Supplementary materials and methods

(a) Diapause incidence

In two replicates, 100 adult females from Jp-WT, Jp2-WT, Jp-lemon, and Jp-inbred-lemon were transferred to a detached bean leaf where they were allowed to lay eggs for 24 h. Upon egg hatching, larvae were placed in short-day conditions (17°C and 8:16 L:D photoperiod). The incidence of diapause was scored using the body coloration of adult female mites, a marker for the diapause state. To confirm this association in the current study, ten mites of the wild-type and mutant phenotypes that exhibited diapausing coloration were placed on individual leaf discs, and the number of eggs laid per female was recorded until mortality. The equality of diapause incidence between Jp-lemon and its parental strain, Jp-WT, was tested by a χ^2 test of independence in R (version 3.4.3) [1].

(b) Final extractions for TLC analysis

For the HPTLC plate shown in figure 2a, 0.90 g, 0.30 g and 0.45 g of kidney bean, Jp-inbred-lemon mites and Jp-WT mites were used, respectively. For the HPTLC plate shown in figure 2b, 0.95 g, 0.40 g and 0.45 g of kidney bean, Jp-inbred-lemon mites and Jp-WT mites were used, respectively.

(c) HPLC set-up

The HPLC system consisted of a reverse-phase column (2.0×250 mm, 5 µm; ODS-80Ts QA; Tosoh Co., Tokyo, Japan) and a photodiode array (SPD-M20A; Shimadzu Co., Kyoto, Japan). The mobile phase consisted of mixtures of methanol/water (90/10 v/v) containing 0.1% ammonium acetate (A) and ethyl acetate/methanol (30/70 v/v) containing 0.1% ammonium acetate (B). The flow rate was 0.2 ml/min, and the gradient elution was performed as follows: 0 to 5 min, 0% B; 5 to 20 min, 0 to 100% B linear; 20 to 35 min, 100% B; 35 to 40 min, 100 to 0% B linear; 40 to 45 min, 0% B.

(d) The generation of the segregating mite populations for BSA and fine-mapping

A single (haploid) male from the Jp2-WT strain was placed on a detached bean leaf and was offered three (diploid) females of the Jp-inbred-lemon line on a daily basis. A total of 30 females was fertilized and only produced wild-type female F1 mites (no lemon F1 females were observed). F1 males and females were allowed to sib-mate and produced wild-type and lemon F2 females. Due to our crossing scheme and the monogenic, recessive basis for the lemon phenotype (Table 1), wild-type F2 females were heterozygous for the lemon locus and over 1000 wild-type F2 females were selected to found the final segregating populations. The two populations were expanded and kept at very high population sizes (over 1000 mites) for approximately four months (roughly 10-12 generations) under laboratory conditions (28°C, 60% RH, and 16:8 L:D photoperiod).

(e) DNA preparation

Pooled samples were homogenized in 790 µl SDS buffer (200 mM Tris·HCl, 400 mM NaCl, 10 mM EDTA and 2% w/v SDS at pH 8.2) and 7 µl proteinase K (20mg/ml) and were incubated at 60°C for 2 h on a shaker (300 rpm). RNAse A (3µl of 100 mg/ml) was added and samples were incubated at 37°C for 2 h on a shaker (300 rpm). Total DNA was extracted using phenol:chloroform:isoamyl alcohol solution (25:24:1), and chloroform:isoamyl solution 49:1. After adding 0.7 volume of isopropanol, samples were centrifuged at 7,000 rcf for 1 hour. DNA pellets were washed with 70% ethanol and dissolved in 30 µl of DNase-free water. DNA samples were finally purified with an EZNA Cycle Pure Kit (Omega Bio-tek) and quantified with an Agilent TapeStation.

(f) SNP filtering for the BSA approach

To be used in our analysis, a variant had to have: (1) a quality score normalized by allele depth (QD) of 2 or higher, (2) mean root square mapping quality (MQ) of at least 50, (3) strand odds ratio (SOR) below 3, (4) mapping quality rank sum (MQRankSum) higher than or equal to -8, (5) rank sum for relative positioning of alleles in reads (ReadPosRankSumTest) of at least -8, and (6) be within 25% and 150% of the sample's genome-wide mean SNP read coverage to exclude potential false positive variants at loci with copy number variation (this was calculated using total depth per allele per sample, or AD).

(g) Single-mite DNA extraction

Individual mites were homogenized in 20 µl of PCR buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, with pH 8) containing 2 µl of proteinase K (20 mg/ml). Homogenates were incubated at 37°C for 30 min after which proteinase K was inactivated by incubating each sample at 95°C for 10 min.

(h) Fine-mapping the lemon locus

The SNP-based genetic markers were selected based on the BWA-alignments against the reference *Tetranychus urticae* genome assembly [2] and the resulting consensus sequences that were generated using SAMtools [3]. Illumina reads that aligned to these marker positions were extracted from the BAM files using count_coverage function of pysam with default settings (all reads used with a minimum base quality of 15). Frequencies of the nucleotide variants of Jp-inbred-lemon were determined for both parents based on the aligned reads and plotted along the ~630 kb region (supplementary figure 3 and supplementary table 1). Using contig 441 from the *de novo* assembly 'lemonBSA2_Assembly', a polymorphic fragment located ~250 bp upstream from the start of the *CYP384A1* coding sequence was identified and amplified using primers listed in supplementary table 1. Sanger sequence trace data was visualized using Chromas and sangerseqR in R (version 3.4.3) [1].

(i) Annotation of variation in the minimal candidate region

The effects of variants (SNPs and indels) on coding sequences in the minimal candidate region for lemon pigmentation, as assessed from variant calling, were predicted using SnpEff version 4.3 [4]. In addition, alignments within the candidate region for the lemon mutation were visually scanned using the Integrated Genomics Viewer (IGV, version 2.3.90) [5,6] as GATK does not reliably predict large indels and structural rearrangements.

(j) Sequencing *CYP384A1*

Using forward primer 5'-GTCGAAGTTGCTGTTTTTGC-3' and reverse primer 5'-TAAATTGTGAAGCCGACTGAAG-3', a *CYP384A1* DNA fragment was PCR-amplified that spanned the predicted exonic deletion for the 429 lemon and 50 wild-type adult females that were used to fine-map the lemon phenotype. RNA was collected from the Jp-lemon and Jp-WT strain as described in materials and methods. After DNase treatment (Ambion Turbo DNA-free kit, Ambion), 1 µg of cDNA was synthesized using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). The complete *CYP384A1* coding sequence of the two *T. kanzawai* strains was PCR-amplified using 5'-ATGACTCTTTTAACCTCGTTAACTTC-3' and 5'-TTACAATTTAATTGGACGGAGTAT-3' as the forward and reverse primer, respectively (DreamTaq, Thermo Fisher Scientific). PCR products were purified (Cycle Pure Kit, Omega Bio-tek) and cloned using the pGEM-T Easy Vector System (Promega). Plasmids were purified

using the Plasmid Mini Kit (Omega Bio-tek) and Sanger sequenced at Macrogen, Amsterdam, the Netherlands.

(k) Selection of cytochrome P450s for phylogenetic analysis

CYP3 clan members were retrieved from the genome assemblies of *T. urticae*, *Panonychus ulmi* (transcriptome), *Dermatophagoides farinae*, *Metaseiulus occidentalis*, *Ixodes scapularis*, *Tropilaelaps mercedesae*, *Parasteatoda tepidariorum*, *Limulus polyphemus*, *Daphnia pulex*, and *Hypsibius dujardini* [7–14]. Only the transcribed *T. mercedesae* CYP genes and annotated *I. scapularis* CYP copies were included [9,11]. The genome-based proteomes of *Sarcoptes scabiei* (version SscaA1.2, https://www.vectorbase.org), *Leptotrombidium deliense*, and *Dinothrombium tinctorium* were mined for homologues to CYP384A1 by a local BLASTp-search (E-value cut-off E-30) using *T. kanzawai* CYP384A1 as query [15,16]. To root the phylogenetic tree, CYP2 clan members of *T. urticae*, *M. occidentalis*, *L. polyphemus, D. pulex*, and *H. dujardini* were selected [8,10,12,14]. Pseudogenized *CYP* copies were removed and in the case of allelic variants or multiple sequences for a specific copy, the longest sequence was retained. Finally, the sequence set was filtered for a minimum sequence length of 200 amino acids.

Supplementary references

1. R Core Team. 2017 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

2. Wybouw N *et al.* 2019 Long-Term Population Studies Uncover the Genome Structure and Genetic Basis of Xenobiotic and Host Plant Adaptation in the Herbivore *Tetranychus urticae*. *Genetics* , genetics.301803.2018. (doi:10.1534/genetics.118.301803)

3. Li H *et al.* 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)

4. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w ^{1118} ; iso-2; iso-3. *Fly (Austin)* **6**, 80–92. (doi:10.4161/fly.19695)

5. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011 Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26. (doi:10.1038/nbt.1754)

6. Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013 Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192. (doi:10.1093/bib/bbs017)

7. Bajda S, Dermauw W, Greenhalgh R, Nauen R, Tirry L, Clark RM, Van Leeuwen T. 2015 Transcriptome profiling of a spirodiclofen susceptible and resistant strain of the European red mite *Panonychus ulmi* using strand-specific RNA-seq. *BMC Genomics* **16**. (doi:10.1186/s12864-015-2157-1)

8. Baldwin WS, Marko PB, Nelson DR. 2009 The cytochrome P450 (CYP) gene superfamily in *Daphnia pulex*. *BMC Genomics* **10**, 169. (doi:10.1186/1471-2164-10-169)

9. Dong X, Armstrong SD, Xia D, Makepeace BL, Darby AC, Kadowaki T. 2017 Draft genome of the honey bee ectoparasitic mite, *Tropilaelaps mercedesae*, is shaped by the parasitic life history. *GigaScience* **6**, 1–17. (doi:10.1093/gigascience/gix008)

10. Grbić M *et al.* 2011 The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* **479**, 487–492. (doi:10.1038/nature10640)

11. Gulia-Nuss M *et al.* 2016 Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat. Commun.* **7**, 10507. (doi:10.1038/ncomms10507)

12. Nelson DR. 2018 Cytochrome P450 diversity in the tree of life. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1866**, 141–154. (doi:10.1016/j.bbapap.2017.05.003)

13. Van Leeuwen T, Dermauw W. 2016 The Molecular Evolution of Xenobiotic Metabolism and Resistance in Chelicerate Mites. *Annu. Rev. Entomol.* **61**, 475–498. (doi:10.1146/annurev-ento-010715-023907)

14. Wu K, Hoy MA. 2016 The Glutathione-S-Transferase, Cytochrome P450 and Carboxyl/Cholinesterase Gene Superfamilies in Predatory Mite *Metaseiulus occidentalis*. *PLOS ONE* **11**, e0160009. (doi:10.1371/journal.pone.0160009)

15. Arlian LG, Morgan MS, Rider SD. 2016 *Sarcoptes scabiei*: genomics to proteomics to biology. *Parasit. Vectors* **9**. (doi:10.1186/s13071-016-1663-6)

16. Dong X *et al.* 2018 Genomes of trombidid mites reveal novel predicted allergens and laterally-transferred genes associated with secondary metabolism. *GigaScience* (doi:10.1093/gigascience/giy127)