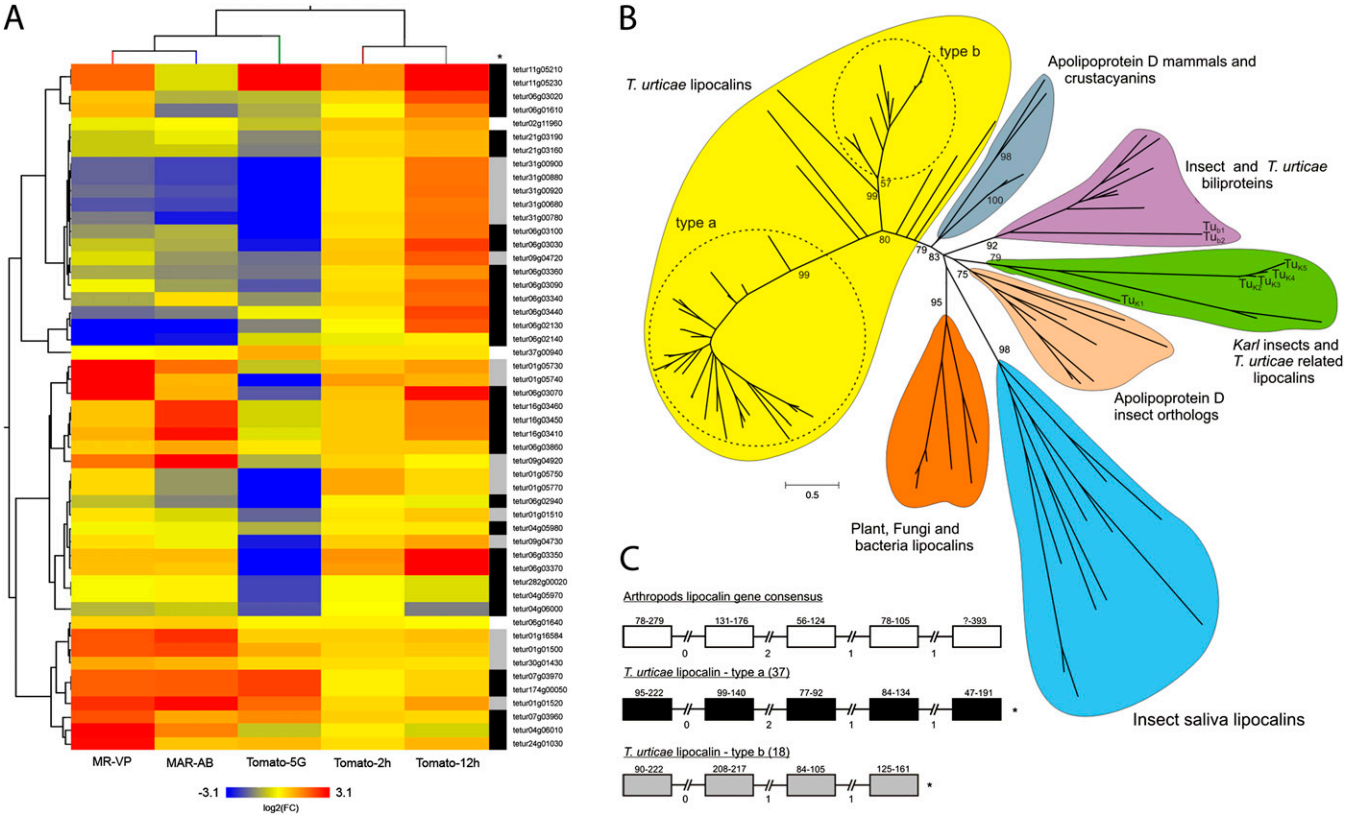

Fig. S2. Lipocalins in *T. urticae*. (*A*) Clustering of lipocalin gene expression between conditions [MAR-AB, MR-VP, and host change (Tomato-2h, Tomato-12h, and Tomato-5G)] reveals correlation of expression levels. Clustering was performed in GeneSpring GX11.0, using the hierarchical clustering algorithm with Pearson's centered distance metric and complete linkage rule. The color bar with corresponding $\log_2$(FC) values is shown at the bottom. The bar under the asterisk indicates the number and phases of introns of each *T. urticae* lipocalin (as explained in *C*: black, type a; gray, type b; white, neither type a nor type b). (*B*) Maximun-likelihood unrooted tree depicting the phylogenetic relationship of the expanded (58 genes) lipocalin family of *T. urticae*. Most of *T. urticae* lipocalins cluster with mammalian Apolipoprotein D and crustacyanins. *T. urticae* homologs of insect biliproteins and Karl of *D. melanogaster* are indicated as Tu_{b1} (tetur07g03790), Tu_{b2} (tetur174g00050) and Tu_{K1} (tetur02g11960), Tu_{K2} (tetur01g01510), Tu_{K3} (tetur01g01500), Tu_{K4} (tetur01g16584), and Tu_{K5} (tetur01g01520), respectively. Members within two subclades of the *T. urticae* lipocalins have similar numbers and phases of introns (type a or b, see *C*) and are depicted by circles. The highly divergent tick lipocalin protein sequences were not included in our phylogenetic analysis to decrease the risk of long-branch artifacts (1). (*C*) Comparison of lipocalin gene structure consensus between arthropods (2) and *T. urticae*, revealing a new *T. urticae* 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.

1. Ganfornina MD, Kayser H, Sanchez D (2006) *Lipocalins*, eds Åkerström B, Borregaard N, Flower DR, Salier JP (Landes Bioscience, Austin, TX), pp 49–74.
2. Sanchez D, et al. (2006) *Lipocalins*, eds Åkerström B, Borregaard N, Flower DR, Salier JP (Landes Bioscience, Austin, TX), pp 5–16.

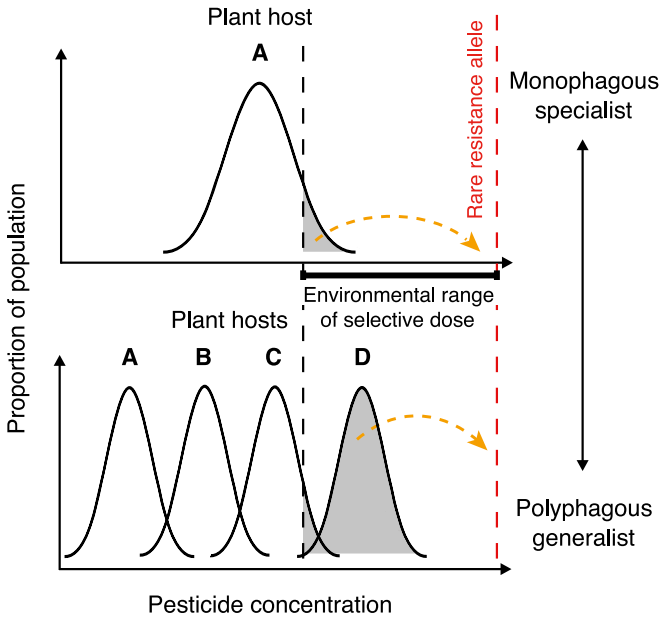

Fig. S3. Model for rapid evolution of pesticide resistance in generalist herbivores compared with specialists (adapted and modified from ref. 1). Preadaptation to multiple host plants is postulated to increase polymorphism in environmental responses leading to several subsets of alleles. The initial stages of selection by a pesticide mimic a host plant shift, resulting in rapid selection of the best-adapted subset of environmental response alleles. This in turn provides a larger initial population from which a rare (high) resistance allele can be selected, thus accelerating the development of agriculturally significant resistance levels compared with specialists. After both types of selection (host shift, pesticide), the transcriptome signature is similar, because it is drawn from a similar subset of genotypes.

1. McKenzie JA, Batterham P (1994) The genetic, molecular and phenotypic consequences of selection for insecticide resistance. *Trends Ecol Evol* 9(5):166–169.

Table S1. GenBank accession numbers of ID-RCDs used for phylogenetic analysis

Genus*	Accession no.†	S_p‡	Genus*	Accession no.†	S_p‡	Genus*	Accession no.†	S_p‡
Tetranychus sp. ID-RCDs								
Tetranychus urticae	tetur01g00490	Y	*T. urticae*	tetur07g05940	Y	*T. urticae*	tetur20g01790	Y
T. urticae	tetur04g00150	Y	*T. urticae*	tetur07g06560	Y	*T. urticae*	tetur28g01250	Y
T. urticae	tetur04g08620	Y	*T. urticae*	tetur12g04761	Y	*T. urticae*	tetur44g00140	Y
T. urticae	tetur06g00450	Y	*T. urticae*	tetur13g04550	Y	*Tetranychus evansi*	JQ736355	
T. urticae	tetur06g00460	Y	*T. urticae*	tetur19g02300	Y	*T. evansi*	JQ736356	
T. urticae	tetur07g02040	Y	*T. urticae*	tetur19g03360	Y	*T. evansi*	JQ736357	
T. urticae	tetur07g05930	Y	*T. urticae*	tetur20g01160	Y	*T. evansi*	JQ736358	
						T. evansi	JQ736359	
"Classical" ID-RCDs								
Acinetobacter	Q1WCN0	N	*Chelativorans*	Q11GN2	N	*Pyrenophora*	B2WPE2	N
Agrobacterium	Q7CV76	N	*Escherichia*	B7LQX6	N	*Ralstonia*	Q0KAA6	N
Ajellomyces	C5JTU1	N	*Jannaschia*	Q28JN2	N	*Rhodococcus*	Q6F4M7	N
Arthrobacter	A1RBV0	N	*Pandoraea*	A7KX07	N	*Streptomyces*	Q9APK9	N
Arthrobacter	Q76CC4	N	*Paracoccidioides*	C1GHP3	N	*Talaromyces*	B8M1V8	N
Aspergillus	B8NN75	N	*Pseudomonas*	A6V876	N			
Candidatus	Q02CF8	N	*Pyrenophora*	B2VWY2	N			
NCBI BLASTp hits (E-value $\leq e^{-10}$) using *T. urticae* ID-RCDs as query								
Acaryochloris	YP_001515498	N	*Frankia*	YP_479231	N	*Phytophthora*	XP_002905783	Y
Acidobacterium	ZP_07065922	Y	*Frankia*	YP_001508082	N	*Podospora*	CAP65306	Y
Acidobacterium	ZP_07648583	N	*Frankia*	YP_001508663	N	*Podospora*	XP_001903249	Y
Actinosynnema	YP_003100310	N	*Frankia*	YP_004019605	N	*Pseudomonas*	YP_350088	N
Actinosynnema	YP_003100966	N	*Frankia*	ZP_06410955	N	*Puccinia*	EFP84436	Y
Ajellomyces	EEH02815	Y	*Frankia*	ZP_06475430	N	*Pyrenophora*	EFQ91909	Y
Ajellomyces	EER41332	Y	*Fulvimarina*	ZP_01438862	Y	*Pyrenophora*	XP_001940623	Y
Ajellomyces	EEQ89061	Y	*Gemmatimonas*	YP_002762722	N	*Ralstonia*	YP_728952	N
Alteromonas	ZP_04713930	N	*Gibberella*	XP_382158	Y	*Ralstonia*	YP_299431	N
Amycolatopsis	YP_003765631	Y	*Gibberella*	XP_384861	Y	*Rhizobium*	ZP_03505301	Y
Arthrobacter	YP_949837	Y	*Gibberella*	XP_388213	Y	*Rhizobium*	YP_470103	Y
Arthrobacter	YP_002489952	N	*Glomerella*	EFQ27291	Y	*Rhizobium*	YP_001978859	Y
Arthrobacter	YP_833604	Y	*Glomerella*	EFQ34262	N	*Rhizobium*	ZP_03530195	Y
Aspergillus	CBF87292	N	*Glomerella*	EFQ35214	Y	*Rhizobium*	ZP_03524635	N
Aspergillus	XP_658905	Y	*Haloferax*	YP_003536123	N	*Rhodococcus*	YP_004008885	Y
Aspergillus	XP_661807	Y	*Herpetosiphon*	YP_001546287	Y	*Rhodococcus*	YP_004008886	Y
Aspergillus	XP_662932	Y	*Herpetosiphon*	YP_001547597	N	*Rhodococcus*	ZP_06830704	Y
Aspergillus	XP_663002	Y	*Hyphomicrobium*	YP_003756407	N	*Rhodococcus*	ZP_06830705	Y
Aspergillus	XP_664765	Y	*Janibacter*	ZP_00995526	N	*Rhodococcus*	AAN37493	N
Aspergillus	XP_001389482	N	*Janibacter*	ZP_00995926	N	*Rhodococcus*	AAR27826	Y
Aspergillus	XP_001389508	Y	*Ketogulonicigenium*	YP_003964856	Y	*Riemerella*	YP_004044931	N
Aspergillus	XP_001390849	Y	*Kineococcus*	YP_001363667	N	*Roseibium*	ZP_07660688	Y
Aspergillus	XP_001391162	N	*Kitasatospora*	BAJ27524	N	*Roseiflexus*	YP_001431698	Y
Aspergillus	XP_001393826	Y	*Kitasatospora*	BAJ31863	N	*Roseiflexus*	YP_001275304	Y
Aspergillus	XP_001395585	N	*Kribbella*	YP_003381075	N	*Saccharopolyspora*	YP_001104202	Y
Aspergillus	XP_001395793	Y	*Laccaria*	XP_001879555	Y	*Saccharopolyspora*	YP_001105681	Y
Aspergillus	XP_001396120	Y	*Limnobacter*	ZP_01915839	Y	*Sanguibacter*	YP_003314415	N
Aspergillus	XP_001399989	Y	*Magnaporthe*	XP_363006	Y	*Schistosoma*	XP_002569539	N
Aspergillus	CAK37271	Y	*Marinomonas*	ZP_01076090	Y	*Schizophyllum*	XP_003026007	Y
Aspergillus	CAK40367	Y	*Maritimibacter*	ZP_01013228	Y	*Schizophyllum*	XP_003026084	Y

Table S1. Cont.

Genus*	Accession no.†	S_p‡	Genus*	Accession no.†	S_p‡	Genus*	Accession no.†	S_p‡
Aspergillus	CAK41421	Y	*Meiothermus*	YP_003505838	N	*Schizophyllum*	XP_003029142	N
Aspergillus	CAK44937	Y	*Mesorhizobium*	YP_004139576	Y	*Schizophyllum*	XP_003029311	N
Aspergillus	CAK46228	Y	*Mesorhizobium*	YP_675384	Y	*Schizophyllum*	XP_003031024	Y
Asticcacaulis	YP_004087964	N	*Methylobacterium*	YP_002423474	Y	*Schizophyllum*	XP_003035277	N
Beutenbergia	YP_002880947	N	*Methylosinus*	ZP_06887255	Y	*Sclerotinia*	XP_001584708	Y
Botryotinia	XP_001547782	N	*Methylovorus*	YP_004039767	Y	*Sclerotinia*	XP_001590335	Y
Botryotinia	XP_001550224	Y	*Methylovorus*	YP_003051066	Y	*Sinorhizobium*	YP_002823995	Y
Bradyrhizobium	YP_001243151	Y	*Micrococcus*	YP_002956532	N	*Sinorhizobium*	YP_001312810	Y
Bradyrhizobium	YP_001202885	Y	*Micrococcus*	ZP_06500982	N	*Sorangium*	YP_001617686	N
Brevibacterium	ZP_05914095	N	*Micromonospora*	ZP_04603856	N	*Sorangium*	YP_001617690	N
Burkholderia	ZP_03270367	N	*Moniliophthora*	XP_002390944	N	*Spirosoma*	YP_003387063	Y
Caulobacter	YP_003594737	Y	*Mucilaginibacter*	ZP_07746550	Y	*Stigmatella*	ZP_01459590	Y
Cellvibrio	YP_001983304	N	*Naegleria*	XP_002681734	Y	*Streptomyces*	YP_003490106	N
Chaetomium	XP_001223092	Y	*Nakamurella*	YP_003203127	N	*Streptomyces*	YP_003490270	Y
Chaetomium	XP_001224578	Y	*Nectria*	XP_003040987	Y	*Streptomyces*	YP_003490910	N
Chaetomium	XP_001228656	Y	*Nectria*	XP_003043843	Y	*Streptomyces*	ZP_06275428	N
Chitinophaga	YP_003120875	Y	*Nectria*	XP_003047595	Y	*Streptomyces*	ZP_07284524	N
Chloroflexus	YP_002463143	N	*Nectria*	XP_003052962	Y	*Streptomyces*	ZP_07285769	N
Chloroflexus	YP_001635170	N	*Neosartorya*	XP_001257499	Y	*Streptomyces*	ZP_06708934	N
Clavibacter	YP_001221487	N	*Neosartorya*	XP_001258835	Y	*Streptomyces*	ZP_06712030	N
Comamonas	YP_003280206	N	*Neosartorya*	XP_001263001	Y	*Streptomyces*	ZP_04999238	Y
Comamonas	ZP_03541192	N	*Neosartorya*	XP_001265559	Y	*Streptomyces*	ZP_06916827	N
Comamonas	ZP_07046125	N	*Neurospora*	XP_963883	Y	*Streptomyces*	ZP_06918248	Y
Conexibacter	YP_003395019	N	*Nitrosococcus*	YP_003527272	N	*Streptomyces*	ZP_06918456	N
Coprinopsis	XP_001831803	Y	*Nitrosococcus*	YP_343162_Y	Y	*Streptomyces*	ZP_06918464	Y
Coprinopsis	XP_001834886	Y	*Nitrosococcus*	YP_344402	N	*Streptosporangium*	YP_003343260	N
Corynebacterium	NP_599490	N	*Nitrosospira*	YP_412780	Y	*Talaromyces*	XP_002478049	Y
Corynebacterium	YP_001137177	N	*Nocardia*	YP_120510	N	*Talaromyces*	XP_002488138	Y
Corynebacterium	ZP_07404061	Y	*Novosphingobium*	YP_497150	N	*Thermobaculum*	YP_003323670	N
Corynebacterium	ZP_07990861	N	*Paracoccidioides*	XP_002796777	Y	*Thermus*	ZP_03495684	Y
Cryptococcus	XP_572061	Y	*Paracoccidioides*	EEH19263	Y	*Thermus*	YP_005116	N
Cupriavidus	YP_587403	Y	*Paracoccidioides*	EEH48111	Y	*Thiobacillus*	YP_314960	N
Cupriavidus	YP_002007309	Y	*Pectobacterium*	YP_003019323	Y	*Tuber*	XP_002840547	Y
Debaryomyces	XP_002770542	Y	*Pelagibaca*	ZP_01444072	N	*Ustilago*	XP_760258	Y
Deinococcus	YP_002784902	N	*Penicillium*	XP_002556982	Y	*Ustilago*	XP_762135	Y
Deinococcus	DYP_002787243	N	*Penicillium*	XP_002559399	Y	*Verticillium*	XP_003005923	Y
Deinococcus	YP_594344	N	*Penicillium*	XP_002563845	Y	*Xenorhabdus*	YP_003467843	Y
Delftia	YP_001564166	N	*Phaeosphaeria*	XP_001791242	Y	*Xenorhabdus*	YP_003712511	Y
Dyadobacter	YP_003088138	Y	*Phaeosphaeria*	XP_001791707	Y	*Yersinia*	NP_670676	N
Flavobacteriales	ZP_02181399	Y	*Photorhabdus*	YP_003042719	N	*Yersinia*	YP_001399705	N
Flavobacterium	YP_001194031	Y	*Photorhabdus*	NP_931592	Y	*Yersinia*	YP_071754	N

**T. urticae* accession numbers can be accessed through the BOGAS genome portal (http://bioinformatics.psb.ugent.be/webtools/bogas/), other accession numbers are available at GenBank.

†Color code: bacteria, green; fungi, purple; Archaea, red; Chromalveolata, dark blue; Excavata, light blue; Animalia, yellow.

‡S_p: predicted with (yes, Y) or without (no, N) a signal peptide according to SignalP 3.0 (8).

Table S2. *T. urticae* lipocalin properties

Tetur ID*	β-strands/helices†	Intron phase pattern‡	S_p§	Length, aa	Strand	OrthoMCL
tetur01g01500	7+1/1+1	0_1_1	Y	190	+	19288
tetur01g01510	8+1/1+1	0_1_1	Y	195	+	19288
tetur01g01520	9+1/1+1	0_1_1	Y	187	+	19288
tetur01g05730	8+1/2+1	0_1_1	Y	208	+	10134
tetur01g05740	8+1/1+1	0_1_1	Y	192	+	10134
tetur01g05750	8+1/1+1	0_1_1	Y	187	–	10134
tetur01g05770	8+1/1+1	0_1_1	Y	187	+	10134
tetur01g16584	8+1/1+1	0_1_1	Y	195	+	10134
tetur06g03550	8/1+1	0_1_1	Y	192	+	10134
tetur09g04720	8+1/1+1	0_1_1	Y	198	+	10134
tetur09g04730	8+1/1+1	0_1_1	Y	195	+	10134
tetur09g04920	8+1/1+1	0_1_1	Y	194	–	10134
tetur30g01430	8+1/2+1	0_1_1	Y	218	–	10134
tetur31g00680	8+1/1+1	0_1_1	Y	199	–	10134
tetur31g00780	8+1/1+1	0_1_1	N	177	+	10134
tetur31g00880	8+1/1+1	0_1_1	Y	199	+	10134
tetur31g00900	8+1/1+1	0_1_1	Y	199	+	10134
tetur31g00920	8+1/1+1	0_1_1	Y	199	+	10134
tetur02g11960	8+2/1+1	0_2_1	Y	218	–	15855
tetur37g00940	8+1/1+1+1	0_2_1	Y	213	+	19721
tetur04g05970	8+1/1+1	0_2_1_1	Y	193	+	10107
tetur04g05980	8+1/1+1	0_2_1_1	Y	187	+	10107
tetur04g06000	8+1/1+1	0_2_1_1	Y	194	+	10107
tetur04g06010	8/1+1	0_2_1_1	Y	182	–	10107
tetur05g07070	8+1/1+1	0_2_1_1	Y	194	+	10134
tetur06g01610	8/1+1	0_2_1_1	Y	167	–	10107
tetur06g01640	8+1/1+1	2_1_1	Y	184	–	10134
tetur06g02130	8+1/1+1	0_2_1_1	Y	183	–	10107
tetur06g02140	8+1/1+1	0_2_1_1	Y	187	–	10107
tetur06g02670	8/1+1	0_2_1_1	Y	179	–	10107
tetur06g02940	2+1+1/2+2+1	0_2_1_1	Y	149	+	10107
tetur06g03020	8+2/1+1	0_2_1_1	Y	192	–	10107
tetur06g03030	3+1+1/1+2+1	0_2_1_1	Y	162	–	10107
tetur06g03070	3+1+1/2+2+1+1	0_2_1_1	Y	172	+	10107
tetur06g03090	8/1+1	0_2_1_1	Y	165	+	10107
tetur06g03100	8+1/1+1	0_2_1_1	Y	184	+	10107
tetur06g03340	8+1/1+1	0_2_1_1	Y	185	–	10107
tetur06g03350	8/1+1	0_2_1_1	Y	179	+	10107
tetur06g03360	8+1/1+1	0_2_1_1	Y	183	–	10107
tetur06g03370	8+1/1+1	0_2_1_1	Y	183	+	10107
tetur06g03440	8+1/1+1	0_2_1_1	Y	188	–	10107
tetur06g03530	8/2+2	0_2_1_1	Y	181	+	10107
tetur06g03860	8+1/1+1	0_2_1_1	Y	191	+	10134
tetur06g06691	8+1/1+1	0_2_1_1	Y	172	+	10107
tetur07g03960	9+1/1+1	0_2_1_1	Y	201	–	10134
tetur07g03970	8+1/1+1	0_2_1_1	Y	199	+	21421
tetur11g05210	8+1/1+1	0_2_1_1	Y	187	–	10107
tetur11g05230	8+1/1+1	0_2_1_1	Y	187	–	10107
tetur16g03410	8+1/1+1	0_2_1_1	Y	184	+	10107
tetur16g03450	8+1/1+1	0_2_1_1	Y	184	–	10107
tetur16g03460	8+1/1+1	0_2_1_1	Y	184	+	10107
tetur174g00050	8+1/1+1	0_2_1_1	Y	199	–	21421
tetur18g00900	8+1/1+1	0_2_1_1	Y	200	–	10107
tetur21g03160	7/1+1	0_2_1_1	Y	174	–	10107
tetur21g03190	8/1+1	0_2_1_1	Y	171	–	10107
tetur21g03340	8/1+1	0_2_1_1	Y	175	+	10107
tetur24g01030	8+1/3+1	0_2_1_1	Y	227	+	10134
tetur282g00020	8+1/1+1	0_2_1_1	Y	193	+	10107

**T. urticae* accession numbers can be accessed through the BOGAS genome portal (http://bioinformatics.psb.ugent.be/webtools/bogas/).

†Secondary structures were predicted using JPred3 (18); predicted β strands ("E") and helices ("H") were counted if they were at least 3 aa long and separated by at least 3 aa; "8+1/1+1" represents the typical repeated +1 topology β-barrel structure of lipocalins with a 3_{10}-like and α-helix at the N-and C-terminal ends, respectively.

‡The intron phase of each intron is separated by an underscore.

§S_p: predicted with (Y) or without (N) a signal peptide using SignalP 3.0 (8).

Table S3. Accession numbers of lipocalins used for phylogenetic analysis

Species	Taxonomy	Name	Accession no.
Tetranychus urticae (58)	Arachnida: Acari: Trombidiformes	—	see Table S2
Ahrensia sp.	Alphaproteobacteria: Rhodobacteriales	Outer membrane lipoprotein Blc	ZP_07375698.1
Homarus gammarus	Crustacea: Decapoda	Crustacyanin-A1	P58989.1
H. gammarus	Crustacea: Decapoda	Crustacyanin-A2	P80007.1
Debaromyces hansenii	Fungi: Ascomycota	—	XP_460369.1
Klebsiella pneumoniae	Proteobacteria: Enterobacteriales	Lipoprotein Blc	ZP_06018160.1
Anopheles gambiae	Insecta: Diptera	Agam.Lip1	XP_320076.4
Drosophila melanogaster	Insecta: Diptera	Neural lazarillo	AAF51378.2
Drosophila melanogaster	Insecta: Diptera	Glial lazarillo	NP_523727.2
D. melanogaster	Insecta: Diptera	Karl	AAK72697.1
Rhodnius prolixus	Insecta: Hemiptera	Biogenic amine-binding protein	AAO25746.1
R. prolixus	Insecta: Hemiptera	*Rhodnius* platelet aggregation inhibitor 3	AAQ20817.1
R. prolixus	Insecta: Hemiptera	*Rhodnius* platelet aggregation inhibitor 4	AAQ20818.1
R. prolixus	Insecta: Hemiptera	Nitrophorin-1	Q26239.1
R. prolixus	Insecta: Hemiptera	Pallidipin salivary platelet aggregation inhibitor 1	AAB09090.1
R. prolixus	Insecta: Hemiptera	Triabin-like lipocalin 1	AAQ20821.1
Triatoma brasiliensis	Insecta: Hemiptera	Salivary triabin 1	ABH09425.1
T. brasiliensis	Insecta: Hemiptera	Pallidipin precursor	ABH09434.1
Triatoma pallidipennis	Insecta: Hemiptera	Pallidipin 2	AAA30329.1
T. pallidipennis	Insecta: Hemiptera	Triabin	CAA56540.1
Apis mellifera	Insecta: Hymenoptera	Amel_Lip1	XP_392555.1
A. mellifera	Insecta: Hymenoptera	Amel_Lip2	XP_623787.2
Harpegnathos saltator	Insecta: Hymenoptera	Apolipoprotein D	EFN80483.1
Bombyx mori	Insecta: Lepidoptera	Bombyrin	NP_001036872.1
B. mori	Insecta: Lepidoptera	32-kDa apolipoprotein	NP_001140192.1
Galleria mellonella	Insecta: Lepidoptera	Gallerin	AAA85089.1
Hyphantria cunea	Insecta: Lepidoptera	Hyphantrin	AAM18117.2
Lonomia obliqua	Insecta: Lepidoptera	Lipocalin 1 (biliprotein)	AAV91447.1
L. obliqua	Insecta: Lepidoptera	Lipocalin 4 (biliprotein)	AAV91423.1
Manduca sexta	Insecta: Lepidoptera	Insecticyanin 1	P00305.1
M. sexta	Insecta: Lepidoptera	Insecticyanin 2	Q00630.1
Pieris brassicae	Insecta: Lepidoptera	Bilin-binding protein	CAA54063.1
Samia cynthia ricini	Insecta: Lepidoptera	Biliverdin-binding protein I	BAB85482.1
S. cynthia ricini	Insecta: Lepidoptera	Biliverdin-binding protein II	BAB84676.1
Shistocerca americana	Insecta: Orthoptera	Lazarillo	CAA86216.1
Arabidopsis thaliana	Plantae: Brassicaceae	—	AAM62904.1
Zea mays	Plantae: Poaceae	—	NP_001140887.1
Oryza sativa	Plantae: Poaceae	—	NP_001047416.1
Bos taurus	Vertebrata: Mammalia	Apolipoprotein D	DAA33372.1
Homo sapiens	Vertebrata: Mammalia	Apolipoprotein D	NP_001638.1
Mus musculus	Vertebrata: Mammalia	Apolipoprotein D	CAA57974.1

Table S4. Classification of *T. urticae* MFS transporter genes (from OrthoMCL 10032, 10082, and 10236) determined by BLASTp in the Transporter Classification DataBase (20)

Tetur ID*	Length, aa	TM†	S_p‡	TCID name of best blastp hit at TCDB§	TCID	*E*-value	TCDB family
10236							
tetur04g02320	505	10	N	Sialin	2.A.1.14.10	*e*-52	Anion:Cation Symporter (ACS) family
tetur05g04460	538	10	N	Brain synaptic vesicle anion:Na^+ symporter	2.A.1.14.13	*e*-42	ACS family
tetur08g06320	530	11	N	Brain synaptic vesicle anion:Na^+ symporter	2.A.1.14.13	*e*-43	ACS family
tetur08g06330	508	12	N	Sialin	2.A.1.14.10	*e*-52	ACS family
tetur08g06340	527	11	N	Sialin	2.A.1.14.10	*e*-52	ACS family
tetur08g06350	512	11	N	Sialin	2.A.1.14.10	*e*-46	ACS family
tetur08g06360	503	11	N	Sialin	2.A.1.14.10	*e*-46	ACS family
tetur08g06370	508	9	N	Sialin	2.A.1.14.10	*e*-55	ACS family
tetur08g06410	512	11	N	Sialin	2.A.1.14.10	*e*-57	ACS family
tetur08g06870	529	8	N	Sialin	2.A.1.14.10	*e*-58	ACS family
tetur08g06890	527	9	N	Sialin	2.A.1.14.10	*e*-61	ACS family
tetur09g02840	489	12	N	Sialin	2.A.1.14.10	*e*-43	ACS family
tetur17g01270	522	11	N	Sialin	2.A.1.14.10	*e*-54	ACS family
tetur17g01300	530	10	N	Sialin	2.A.1.14.10	*e*-50	ACS family
tetur05g04570	353	5	N	Sialin	2.A.1.14.10	*e*-27	ACS family
tetur05g04590	498	11	N	Sialin	2.A.1.14.10	*e*-36	ACS family
tetur05g04600	280	6	N	brain synaptic vesicle anion:Na^+ symporter	2.A.1.14.13	*e*-17	ACS family
tetur08g06400	447	7	N	brain synaptic vesicle anion:Na^+ symporter	2.A.1.14.13	*e*-45	ACS family
tetur60g00010	240	6	N	Sialin	2.A.1.14.10	*e*-28	ACS family
tetur533g00010	265	5	N	Sialin	2.A.1.14.10	*e*-32	ACS family
10082							
tetur01g00770	486	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-20	Fucose: H^+ Symporter (FHS) family
tetur01g06570	444	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-13	FHS family
tetur01g06580	441	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-22	FHS family
tetur01g06600	441	10	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-26	FHS family
tetur01g08660	463	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-18	FHS family
tetur01g10420	457	10	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-21	FHS family
tetur01g10430	457	10	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-20	FHS family
tetur01g15760	460	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-24	FHS family
tetur02g11610	453	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-16	FHS family
tetur02g13340	467	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-09	FHS family
tetur04g03830	452	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-16	FHS family
tetur04g07010	458	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-21	FHS family
tetur05g05100	443	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-22	FHS family
tetur07g06710	443	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-19	FHS family
tetur07g06730	448	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-21	FHS family
tetur07g06740	447	10	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-24	FHS family
tetur10g00280	136	2	N	Putative vanillate porter	2.A.1.15.6	0.0025	FHS family
tetur10g00320	453	11	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-17	FHS family
tetur13g02590	442	10	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-05	FHS family
tetur13g02600	442	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-10	FHS family
tetur13g02620	442	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-14	FHS family
tetur15g00220	465	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-22	FHS family
tetur18g01050	445	10	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-19	FHS family
tetur19g02710	455	11	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-12	FHS family
tetur21g00030	475	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-25	FHS family
tetur21g00050	475	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-21	FHS family
tetur21g00410	458	10	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-17	FHS family
tetur21g00420	461	10	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-22	FHS family
tetur21g00500	461	10	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-18	FHS family
tetur21g00590	142	4	N	E-protein viroporin	1.A.65.1.2	0.0685	E-protein viroporin
tetur22g02590	449	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-13	FHS family
tetur24g01860	448	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-15	FHS family
tetur24g02320	449	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-09	FHS family
tetur39g00620	444	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-14	FHS family
tetur40g00020	484	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-37	FHS family
tetur40g00030	436	12	N	Na^+-dependent glucose transporter	2.A.1.7.4	*e*-20	FHS family

Table S4. Cont.

Tetur ID*	Length, aa	TM†	S_p‡	TCID name of best blastp hit at TCDB§	TCID	*E*-value	TCDB family
10032							
tetur02g07750	490	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0001	PCFT/HCP family
tetur02g07790	476	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0033	PCFT/HCP family
tetur02g07920	490	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0002	PCFT/HCP family
tetur02g10410	559	12	N	Quinolone resistance protein norA	2.A.1.2.10	0.0117	Drug:H^+ Antiporter-1 (DHA1) family
tetur03g00780	507	11	N	Tetracycline:H^+ antiporter	2.A.1.2.4	0.3359	DHA1 family
tetur03g02680	508	10	N	Solute carrier family 46, member 3	2.A.1.50.3	*e*-06	PCFT/HCP family
tetur03g02740	507	11	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	*e*-05	PCFT/HCP family
tetur03g02780	506	10	N	TSCOT	2.A.1.50.2	0.0002	PCFT/HCP family
tetur03g02800	512	12	N	Tetracycline:H^+ antiporter	2.A.1.2.4	*e*-05	DHA1 family
tetur03g02880	508	10	N	Proton-coupled folate transporter/Heme carrier protein	2.A.1.50.1	*e*-05	PCFT/HCP family
tetur03g02890	508	10	N	Tetracycline:H^+ antiporter	2.A.1.2.4	*e*-05	DHA1 family
tetur03g04410	574	12	N	Solute carrier family 46, member 3	2.A.1.50.3	*e*-05	PCFT/HCP family
tetur04g02440	491	10	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0002	PCFT/HCP family
tetur04g06740	550	12	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0001	PCFT/HCP family
tetur06g05250	599	10	N	Tetracycline-specific exporter, TetA39 (most like 2.A.1.2.4)	2.A.1.2.38	0.0044	DHA1 family
tetur08g00460	496	12	N	Solute carrier family 46, member 3	2.A.1.50.3	*e*-05	PCFT/HCP family
tetur08g00470	501	11	N	Proton-coupled folate transporter/Heme carrier protein	2.A.1.50.1	*e*-09	PCFT/HCP family
tetur08g00480	534	10	N	Proton-coupled folate transporter/Heme carrier protein	2.A.1.50.1	*e*-05	PCFT/HCP family
tetur08g02850	617	11	N	TSCOT	2.A.1.50.2	*e*-06	PCFT/HCP family
tetur08g04870	511	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0003	PCFT/HCP family
tetur09g02400	492	11	N	Proton-coupled folate transporter/Heme carrier protein	2.A.1.50.1	0.0502	PCFT/HCP family
tetur09g02410	525	11	N	Glycolipid translocase (floppase) Spr1816/Spr1817	3.A.1.142.1	0.0541	Glycolipid Flippase (G.L.Flippase) family
tetur09g02420	510	10	N	Cu+ /Ag+ efflux pump	2.A.6.1.4	0.0682	Heavy Metal Efflux (HME) family
tetur09g02430	525	10	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0109	PCFT/HCP Family
tetur11g00550	537	8	N	(Mono- and divalent organocation):H^+ antiporter.	2.A.1.3.4	0.0017	Drug:H^+ Antiporter-2 (14 Spanner) (DHA2) family
tetur11g05100	501	11	N	TSCOT	2.A.1.50.2	*e*-05	PCFT/HCP family
tetur11g05110	500	11	N	TSCOT	2.A.1.50.2	0.0001	PCFT/HCP family
tetur11g05410	506	10	N	Solute carrier family 46, member 3	2.A.1.50.3	*e*-05	PCFT/HCP family
tetur11g06110	528	9	N	TSCOT	2.A.1.50.2	*e*-06	PCFT/HCP family
tetur13g03400	502	10	N	Solute carrier family 46, member 3	2.A.1.50.3	*e*-05	PCFT/HCP family
tetur14g02020	564	12	N	TSCOT	2.A.1.50.2	0.0014	PCFT/HCP family
tetur15g01050	489	12	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0131	PCFT/HCP family
tetur16g01660	566	12	N	MFS permease of unknown function	2.A.1.66.1	0.007	Unidentified Major Facilitator-15 (UMF15) Family
tetur16g03200	499	8	N	TetA42 from a deep terrestrial subsurface bacterium	2.A.1.2.41	0.0228	DHA1 family
tetur16g03270	497	10	N	Solute carrier family 45, member 4	2.A.2.4.7	0.0078	GPH:Cation Symporter family
tetur17g03580	499	12	N	TSCOT	2.A.1.50.3	*e*-10	PCFT/HCP family
tetur17g00880	606	12	N	TSCOT	2.A.1.50.2	0.0375	PCFT/HCP family
tetur17g00890	541	12	N	Solute carrier family 45, member 4	2.A.1.50.3	0.0051	PCFT/HCP family
tetur20g03210	501	11	N	Tetracycline:H^+ antiporter	2.A.1.2.4	0.0021	DHA1 family
tetur30g01490	471	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0002	PCFT/HCP family
tetur30g02070	549	12	N	Proton-coupled folate transporter/Heme carrier protein	2.A.1.50.1	0.0256	PCFT/HCP family
tetur02g07760	432	10	Y	Solute carrier family 46, member 3	2.A.1.50.3	0.0008	PCFT/HCP family
tetur03g02670	372	7	N	TSCOT	2.A.1.50.2	0.0002	PCFT/HCP family
tetur03g02770	166	4	N	—	—	—	
tetur03g02820	342	10	N	TSCOT	2.A.1.50.2	0.0003	PCFT/HCP PCFT/HCP family
tetur03g04360	194	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0036	PCFT/HCP PCFT/HCP family

Table S4. Cont.

Tetur ID*	Length, aa	TM†	S_p‡	TCID name of best blastp hit at TCDB§	TCID	*E*-value	TCDB family
tetur03g04370	157	3	N	Tetracycline resistance protein OS...	2.A.1.2.41	0.8993	DHA1 family
tetur03g04390	219	6	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0012	PCFT/HCP PCFT/HCP family
tetur11g05550	508	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.003	PCFT/HCP PCFT/HCP family
tetur16g03180	483	9	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0026	PCFT/HCP PCFT/HCP family
tetur17g03650	307	4	N	Lipopolysaccharide (colanic acid) exporter, WzxC	2.A.66.2.7	0.8899	Polysaccharide Transport (PST) family
tetur46g00180	355	8	N	Inner membrane protein YqcE	2.A.1.52.2	0.3687	Unknown Major Facilitator-8 (UMF8) family
tetur623g00010	141	4	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0012	PCFT/HCP PCFT/HCP family

*_T. urticae_ accession numbers can be accessed through the BOGAS genome portal (http://bioinformatics.psb.ugent.be/webtools/bogas/).
†Transmembrane regions were predicted using TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/).
‡S_p: Signal peptides were predicted using SignalP 3.0 (8).
§TCID, Transporter Class ID at the Transporter Class DataBase (TCDB) (20).

Table S5. Differentially expressed transcription factors in MR-VP, MAR-AB, and Tomato-5G

Tetur ID*	OrthoMCL cluster	PFAM domain (name) $\leq e^{-5}$	$\log_2$(FC)		
			MR-VP	MAR-AB	Tomato-5G
tetur08g07430	arthro11677	PF00010.21 (Helix–loop–helix DNA-binding domain), PF08447.6 (PAS fold domain)	0.00	0.00	−1.22
tetur08g07600	arthro13628	PF00010.21 (Helix–loop–helix DNA-binding domain), PF08447.6 (PAS fold domain)	0.00	−1.01	0.00
tetur03g03150	**arthro13720**	**PF00010.21 (Helix–loop–helix DNA-binding domain)**	**−1.08**	**−1.26**	**−1.67**
tetur01g11390	arthro18772	PF00010.21 (Helix–loop–helix DNA-binding domain)	1.62	0.00	1.96
tetur01g11460	arthro18772	PF00010.21 (Helix–loop–helix DNA-binding domain)	1.65	0.00	1.89
tetur01g11400	arthro18772	PF00010.21 (Helix–loop–helix DNA-binding domain)	0.00	−1.33	−1.09
tetur07g01420	Singleton	PF00010.21 (Helix–loop–helix DNA-binding domain)	1.61	0.00	0.00
tetur02g09640	arthro13477	PF00046.24 (Homeobox)	0.00	−1.21	0.00
tetur09g01820	arthro20299	PF00046.24 (Homeobox), PF00157.12 (Pou domain)	0.00	0.00	−1.03
tetur01g13830	arthro21308	PF00046.24 (Homeobox)	0.00	0.00	−1.17
tetur24g02450	**Singleton**	**PF00046.24 (Homeobox)**	**1.77**	**1.29**	**1.78**
tetur21g02710	Singleton	PF00046.24 (Homeobox)	0.00	−2.40	0.00
tetur01g09330	arthro10817	PF00076.17 (RNA recognition motif 1)	0.00	−1.19	0.00
tetur36g00260	**arthro18765**	**PF00104.25 (Ligand-binding domain of nuclear hormone receptor)**	**1.08**	**1.80**	**1.26**
tetur03g02550	Singleton	PF00104.25 (Ligand-binding domain of nuclear hormone receptor)	−4.16	0.00	0.00
tetur34g00430	Singleton	PF00105.13 (Zinc finger, C4 type)	0.00	0.00	−1.04
tetur03g03610	arthro21359	PF00170.16 (bZIP transcription factor)	0.00	0.00	1.15
tetur02g02610	arthro13949	PF00250.13 (Fork-head domain)	1.05	0.00	1.36
tetur14g02480	Singleton	PF00250.13 (Fork-head domain)	0.00	0.00	−1.06
tetur13g04490	Singleton	PF00412.17 (Lim domain)	1.12	0.00	0.00
tetur18g00570	arthro17249	PF00628.24 (PHD finger)	1.33	0.00	0.00
tetur18g00520	arthro17249	PF00628.24 (PHD finger)	0.00	−1.21	0.00
tetur07g01800	**arthro15823**	**PF00651.26 (BTB/POZ domain)**	**1.20**	**1.08**	**1.24**
tetur11g00680	Singleton	PF00651.26 (BTB/POZ domain)	−1.26	0.00	0.00
tetur11g01870	Singleton	PF00651.26 (BTB/POZ domain)	0.00	−1.14	0.00
tetur41g00270	Singleton	PF00751.13 (DM DNA-binding domain)	0.00	0.00	−1.02
tetur03g01600	arthro14230	PF08447.6 (PAS fold domain), PF00989.19 (PAS domain)	0.00	−1.01	0.00
tetur09g05790	Singleton	PF01388.16 (ARID/BRIGHT DNA-binding domain)	0.00	1.07	1.02
tetur21g01360	arthro18850	PF01426.13 (Bromo adjacent homology)	0.00	1.22	0.00
tetur35g00700	Singleton	PF02257.10 RFX (DNA-binding domain)	0.00	0.00	−1.00
tetur08g07670	Singleton	PF03166.9 (MAD homology 2 domain)	0.00	1.16	1.25
tetur05g03220	arthro18519	PF05920.6 (Homeobox KN domain)	0.00	0.00	−1.17
tetur01g09160	arthro15791	PF07716.10 (Basic Leucine Zipper Domain)	0.00	−1.14	0.00
tetur27g00310	arthro13959	PF10163.4 (Transcription factor e(y)2)	0.00	−1.03	0.00
tetur12g02760	arthro10838	PF13465.1 (Zinc-finger double domain)	−1.39	0.00	0.00
tetur01g07180	arthro21295	PF13465.1 (Zinc-finger double domain)	0.00	−1.26	0.00
tetur04g03720	arthro21379	PF13465.1 (Zinc-finger double domain)	0.00	0.00	−1.81
tetur01g09970	Singleton	PF13465.1 (Zinc-finger double domain)	0.00	0.00	−1.03
tetur03g05420	Singleton	PF13551.1 (Winged helix-turn helix)	0.00	−1.40	−1.26
tetur30g00520	Singleton	PF13815.1 (Iguana/Dzip1-like DAZ-interacting protein N-terminal)	0.00	1.10	0.00
tetur07g07990	arthro12151	PF13873.1 (Myb/SANT-like DNA-binding domain)	0.00	0.00	−1.35
tetur34g00350	arthro12151	PF13873.1 (Myb/SANT-like DNA-binding domain)	0.00	0.00	−1.35
tetur46g00100	arthro10157	—	0.00	0.00	1.37
tetur86g00030	arthro10157	—	−1.34	0.00	−1.01
tetur01g08230	arthro10157	—	−1.29	0.00	0.00
tetur01g08270	arthro10157	—	−1.27	0.00	0.00
tetur86g00020	arthro10157	—	−1.34	0.00	0.00
tetur12g02560	arthro15835	—	0.00	0.00	−1.28

Transcription factors differentially expressed in MR-VP, MAR-AB, and Tomato-5G are indicated in boldface type.

T. urticae accession numbers can be accessed through the BOGAS genome portal (http://bioinformatics.psb.ugent.be/webtools/bogas/); Tetur IDs shaded in gray are nuclear receptors according to ref. 2, and Tetur IDs indicated in boldface type are differentially expressed in MR-VP, MAR-AB, and Tomato-5G.

Table S6. Primers used in this study

Tetur ID	Gene name/family	Primer	Sequence, 5′–3′	qPCR product, nt	T_m, °C
tetur18g03590	Ribosomal protein 49	18g03590_q_F	CTTCAAGCGGCATCAGAGC	105	62.19
		18g03590_q_R	CGCATCTGACCCTTGAACTTC		62.09
tetur03g09480	Actin	03g09480_q_F	GCCATCCTTCGTTTGGATTTGGCT	113	69.94
		03g09480_q_R	TCTCGGACAATTTCTCGCTCAGCA		69.43
tetur03g04990	CYP392D2	03g04990_q_F	TTTAAAATCAGCACAGGGTAAATTG	152	60.58
		03g04990_q_R	CACTAACTTGCTTCATGTTGTGTCTT		60.92
tetur03g05070	CYP392D8	03g05070_q_F	TGAGCTCAGAAACGCGAAT	100	59.69
		03g05070_q_R	CTCGATTTCATGGGTTGCTT		60.07
tetur03g05110	CYP392D10	03g05110_q_F	ATTGGATTCGAACGTCAACC	109	59.80
		03g05110_q_R	GTAAATTAAGAGGAGTGATTGTTGCT		58.48
tetur06g04520	CYP392A16	06g04520_q_F	AAATACCGAGGTCGGACGTA	117	59.45
		06g04520_q_R	AAGCACTTTTTCAATCTGGTCAC		59.69
tetur06g04970	Short chain reductases	06g04970_q_F	TGCTGGTCTTCATGGTTTCC	146	61.05
		06g04970_q_R	TGACTATGTTTCCTGCAGATTTGT		60.06
tetur29g00220	TuGSTd14	29g00220_q_F	CTTTGGCAGATCTCACCGTAA	119	60.26
		29g00220_q_R	GTCAGCGTAGTTAGTTTCCAGTTG		59.43
tetur02g09840	Glycosyltransferases	02g09840_q_F	GTTTACTCAGCAAATCCTCTTGC	104	59.45
		02g09840_q_R	TCTCGGTAAATTTTCCAATAATCTC		59.02
tetur16g03220	MFS	16g03220_q_F	ATATCCGTGATCAGTGCAACA	140	59.02
		16g03220_q_R	AATTGGAACTATTGGCACAGC		59.10
tetur01g00490	ID-RCD	01g04900_q_F	CCGAAAAGCTCACCAACATTCAAG	81	66.04
		01g04900_q_R	CGTTTCAAGTCATCGGGAAGAAAG		65.20
tetur13g04550	ID-RCD	13g04550_q_F	CTGGCAAGCCAATGCTTTA	100	59.96
		13g04550_q_R	ACCTCTGAGGAATCTTTCACCA		60.11
tetur01g00490	ID-RCD	01g00490_s_F	CCTTTGTCTTGTTTTCATTACCG	712	59.92
		01g00490_s_R	TGGATCAATGGCGACTGTG		61.71
tetur04g08620	ID-RCD	04g08620_s_F	TCCGATCCCGATTATGTCTC	686	59.85
		04g08620_s_R	CTGGTAGGCTCAATACCAAGTG		58.79
tetur06g00460	ID-RCD	06g00460_s_F	GAAAGACCTGGTAAATTCGTTG	680	57.80
		06g00460_s_R	TCAGGATCAATTCCCAAAGTG		59.92
tetur07g02040	ID-RCD	07g02040_s_F	GCCTGTTGATTACTTCTTCCTTG	737	59.32
		07g02040_s_R	GGCCCACATTAAATTTGACCTATG		62.63
tetur07g05940	ID-RCD	07g05940_s_F	CCATTCCGAAAGATCCATTG	660	60.27
		07g05940_s_R	GCACTGTAACCATTACCTTCTGG		59.94
tetur12g04671	ID-RCD	12g04671_s_F	GATCGTACAATAGTTGATTGTGCTC	624	59.14
		12g04671_s_R	TCTGAAACCATTTTCCTGTGG		59.96
tetur19g03360	ID-RCD	19g03360_s_F	CGTTATTGTTACAGCCGATCC	831	59.49
		19g03360_s_R	GGATCGATACCGAATATCATGG		60.38
tetur20g01160	ID-RCD	20g01160_s_F	CGTACTGATTAGCCACCAATCC	730	60.74
		20g01160_s_R	AGATGTCATTCCCTCGGTTG		59.93
tetur20g01790	ID-RCD	20g01790_s_F	TTGGCTGTTACTTTTGGTGAC	812	57.80
		20g01790_s_R	GAACCAAAGTATCCTGTTTCTCG		59.19
tetur28g01250	ID-RCD	28g01250_s_F	TTTCCACTCATTCAGAGAGACC	710	58.38
		28g01250_s_R	TGTTAAAGTAGTGACTGTTGGATCG		59.65
tetur44g00140	ID-RCD	44g00140_s_F	TGAACCCAGGCCTATTGATG	850	60.85
		44g00140_s_R	TCCGAGTGTAATGATTCCTTG		57.69