Genome streamlining in a minute herbivore that manipulates its host plant

Robert Greenhalgh[1†], Wannes Dermauw[2†‡*], Joris J Glas[3§],
Stephane Rombauts[4,5], Nicky Wybouw[2], Jainy Thomas[6], Juan M Alba[3],
Ellen J Pritham[6], Saioa Legarrea[3], René Feyereisen[2,7], Yves Van de Peer[4,5,8],
Thomas Van Leeuwen[2], Richard M Clark[1,9*], Merijn R Kant[3*]

[1]School of Biological Sciences, University of Utah, Salt Lake City, United States;
[2]Laboratory of Agrozoology, Department of Plants and Crops, Faculty of Bioscience
Engineering, Ghent University, Ghent, Belgium; [3]Department of Evolutionary and
Population Biology, Institute for Biodiversity and Ecosystem Dynamics, University of
Amsterdam, Amsterdam, Netherlands; [4]Department of Plant Biotechnology and
Bioinformatics, Ghent University, Ghent, Belgium; [5]Center for Plant Systems
Biology, VIB, Ghent, Belgium; [6]Department of Human Genetics, University of Utah
School of Medicine, Salt Lake City, United States; [7]Department of Plant and
Environmental Sciences, University of Copenhagen, Copenhagen, Denmark; [8]Centre
for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and
Microbiology, University of Pretoria, Pretoria, South Africa; [9]Henry Eyring Center
for Cell and Genome Science, University of Utah, Salt Lake City, United States

*For correspondence:
wannes.dermauw@ugent.be (WD);
clark@biology.utah.edu (RMC);
M.Kant@uva.nl (MRK)

[†]These authors contributed
equally to this work

Present address: [‡]Flanders
Research Institute for
Agriculture, Fisheries and Food
(ILVO), Plant Sciences Unit,
Merelbeke, Belgium; [§]Rijk Zwaan
Breeding BV, De Lier, The
Netherlands

Abstract The tomato russet mite, *Aculops lycopersici,* is among the smallest animals on earth. It
is a worldwide pest on tomato and can potently suppress the host's natural resistance. We
sequenced its genome, the first of an eriophyoid, and explored whether there are genomic
features associated with the mite's minute size and lifestyle. At only 32.5 Mb, the genome is the
smallest yet reported for any arthropod and, reminiscent of microbial eukaryotes, exceptionally
streamlined. It has few transposable elements, tiny intergenic regions, and is remarkably intron-
poor, as more than 80% of coding genes are intronless. Furthermore, in accordance with ecological
specialization theory, this defense-suppressing herbivore has extremely reduced environmental
response gene families such as those involved in chemoreception and detoxification. Other losses
associate with this species' highly derived body plan. Our findings accelerate the understanding of
evolutionary forces underpinning metazoan life at the limits of small physical and genome size.

Introduction

The free-living microarthropod *Aculops lycopersici* (Tryon) belongs to the superfamily of the Erio-
phyoidea (Arthropoda: Chelicerata: Acari: Acariformes) that harbors the smallest plant-eating ani-
mals on earth (*Keifer, 1946*; *Navia et al., 2010*; *Sabelis and Bruin, 1996*). Eriophyoids are known
by many names including gall, blister, bud, and rust mites, depending on the type of damage they
cause (*Hoy, 2004*). Since the 1930s, the tomato russet mite *A. lycopersici* has been reported as a
minor pest of cultivated tomato (*Solanum lycopersicum* L.) worldwide (*Massee, 1937*). For unknown
reasons, it has emerged in recent years as a significant pest of tomatoes in European greenhouses
(*Moerkens et al., 2018*). While it is extremely small – only ~50 μm wide and 175 μm in length
(*Figure 1a,b*) – it can reach high population densities (*Figure 1c*). The damage it causes to plants
superficially resembles that of microbial disease (*Figure 1d*), for which it is often misdiagnosed, and
controlling it is troublesome (*Gerson and Weintraub, 2012*; *Van Leeuwen et al., 2010*).

eLife digest Arthropods are a group of invertebrates that include insects – such as flies or beetles – arachnids – like spiders or scorpions – and crustaceans – including shrimp and woodlice. One of the tiniest species of arthropods, measuring less than 0.2 millimeters, is the tomato russet mite *Aculops lycopersici*. This arachnid is among the smallest animals on Earth, even smaller than some single-celled organisms, and only has four legs, unlike other arachnids. It is a major pest on tomato plants, which are toxic to many other animals, and it feeds on the top cell layer of the stems and leaves. Tomato growers need a way to identify and treat tomato russet mite infestations, but this tiny species remains something of a mystery.

One way to tackle this pest may be to take a closer look at its genome, as this could reveal what genes the mite uses to detoxify its diet. Examining the mite's genome could also reveal information about how evolution handles creatures becoming smaller. An area of particular interest is the overall size of its genome. Not all of the DNA in a genome is part of genes that code for proteins; there are also sections of so-called 'non-coding' DNA. These sequences play important roles in controlling how and when cells use their genes. In the human genome, for example, just 1% of the DNA codes for protein. In fact, most human protein-coding genes are interrupted by sequences of non-coding DNA, called introns.

Here, Greenhalgh, Dermauw et al. sequence the entire tomato russet mite genome and reveal that not only is the mite's body size miniature: these tiny animals have the smallest arthropod genome reported to date, almost a hundred times smaller than the human genome. Part of this genetic miniaturization seems to be down to massive loss of non-coding DNA. Around 40% of the mite genome codes for protein, and 80% of its protein coding genes contain no introns. The rest of the miniaturization involves loss of genes themselves. The mites have lost some of the genes that determine body structure, which could explain why they have fewer legs than other arachnids. Additionally, they only carry a small set of genes involved in sensing chemicals and clearing toxins, which could explain why they are mostly found on tomato plants.

Greenhalgh, Dermauw et al.'s findings shed light on what may happen to the genome at the extremes of size evolution. Sequencing the genomes of other mites could reveal when in evolutionary history this genetic miniaturization occurred. Furthermore, a better understanding of the tomato russet mite genome could lead to the development of methods to detect the infestation of plants earlier and be highly beneficial for tomato agriculture.

The mite feeds on plant epidermal cells (*Royalty and Perring, 1988*), which are relatively low in nutrients, with needle-shaped mouth parts (stylets) that allow the transfer of saliva and the uptake of cell contents (*Nuzzaci and Alberti, 1996*). The first visible signs of a russet mite infestation are a rapid local collapse of the leaf hairs (trichomes) on the stem, leaflet or petiole upon which the mites are feeding (*van Houten et al., 2013*). This is followed by withering and necrosis of infested leaves, which ultimately leads to a bronzed or russet color, from which the mite owes its name (*Jeppson et al., 1975*; *Kawai and Haque, 2004*). Although it is now a global pest on tomato, it can survive on many related solanaceous plants (nightshade family) such as potato, tobacco, petunia, nightshade, and various peppers (*Perring and Farrar, 1986*), as well as on a few hosts outside the nightshade family (*Perring and Royalty, 1996*; *Rice and Strong, 1962*).

The Eriophyoidea belong to the Chelicerata, a subphylum of Arthropoda which includes spiders, scorpions, ticks, and mites. The Eriophyoidea consists of three families – Phytoptidae, Eriophyidae (or eriophyids, to which *A. lycopersici* belongs), and Diptilomiopidae, and comprises 357 herbivorous genera found on more than 1800 different plant species (*Oldfield, 1996*; *Zhang, 2011*). Eriophyoids are known to manipulate host plant resource allocation and resistance, and many species do so by inducing the formation of plant galls (*de Lillo et al., 2018*), possibly by secreting molecular mimics of plant hormones in their saliva (*De Lillo and Monfreda, 2004*; *de Lillo and Skoracka, 2010*). Although *A. lycopersici* is not a gall-inducing species, it nevertheless manipulates the defense mechanisms of its tomato host to its benefit. Through an unknown mechanism during feeding, this mite suppresses the jasmonic acid (JA) signaling pathway (*Glas et al., 2014*; *Schimmel et al., 2018*). This blocks the ability of the tomato host plant to produce defensive metabolites and proteins against
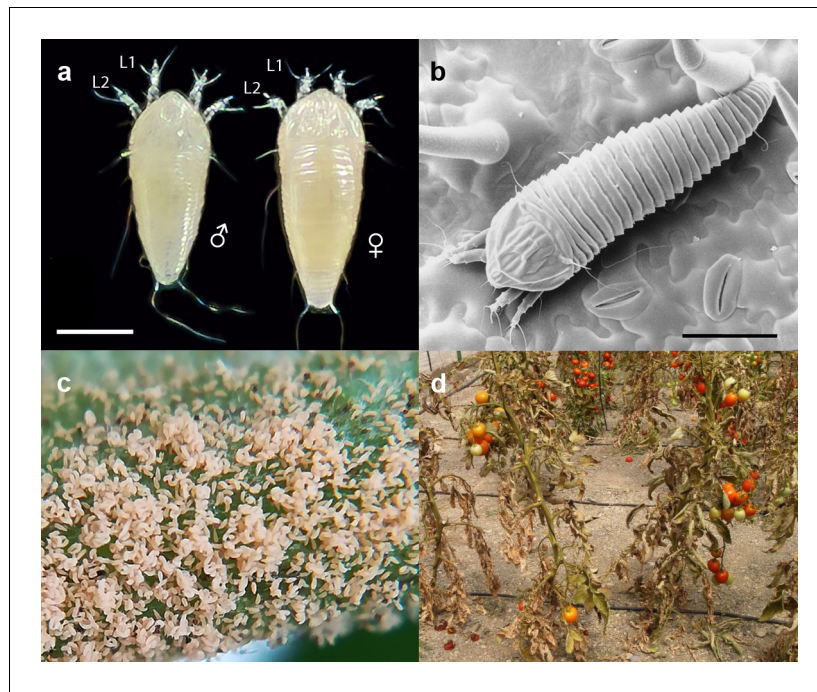
Figure 1. The tomato russet mite *Aculops lycopersici* is a devastating pest of tomato. (a) Habitus of the eriophyoid mite *A. lycopersici*. Male (left) and female (right) mites are slender, worm-like animals bearing, in contrast to non-eriophyoid mites with four pairs of legs, only two pairs of small legs (indicated by L1 and L2). (b) Low temperature (LT) - scanning electron microscopy (SEM) image of *A. lycopersici* on a leaf of *S. lycopersicum*. (c) *A. lycopersici* populations can rapidly build to extremely large numbers on tomato stems and leaves. (d) *A. lycopersici* damage of heavily infested tomato plants is shown. Scale bars in panels a and b represent 0.05 mm.

herbivorous insects and mites (*Alba et al., 2015*; *Howe and Jander, 2008*), thereby rendering the plant defenseless. The consequences of suppressing host defenses for the herbivore's selective environment may be variable depending on the degree of host specialization (*Blaazer et al., 2018*; *Kant et al., 2015*) but for mite species that can feed on multiple hosts, there are indications of a trade-off between the ability to suppress defenses and the ability to cope with xenobiotics (*Kant et al., 2008*; *Wybouw et al., 2015*). Many species of eriophyoid mites cause little damage to their hosts (*Jeppson et al., 1975*), or alternatively induce damage indirectly as vectors of pathogens (*Navia et al., 2013*). In contrast, while *A. lycopersici* is not known to vector plant diseases, its ability to alter the chemistry and morphology of tomato severely weakens the plants, which are then overwhelmed and killed by exponentially growing *A. lycopersici* populations (*Figure 1c,d*; *Perring, 1996*).

In addition to being a priority pest of tomato, *A. lycopersici* and related eriophyoids are among the most extreme examples of miniaturization in arthropods. As one of the smallest documented animal species (*Bailey and Keifer, 1943*), with dimensions smaller than some single-celled organisms (*Polilov, 2015*), it is not surprising that *A. lycopersici* has a derived morphology. Compared to almost all adult arachnids outside of the Eriophyoidea, which have a body plan with eight legs, *A. lycopersici* has only four legs (*Figure 1a,b*). Further, reproductive structures, which are located at the terminal end in other mites, are positioned in the central ventral region (*Nuzzaci and Alberti, 1996*). This type of morphology has resulted in altered reproductive behavior wherein males, instead of direct insemination, deposit spermatophores (packets of sperm) in the environment that are subsequently picked up by females (*Al-Azzazy and Alhewairini, 2018*; *Oldfield and Michalska, 1996*). Despite these morphological and behavioral innovations, *A. lycopersici* retains the haplodiploid mechanism of sex determination characteristic of many other mite species (*Anderson, 1954*). Further, female *A. lycopersici* mites can lay up to four eggs per day, and the generation time is as little as 5 days under optimal conditions (*Kawai and Haque, 2004*; *Rice and Strong, 1962*). These features, which resemble those of other agriculturally important mite herbivores, result in rapid

overexploitation of the host plant and have undoubtedly contributed to the importance of this species as a field and greenhouse pest of tomato.

Here, we present the genome of *A. lycopersici*, the first for an eriophyoid mite. At only 32.5 Mb, it is the smallest arthropod genome reported to date (*Grbić et al., 2011*; *Waldron et al., 2017*). As revealed by contrasting the genomic architecture of the tomato russet mite with other sequenced arthropods, including the two-spotted spider mite *Tetranychus urticae* (*Grbić et al., 2011*), a generalist herbivore often found in co-infestations alongside *A. lycopersici* (*Glas et al., 2014*), we elucidate mechanisms underlying dramatic genome reduction. In particular, we observed typical features of streamlined genomes (*Arkhipova, 2018*; *Hessen et al., 2010a*), including a marked reduction in the distance between adjacent genes, and few repetitive sequences. Massive loss of introns was apparent. Moreover, reductions in specific genes and gene families, such as environmental response genes, associate with *A. lycopersici*'s ability to suppress host plant defenses as well as its derived morphology. The genome therefore sheds light not only on mechanisms of extreme metazoan genome reduction, but also on the interplay between gene content and the lifestyle of small herbivores that manipulate their environment.

Results

Genome size, assembly, and annotation

We assembled the genome of *A. lycopersici* into seven scaffolds of cumulative length 32.53 Mb, of which 99.98% is represented on scaffolds 1–5 of lengths 12.44, 10.50, 3.66, 3.57 and 2.36 Mb, respectively. The remaining two scaffolds are each <6 kb in length, in addition to a mitochondrial genome scaffold. The observed assembly length is similar to the length estimated by a k-mer analysis with genomic sequence reads (34.81 Mb). Separate genome completeness estimates with CEGMA (*Parra et al., 2007*) and BUSCO (*Simão et al., 2015*) located 90.7% and 86.0% of the expected core eukaryotic genes, respectively; these values are within the same range as those for *T. urticae,* the only other sequenced chelicerate herbivore, and for which a high-quality Sanger assembly is available (95.16% and 92.07%, respectively). As an additional assessment of completeness, we generated a de novo assembly of the *A. lycopersici* transcriptome using deep, paired-end Illumina RNA-seq reads derived from mixed sex and developmental stages, and aligned it to the genome sequence. We found that 98.2% of transcript contigs could be located on the reference sequence. Of the remaining 243 unplaced transcript sequences, only eight had similarity to known arthropod sequences; the others had homology to bacterial, fungal, or plant sequences, or lacked homology to sequences in existing databases.

Features of extreme genome reduction in *A. lycopersici*

Annotation of the *A. lycopersici* genome by automated methods, coupled with extensive manual curation, revealed only 10,263 protein-coding genes. As assessed against other mite genomes, including *T. urticae*, *Dermatophagoides pteronyssinus* (the European house dust mite) (*Waldron et al., 2017*), and *Metaseiulus occidentalis* (a phytoseiid predatory mite) (*Hoy et al., 2016*), as well as the *Drosophila melanogaster* and human genomes, several features of genic organization in *A. lycopersici* stand out (*Table 1*). The fraction of the genome comprising coding sequence is highest in *A. lycopersici*, and the distance between genes is the lowest. Associated with the compact genic landscape of *A. lycopersici* (*Figure 2* and *Figure 2—figure supplements 1–6*), the percentage of the genome consisting of transposable elements was merely 1.54%, which is more than fourfold less than that observed in several other mite genomes, or in the insect *D. melanogaster* (*Figure 2—figure supplement 1*, *Supplementary file 1* — 'Table S1' Tab). Nevertheless, sequences homologous to the major classes of transposable elements, such as DNA transposons, including *Helitrons*, as well as both long terminal repeat (LTR) and non-LTR retrotransposons, were detected (*Supplementary file 1* — 'Table S1' Tab and 'Table S2' Tab). Across the *A. lycopersici* genome, extended regions of low genic composition and high TE density were not observed (*Figure 2—figure supplement 2*), consistent with the purported holocentric chromosome architecture (lack of regional centromeres) of eriophyoid mites (*Helle and Wysoki, 1996*).

We also observed that the *A. lycopersici* genome has only 3057 introns in coding sequences (CDS introns), which is more than an order of magnitude fewer than the 44,881 in the 90 Mb *T. urticae*

Table 1. Genome metrics for *A. lycopersici,* other mite species, *D. melanogaster* and *H. sapiens.*

Species	Genome size (Mb)	PCG*	% intronless†	Coding %‡	Intergenic %§	Intronic %¶	Intergenic M	Intron M
A. lycopersici	32.53	10,263	83.67	42.26	45.12	12.62	538 bp	170 bp
D. pteronyssinus	70.76	12,530	25.29	35.26	46.00	18.73	542 bp	75 bp
T. urticae	90.83	19,086	18.26	22.10	54.12	23.78	1302 bp	94 bp
M. occidentalis	151.90	17,310	24.97	15.25	59.14	25.61	2035 bp	135 bp
D. melanogaster	143.73	13,931	16.37	15.60	57.37	27.03	1228 bp	69 bp
H. sapiens	3088.27	19,636	6.74	1.10	68.14	30.77	23,279 bp	1,505 bp

*PCG: protein coding genes.

†Percent coding genes with no introns.

‡Percentage of genome in coding regions.

§Percentage of genome in between genes.

¶Percentage of genome in introns.

M = Median. See 'Genome metric calculations' in Materials and methods and **Table 1—source data 1** for more information.

The online version of this article includes the following source data for Table 1:

Source data 1. GFF3 annotation file of the *A. lycopersici* genome.

genome, and the 35,841 in the 70.8 Mb *D. pteronyssinus* genome. Strikingly, nearly 84% of *A. lycopersici* protein coding genes were intronless, which is more than threefold higher than observed for the other mite species we analyzed, and more than fivefold higher than for *D. melanogaster* (**Table 1**). To further investigate the dynamics of intron evolution, we evaluated patterns of intron gain and loss in orthologous genes among *A. lycopersici* and 17 other animal genomes using the Malin analysis pipeline (**Csurös, 2008**; **Figure 2**, and **Figure 2—figure supplements 3** and **4**, and **Supplementary file 2**). At 29,447 conserved intron sites (**Figure 2a**), *A. lycopersici* has a mere 207 introns. This is an ~11 fold reduction from that seen in the species with the next lowest counts, the European house dust mite *D. pteronyssinus*, at 2292. Apart from *A. lycopersici*, Acari intron loss rates were broadly similar to those observed for other arthropods, except for *M. occidentalis*, for which high rates of both intron loss and gain were apparent, a finding previously reported (**Hoy et al., 2016**). However, the rate of intron loss in *A. lycopersici* was higher than observed in *M. occidentalis* (**Figure 2b**), and in contrast to *M. occidentalis*, intron gains were minimal (**Figure 2—figure supplement 4**). The only evidence for retention of the minor spliceosome in *A. lycopersici* comes from the presence of a single canonical U12 (minor) intron in the gene *aculy03g00270* that encodes an ultra-conserved calcium channel (splice sites AT-AC in intron one of length 12.5 kb). Splicing of this large intron is supported by RNA-seq read alignments, and the orthologous intron one of the *T. urticae* orthologue of this gene is one of the three U12 introns documented previously in *T. urticae* (**Grbić et al., 2011**).

Although relatively few conserved introns are present in the *A. lycopersici* genome, they exhibit a bias toward 5' gene ends (**Figure 2—figure supplement 5**), and compared to most arthropods, the median intron length is larger (**Table 1** and **Figure 2—figure supplement 6**). In a single copy (orthologous) gene set for which introns were lost in *A. lycopersici*, but conserved in five other closely related or high-quality mite or insect genomes (see Materials and methods), the impact of intron loss on *A. lycopersici*-encoded protein sequences was generally minimal. In fact, in the respective protein sequences spanning 97 of 100 *A. lycopersici*-specific intron loss events (97%), multi-species alignments did not reveal insertions or deletions (indels) of amino acid residues (e.g. **Figure 2c**, and **Supplementary file 1** — 'Table S3' Tab and **Supplementary file 3**); for the remaining few cases (3%), the respective sites of loss events in *A. lycopersici* were coincident with the gain or loss of one or several amino acid residues (e.g. **Figure 2d**). Within this gene set, similar findings were apparent for the larger number of *A. lycopersici* intron losses as compared to intron sites conserved between the two closest relatives (*D. pteronyssinus* and *T. urticae*; **Supplementary file 3**). Despite striking examples of intronless genes arising from the loss of multiple conserved introns, as for *aculy03g01320* (**Figure 2c**), some *A. lycopersici* genes have both lost and retained arthropod conserved introns (i.e. *aculy02g00250*, *aculy03g02140*, and *aculy01g28080*, **Supplementary file 3**).
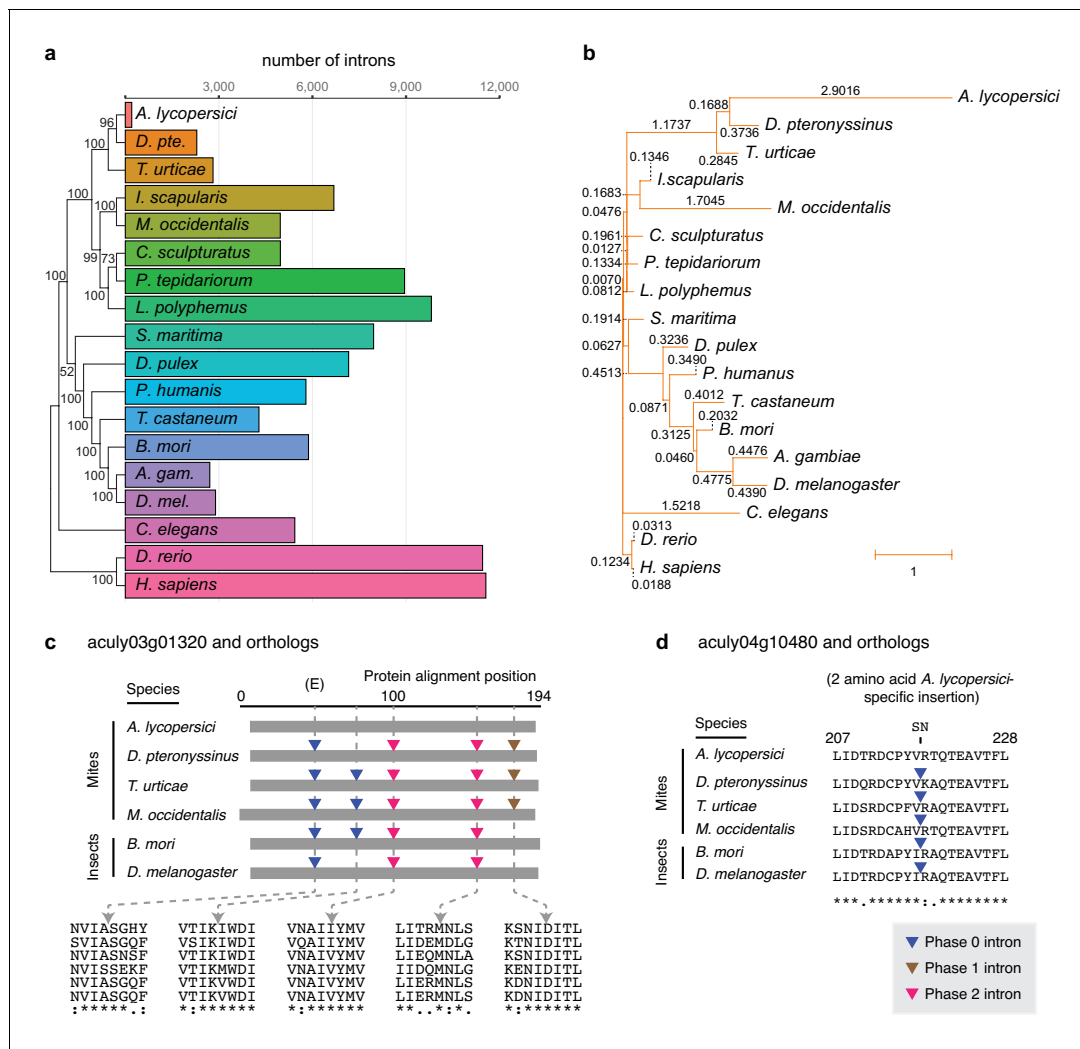
Figure 2. Number of conserved introns and intron loss rate across 18 metazoan species. (a) Phylogenetic tree built from 147 single copy orthologues (left; numbers at nodes indicate bootstrap support), and a histogram of introns present at 29,447 conserved positions identified by the software package Malin (right). (b) Phylogenetic tree with branch lengths labeled and scaled to the intron loss rate calculated by Malin. The unedited tree in both panels is given in *Figure 2—figure supplement 3*, and was, together with 2371 orthologous protein clusters (*Supplementary file 2*), used as input for Malin. (c) Alignment of *A. lycopersici* aculy03g01320 (which encodes an ADP-ribosylation factor-like 8, or Arl8, protein) with single copy orthologues from five other mite and insect species as indicated. Analogous positions of phase 0, 1, and 2 introns are denoted by colored triangles (legend, bottom right), with amino acids at the analogous intronic positions indicated beneath (identity, similarity, and non-similarity are indicated by '*', ':', and '.', respectively, for aculy03g01320 and its orthologue from *D. pteronyssinus,* the most closely related genome; in descending order, the sequence identifiers are aculy03g01320.1, g8154.t1, tetur10g00460, rna18006, BGIBMGA010943-RA, and FBtr0339723). The letter 'E' indicates that this intron position is conserved across other model organisms in Eukaryota; *Dictyostelium purpureum* (GenBank Accession XM_003283650), *C. elegans* (NM_070390.9), *H. sapiens* (NM_018184.3), *Monosiga brevicollis* (XM_001744342.1), and *A. thaliana* (NM_114847.5). (d) Local protein alignment, after panel c, revealing a candidate imprecise intron loss event in *aculy04g10480* (which encodes a polymerase delta-interacting protein) in *A. lycopersici* (insertion of S and N amino acid residues, top). Numbers denote positions in the *A. lycopersici* orthologue; sequence identifiers, in descending order, are aculy04g10480.1, g5664.t1, tetur01g12540, rna9399, BGIBMGA013121-RA, and FBtr0078681. Panels (c) and (d) are drawn based on Malin output. Other findings for intronic features and factors contributing to *A. lycopersici*'s genome reduction, and the supporting analyses, are presented in *Figure 2—figure supplements 1*, *2*, *4*, *5* and *6*.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Transposable element (TE) composition of the genome of *A. lycopersici* as well as that of four other animals.

Figure supplement 2. Gene and TE density along the major *A. lycopersici* genome scaffolds.

Figure supplement 3. Maximum likelihood phylogenetic analysis of 18 metazoan species including *A. lycopersici*.

Figure supplement 4. Intron gain rate across 18 metazoan species including *A. lycopersici*.

Figure supplement 5. Density plot of conserved intron positions identified by Malin in 18 metazoan species.

Figure supplement 6. Median length of all introns in 18 metazoan species.

Gene family contractions predominate in *A. lycopersici*

As revealed by the clustering algorithm implemented in the CAFE software (*Han et al., 2013*), *A. lycopersici* exhibits one of the highest rates of gene family contractions (1725), and by far the lowest rate of gene family expansions (206), among the 18 metazoans we analyzed (*Figure 3*; input data for the analysis are provided in *Supplementary file 4* and *Supplementary file 5*). It also has the lowest average expansion per gene family (*Supplementary file 1* — 'Table S4' Tab). Of the 105 gene families that were identified as 'rapidly evolving' in *A. lycopersici*, only four – as represented by orthogroups (OGs) OG0000007 (containing an Asteroid domain: IPR026832), OG0000546 (containing a Major Facilitator Superfamily, or MFS, domain: IPR011701), OG0000583 (containing a Troponin domain: IPR001978), and OG0002260 (hypothetical proteins) – were identified as expanding. The remaining 101 families were all identified as contracting (*Supplementary file 1* — 'Table S5' Tab). Six of these contracting families did contain more than 10 members in *A. lycopersici* (OG0000000, containing a Zinc finger C2H2-type domain: IPR013087; OG0000003, containing a Homeobox domain: IPR001356; OG0000005, containing a Serine protease, trypsin domain: IPR001254; OG0000014, containing a Cytochrome P450 domain: IPR001128; OG0000015, containing a G-protein-coupled receptor, rhodopsin-like domain: IPR000276; G0000025, containing a Homeobox domain: IPR001356) and, except for OG0000014 containing members of the P450 family, which is known to have only few orthologous relationships (*Feyereisen, 2011*), on average 72.2% of retained *A. lycopersici* genes had an orthologue in the majority of chelicerate species (*Supplementary file 1* — 'Table S6' Tab). Further, among the 101 rapidly contracted gene families we identified families previously implicated in mite and insect xenobiotic detoxification (*Dermauw et al., 2013a*; *Dermauw et al., 2013b*; *Snoeck et al., 2018*; *Van Leeuwen and Dermauw, 2016*) – carboxyl/choline esterases (CCEs: OG0000021 and OG0001201), cytochrome P450 monooxygenases (CYPs: OG0000014, OG0000030 and OG0000052), glutathione-S-transferases (GSTs: OG0000102, OG0000124), short-chain dehydrogenases/reductases (SDRs: OG0000096), ATP-binding cassette
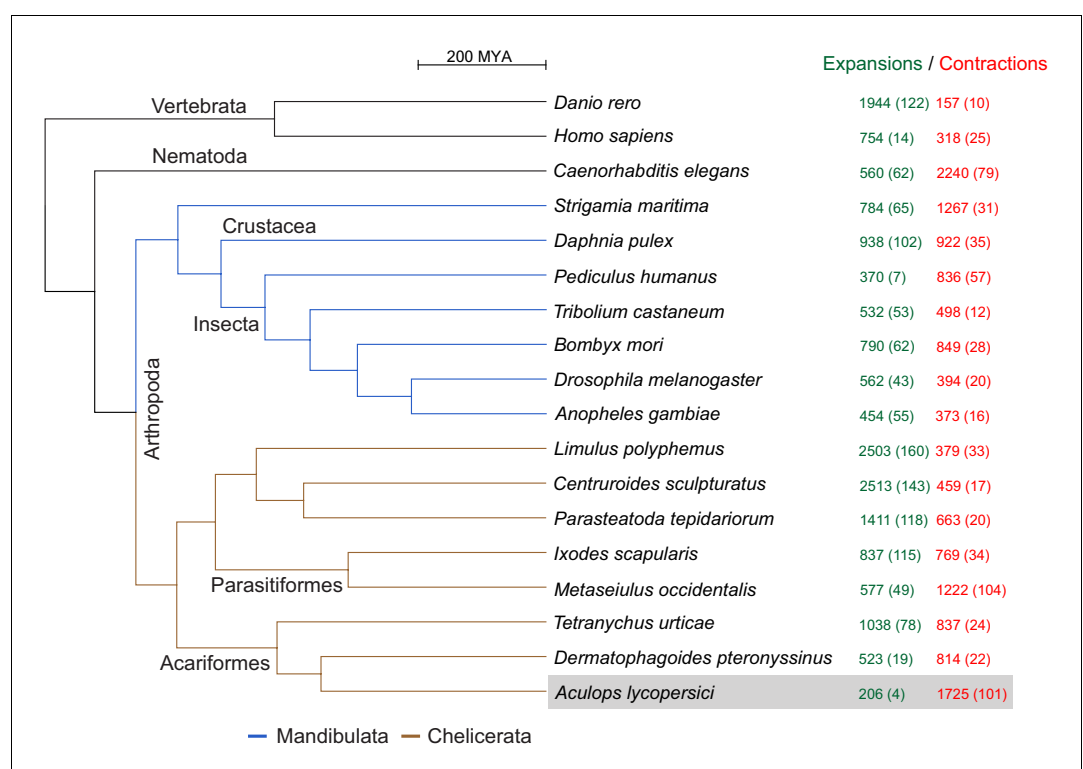

Figure 3. CAFE analysis of 6487 metazoan orthogroups. The number of expanding orthogroups are indicated in green font, while contracting orthogroups are indicated in red font. The number of rapidly expanding or contracting orthogroups (p-value<0.05) is shown in parentheses and details regarding these orthogroups can be found in *Supplementary file 1* — 'Table S5' Tab and 'Table S7' Tab.

(ABC) transporters (ABCs: OG0000051 and OG0000109) and MFS proteins (OG0000029, OG0000071, OG0000099, OG0000187) (*Supplementary file 1* — 'Table S5' Tab and 'Table S7' Tab). Given the role of these families in herbivory and host plant use (*Després et al., 2007*; *Heckel, 2014*; *Van Leeuwen and Dermauw, 2016*), we analyzed a selection of these gene families in detail (see the following sections).

We also found 315 orthogroups with no members in *A. lycopersici* but at least one member in all other arthropod species. This is the highest number of absent orthogroups of all arthropods included in our analysis, is ~2-fold more than those lacking in *D. pteronyssinus* (171), and more than threefold those absent in *T. urticae* (101) (*Supplementary file 1* — 'Table S8' Tab). A gene ontology (GO) enrichment analysis for *D. melanogaster* members within these conserved arthropod orthogroups without *A. lycopersici* members revealed that N-acetylglucosamine metabolic process (GO:0006044), transferase activity (GO:0016740) and Golgi apparatus (GO:000579) were the most highly significantly enriched GO terms within the Biological Process, Molecular Function and Cellular Component GO categories, respectively (*Supplementary file 1* — 'Table S9' Tab). Lastly, we found that 427 *D. melanogaster* essential genes (*Aromolaran et al., 2020*) coded for members of 390 orthogroups. Forty-eight of these essential orthogroups did not have members within the Acariformes, the mite superorder comprising *A. lycopersici*, *D. pteronyssinus,* and *T. urticae*, while 21 (5.4%) orthogroups were absent in *A. lycopersici* but present in other acariform mites (*Supplementary file 1* — 'Table S10' Tab).

Furthermore, in a number of cases, orthogroups absent in *A. lycopersici* harbor conserved genes with potential roles in the development of tissues or structures that are absent or modified in the russet mite relative to other chelicerates or insects (see also Discussion, and Results section, 'Loss of highly conserved transcription factors'). For instance, orthologues of *Drosophila* unkempt, a known developmental regulator, and *Drosophila* dachs, essential for appendage growth, are both absent in *A. lycopersici* but present in all other arthropods (OG0002898 and OG0006002, respectively). Dachs is known to interact with four-jointed (*Buckles et al., 2001*), which is also absent in *A. lycopersici*, even though it is present in all insect and chelicerate species included in our analysis (OG0003305). Finally, *fat* belongs, together with *dachs* and *four-jointed*, to the Fat/Hippo pathway and plays a key-role in tissue proliferation and development in both invertebrates and vertebrates (*Simon et al., 2010*). Although *dachsous*, another player in this pathway, is present (aculy04g02000 in OG0001018), a *fat* orthologue could not be identified in *A. lycopersici* while this orthologue was found in other acariform mites (OG0000383, *Supplementary file 1* — 'Table S7' Tab).

Detoxification genes

We curated the *A. lycopersici* genome for sequences encoding established detoxification enzymes (*Després et al., 2007*; *Heckel, 2014*; *Van Leeuwen and Dermauw, 2016*) including GSTs, CCEs, and CYPs. In *A. lycopersici*, detoxification gene families are especially reduced, with a mere 4 GSTs, 8 CCEs, and only 23 CYPs (*Table 2*, *Figure 4a*, and *Figure 4—figure supplements 1*, *2* and *3*; *Van Leeuwen and Dermauw, 2016*). In particular, the number of GSTs and CCEs is remarkably low (see Discussion). This finding was corroborated by mining of the *A. lycopersici* transcriptome assembly (the 4 GSTs and 8 CCEs present in the genome assembly were also present in transcriptome assembly, with no other transcript contigs with homology to GSTs or CCEs identified). Of note, half of the GSTs and almost all (7 out of 8) CCE genes in *A. lycopersici* are evolutionarily conserved across chelicerates or arthropods (*Figure 4—figure supplements 1* and *2*). We also examined transporters of the ABC family and MFS proteins that have been implicated in detoxification responses in arthropod species, although transporters in both of these families have diverse other roles as well (*de la Paz Celorio-Mancera et al., 2013*; *Dermauw et al., 2013a*; *Dermauw et al., 2013b*; *Dermauw and Van Leeuwen, 2014*; *Govind et al., 2010*). In contrast to genes encoding 'classic' detoxification enzymes like CYPs, CCEs, or GSTs, dramatic reductions in ABC transporter genes were not observed. For example, *A. lycopersici* has 9 ABCC and 16 ABCG transporters, while 22 and 2 are present in *M. occidentalis* and 39 and 23 are present in *T. urticae*, respectively (*Table 2*, *Figure 4—figure supplement 4*). Further, in contrast to the trend for contractions of the classic detoxification gene families, we also observed two *A. lycopersici* expansions - comprising three orthogroups, OG0000024, OG0000546, and OG0006109 - of the MFS, which is involved in membrane-based transport of small molecules (*Figure 4b*, *Figure 4—figure supplement 5*; *Pao et al., 1998*; *Yan, 2015*).

Table 2. Detoxification enzyme (CYPs, GSTs, CCEs) and ABC transporter gene family size in *A. lycopersici*, *T. urticae*, *M. occidentalis*, and *D. melanogaster*.

Detoxification enzyme	A. lycopersici	T. urticae	M. occidentalis	D. melanogaster
CYPs (total)	23	78*	63	86
CYP2	1	38	16	7
CYP3	17	9	23	36
CYP4	2	26	19	32
Mito Clan	3	5	5	11
GSTs (total)	4	31	13	37
Delta/Epsilon	1	16	3	25
Mu	2	12	5	0
Omega	0	2	3	5
Sigma	0	0	0	1
Theta	0	0	0	4
Zeta	1	1	1	2
Unknown	0	0	1	0
CCEs (total)	8	69	44	35
Dietary class (clade A-C)	0	0	0	13
Hormone class				
D (integument CCEs)	0	0	0	3
E (secreted beta-esterases)	0	0	0	2
F (dipteran JHEs[†])	0	0	0	3
F' (chelicerate JHEs)	0	2	1	0
Neurodevelopmental class				
H (glutactins)	0	0	0	4
J (AChE)	1	1	1	1
J' (Acari-specific CCEs)	0	32	19	0
J'' (Acari-specific CCEs)	0	22	15	0
K (gliotactin)	1	1	1	1
L (neuroligins)	2	5	5	4
M (neurotactin)	1	1	0	1
U (unchar. conserv. clade in Acariformes/*L. polyphemus*)	2	3	0	0
I (unchar. conserv. clade in insects)	0	0	0	2
No clear clade assignment	1	2	2	1
ABCs (total)	44	103	55	56
ABCA	4	9	8	10
ABCB-FT[‡]	3	2	1	4
ABCB-HT[§]	1	2	4	4
ABCC	9	39	22	14
ABCD	2	2	4	2
ABCE	1	1	1	1
ABCF	3	3	3	3
ABCG	16	23	2	15
ABCH	5	22	6	3
Unknown	0	0	4	0
Total	79	281	175	214

Numbers and class/clade/subfamily assignments were derived from previous studies (**Grbić et al., 2011**; **Wei et al., 2020**; **Wu and Hoy, 2016**) and this study.

*Of the 81 *T. urticae* CYPs identified by **Grbić et al., 2011**, three CYP genes (*tetur46g00150, tetur46g00170 and tetur47g00090*) and *tetur602g00010* were considered as allelic variants and a pseudogene, respectively, and one new full-length CYP gene (*tetur01g13730*) was identified in this study.

†JHE, juvenile hormone esterases.

‡FT, full transporter.

§HT, half transporter.

Chemosensory and related receptors

To see if *A. lycopersici*'s specialized lifestyle has had a notable impact on chemoreception, we also exhaustively mined and annotated the *A. lycopersici* genome for gustatory receptors (GRs), degenerin/epithelial Na+ channels (ENaCs), ionotropic receptors (IR) and transient receptor potential (TRP) channels. Members of these four families have been previously documented to play important roles in sensing environmental (chemical) cues in other arthropod species (**Damann et al., 2008**; **Hoy et al., 2016**; **Ngoc et al., 2016**; **Robertson et al., 2003**; **Rytz et al., 2013**; **Whiteman and Pierce, 2008**). The GR family, which contains seven transmembrane spanning regions (**Touhara and Vosshall, 2009**) and is linked to the detection of sweet and bitter compounds (**Silbering and Benton, 2010**), was the most strongly reduced, with only two of these genes identified (**Figure 4c**, **Figure 4—figure supplement 6**), as opposed to the 447 intact GRs reported in *T. urticae* (**Ngoc et al., 2016**). Further, only four ENaCs are present in the *A. lycopersici* genome (**Figure 4d**, **Figure 4—figure supplement 7**). Members of this family have recently been shown or suggested to be chemoreceptors for diverse compounds in insects and mites, but some family members likely have highly conserved roles in acid sensing (**Ben-Shahar, 2011**; **Silbering and Benton, 2010**), as well as in the perception of mechanical or osmotic cues (**Ben-Shahar, 2011**; **Zelle et al., 2013**). Of the two ENaCs likely to play these conserved roles in *T. urticae,* one is in a well-supported clade with a single ENaC in the tomato russet mite (aculy04g09940) (**Figure 4** , **Figure 4—figure supplement 7**).

The IR family, which has been linked to odorant detection (**Joseph and Carlson, 2015**), humidity and temperature sensing in *D. melanogaster* (**Enjin et al., 2016**), is markedly reduced in *A. lycopersici* compared to most insects and *M. occidentalis* (**Hoy et al., 2016**). However, the numbers are similar to those in *T. urticae* (each has four putative IRs with strong bootstrap support), including homologues of the highly conserved IR25a and IR93a receptors (**Figure 4—figure supplement 8**). Interestingly, *A. lycopersici* may have as few as six ionotropic glutamate receptors (iGluRs), as compared to 14 in *T. urticae* (**Figure 4—figure supplement 8**); proteins in this family are related to IRs, but have ultra-conserved roles in synaptic transmission in animals (**Benton et al., 2009**).

Finally, we found both expansions and contractions of the TRP family (**Figure 4—figure supplement 9**). Like the other sequenced herbivorous mite, *T. urticae*, no orthologue of TRPA1 was located, but orthologues for TRPgamma, NopmC, and TRPML are present, with three copies of NopmC as compared to *T. urticae*'s two. Unlike *T. urticae*, members of the TRPP and TRPM clades were completely absent in the russet mite, but strikingly, two putative members of the TRPV clade (Inactive and Nanchung), previously thought to be lost in mites and ticks (**Peng et al., 2015**; **Regier et al., 2010**), appear to be present.

Loss of highly conserved transcription factors

Among two vertebrates, one nematode and the 15 arthropod species we analyzed, *A. lycopersici* has the lowest number (364) of transcription factor (TF) genes (**Supplementary file 1** — 'Table S11' Tab). Nevertheless, when accounting for the total number of genes by species, the TF fraction in *A. lycopersici* (3.55%) is higher than that of *T. urticae* (2.98%), and is within the range reported for metazoan animals (4.7% ±1.4, **Charoensawan et al., 2010**). However, a lower number of the PFAM TF domains Zinc finger (zf-C2H2 and zf-CCHC), Forkhead, Homeobox, Hormone (nuclear) receptor, HLH, bZIP_2 and T-box were found in *A. lycopersici* compared to all other species included in our analysis (**Supplementary file 1** — 'Table S11' Tab). In addition, *A. lycopersici* orthologues of the Hairy Orange protein family (hey, cwo and deadpan) have lost the Hairy Orange domain (**Figure 4—figure supplement 10**), while an orthologue of *D. melanogaster* SoxNeuro could not be identified in *A. lycopersici* despite being present in the spider and Acari genomes examined (**Figure 4—figure**
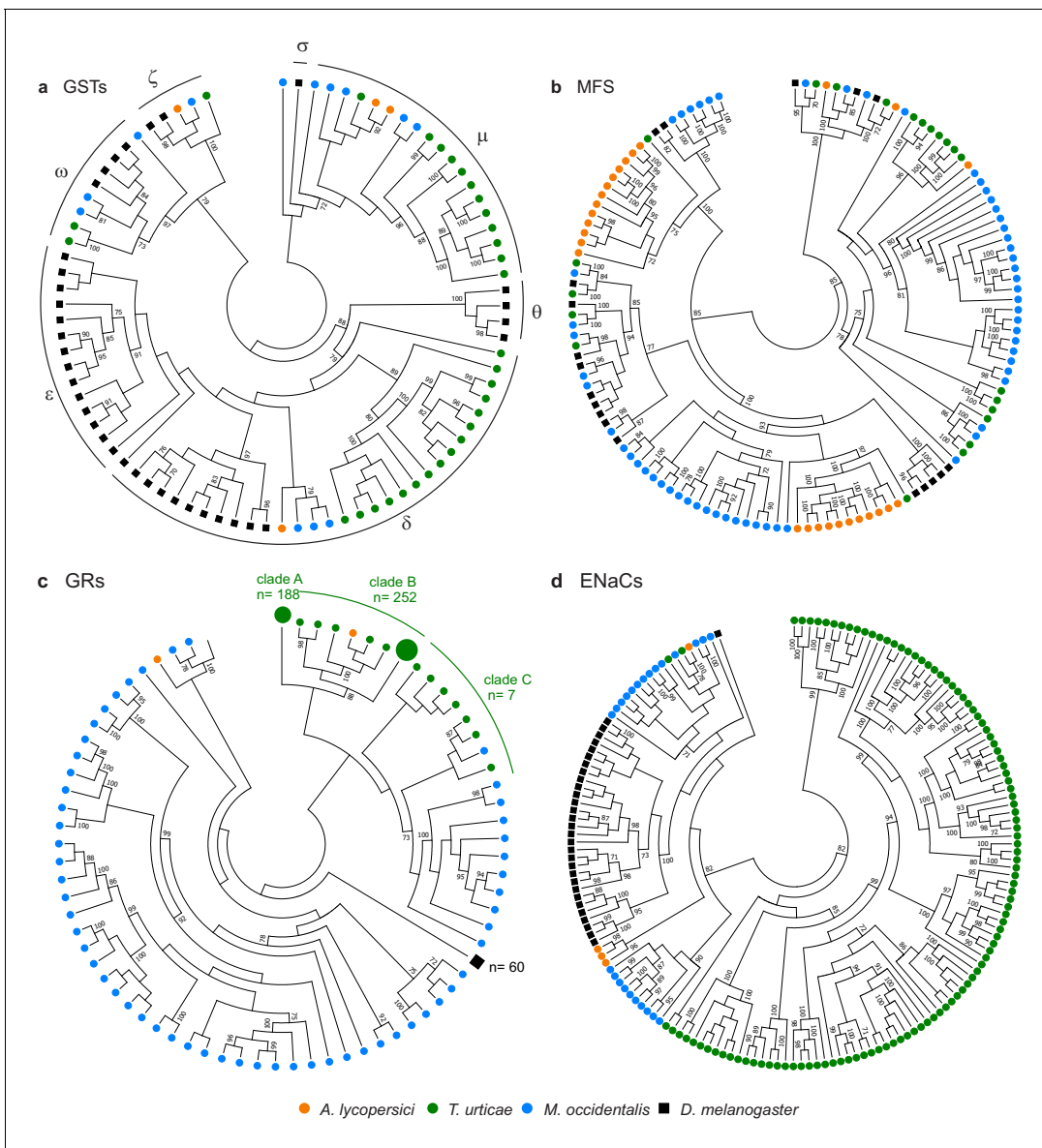
Figure 4. Gene family contractions and mini-expansions in *A. lycopersici*. Maximum likelihood phylogenetic analysis of selected detoxification and chemosensory families among *A. lycopersici*, *T. urticae*, *M. occidentalis* and *D. melanogaster*. (**a**) Glutathione-S-transferases (GSTs); the different GST classes (zeta, theta, delta, epsilon, omega, mu, sigma) are indicated with arches. (**b**) Major facilitator superfamily (MFS). (**c**) Gustatory receptors (GRs). (**d**) Epithelial Na+ Channels (ENaCs). All trees are midpoint rooted and only topology is shown. Gustatory receptors for *D. melanogaster* as well as the species-specific class A and B expansions identified in *T. urticae* are collapsed for clarity. Only bootstrap values above 70 are shown. Phylogenetic reconstructions for gene families, or analyses of domain losses in *A. lycopersici* in arthropod conserved genes, are given in *Figure 4—figure supplements 1–20*. For panels a-d, the detailed versions for each tree, including sequence identifiers, can be found in *Figure 4—figure supplements 1*, *5*, *6* and *7*, respectively. The alignments used for phylogenetic inference can be found in *Supplementary file 7*.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Phylogenetic analysis of GST protein sequences of *A. lycopersici*.

Figure supplement 2. Phylogenetic analysis of CCE protein sequences of *A. lycopersici*.

Figure supplement 3. Phylogenetic analysis of CYP protein sequences of *A. lycopersici*.

Figure supplement 4. Phylogenetic analysis of nucleotide-binding domains of ABC proteins of *A. lycopersici*.

Figure supplement 5. Phylogenetic analysis of MFS protein sequences of *A. lycopersici*.

Figure supplement 6. Phylogenetic analysis of GRs of *A. lycopersici*.

Figure supplement 7. Phylogenetic analysis of ENaCs of *A. lycopersici*.

Figure supplement 8. Phylogenetic analysis of ionotropic and related receptors of *A. lycopersici*.

Figure supplement 9. Phylogenetic analysis of TRP channels of *A. lycopersici*.

Figure 4 continued on next page

Figure 4 continued

Figure supplement 10. Alignment of the Hairy Orange domain region from deadpan, hey and cwo proteins of *A. lycopersici*, *D. pteronyssinus*, and *T. urticae* with deadpan, hey and cwo of *D. melanogaster*.

Figure supplement 11. Bayesian phylogenetic analysis of *A. lycopersici* Sox proteins.

Figure supplement 12. Alignment of the DNA-binding domain and the ligand-binding domain region of *D. melanogaster* E75 with those of *T. urticae*, *D. pteronyssinus* and *A. lycopersici*.

Figure supplement 13. Alignment of the DNA-binding domain and the ligand-binding domain region of *D. melanogaster* HR4 with those of *T. urticae* and *A. lycopersici*.

Figure supplement 14. Alignment of the DNA-binding domain and the ligand-binding domain region of *D. melanogaster* HR38 with those of *T. urticae*, *D. pteronyssinus* and *A. lycopersici*.

Figure supplement 15. Alignment of the DNA-binding domain and the ligand-binding domain region of *D. melanogaster* HR51 with those of *T. urticae*, *D. pteronyssinus*, and *A. lycopersici*.

Figure supplement 16. Alignment of the DNA-binding domain and the ligand-binding domain region of *D. melanogaster* SVP with those of *T. urticae*, *D. pteronyssinus*, and *A. lycopersici*.

Figure supplement 17. Alignment of the DNA-binding domain and the ligand-binding domain region of *D. melanogaster* DSF with those of *T. urticae*, *D. pteronyssinus*, and *A. lycopersici*.

Figure supplement 18. Phylogenetic analysis of *A. lycopersici* protein sequences with a T-box (PF00907) domain.

Figure supplement 19. Phylogenetic analysis of *A. lycopersici* UGT protein sequences.

Figure supplement 20. Phylogenetic analysis of *A. lycopersici* C1A proteases.

supplement 11). Among nuclear receptors (NRs), we identified eight canonical NRs in the *A. lycopersici* genome (E78, HR3, EcR, two RXRs, ERR, FTZ-F1, HR96) that contained both a DNA-binding domain (DBD) and a ligand-binding domain (LBD). However, no homologues of the evolutionary conserved NRs HNF4, HR39, HR78, and HR83 (*Bodofsky et al., 2017*; *Bonneton and Laudet, 2012*), nor a homologue of the *T. urticae* Photoreceptor-specific NR (PNR), were detected in the *A. lycopersici* genome, even though HR78, HNF4, and PNR are present in *D. pteronyssinus* (*Supplementary file 1* — 'Table S12.1' Tab and 'Table S12.2' Tab). Further, for six nuclear receptors (E75, DSF, HR4, HR38, HR51, and SVP) that are evolutionary conserved across arthropods and normally have a canonical (DBD+LBD) structure (*Fahrbach et al., 2012*; *Grbić et al., 2011*; *Hwang et al., 2014*; *Litoff et al., 2014*), an LBD was not predicted for the respective *A. lycopersici* homologues. LBDs for all of these except HR4 were predicted for both the *D. pteronyssinus* and *T. urticae* homologues (*Supplementary file 1* — 'Table S12.1' Tab and 'Table S12.2' Tab, *Figure 4— figure supplements 12–17*).

The basic helix-loop-helix (bHLH) gene family is an ancient family found in fungi, plants, and animals, and members of this family are essential both for organisms to respond to environmental factors, as well as for cellular differentiation during development (*Skinner et al., 2010*). The *D. melanogaster achaete* and *scute* bHLH genes play crucial roles in bristle development (*García-Bellido and de Celis, 2009*). Within the bHLH family group we found that *T. urticae*, *M. occidentalis* and *I. scapularis* have five bHLH proteins with an achaete-scute InterPro domain (IPR015660), while only three were found in both *D. pteronyssinus* (g4111.t1, g7028.t1 and g6164.t1) and *A. lycopersici* (aculy01g18470, aculy01g18540 and aculy02g28230).

A number of other specific transcription factors that are highly conserved among most arthropods are also absent from the *A. lycopersici* genome. For *A. lycopersici*, we were unable to identify *proboscipedia*, a member of the Hox gene family. Members of this family (*labial*, *proboscipedia*, *Hox3/zen*, *Deformed*, *Sex combs reduced*, *fushi tarazu*, *Antennapedia*, *Ultrabithorax*, *abdominal-A*, and *Abdominal-B*) encode homeodomain transcription factors and act to determine the identity of segments along the anterior–posterior axis in arthropods (*Hughes and Kaufman, 2002*). *Proboscipedia* is present in all chelicerate genomes (horseshoe crab, scorpions, spiders, mites and ticks) for which Hox genes have been analyzed (*Figure 5*, *Supplementary file 1* — 'Table S13.1' Tab and 'Table S13.2' Tab, *Supplementary file 6*; *Di et al., 2015*; *Hoy et al., 2016*; *Kenny et al., 2016*; *Schwager et al., 2017*), and is believed to be ancestral to all arthropods (*Pace et al., 2016*). Of particular note, *proboscipedia* is located in close proximity (<35 kb) of *labial* in Acariformes, but in *Aculops labial* was the only Hox gene that was present on scaffold 2 (*Supplementary file 1* — 'Table S14' Tab). Furthermore, *A. lycopersici* lacks a homologue of the T-box encoding gene *org-1* (*Figure 4—figure supplement 18*), which in *D. melanogaster* plays a pivotal role in diversification of circular visceral muscle (*Schaub and Frasch, 2013*). Finally, we also mined the *A. lycopersici* genome
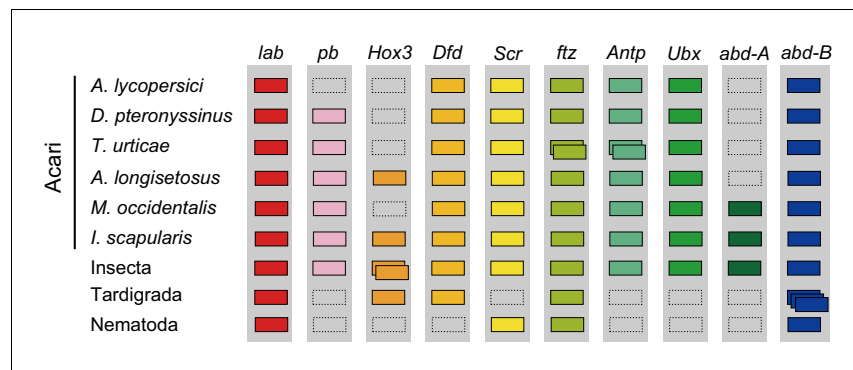
Figure 5. Hox genes in Acari and other ecdysozoan lineages. Hox orthology groups are indicated by different colored boxes. Gray boxes with a dashed outline represent missing Hox genes. Some species have duplications of Hox genes and these are indicated by multiple boxes that overlap. *T. castaneum*, *H. dujardini* and *C. elegans* were selected as representative species for the Hox gene clusters of Insecta, Tardigrada and Nematoda, respectively.

for transcription factors and other genes involved in circadian rhythm (so-called 'clock' genes) (*Supplementary file 1* — 'Table S15' Tab). Orthologues of the helix-loop-helix TFs *cycle*, *Clock* and *tango* and the bZIP TF *vrille* were identified in the *A. lycopersici* genome. However, we did not identify *period* and *timeless*, known negative regulators of *Clock* and *cycle* (*Lee et al., 1999*; *Peschel and Helfrich-Förster, 2011*). Other circadian regulators, like the circadian photoreceptor *cryptochrome* and the bZIP TF *PAR-domain protein 1ε*, were also not identified, even though these are present in *T. urticae* (*Hoy et al., 2016*).

Horizontally transferred genes

We identified 18 putatively intact horizontal gene transfer (HGT) candidate genes (*Supplementary file 1* — 'Table S16' Tab), and performed subsequent phylogenetic analyses that suggested that nine were acquired from a foreign source. Seven of these genes code for UDP-glyco-syltransferases (UGTs), members of which have well documented roles in xenobiotic detoxification (*Snoeck et al., 2019*). Phylogenetic inference with all *T. urticae*, *D. pteronyssinus* and *A. lycopersici* UGTs (80, 27, and 7, respectively) indicated that the seven UGTs in the tomato russet mite genome were the result of a lineage-specific expansion (*Figure 4—figure supplement 19*). Although we did not observe a clear phylogenetic signature of HGT (*Wybouw et al., 2016*), our phylogenetic reconstruction is consistent with previous studies which indicated that, prior to the formation of the Acariformes lineage, an ancestral mite species laterally acquired a *UGT* gene copy from a bacterial source (*Ahn et al., 2014*)(*Wybouw et al., 2018*).

Two intact genes of bacterial origin (*aculy01g38350* and *aculy04g02470*) were also identified in the tomato russet mite genome that are predicted to code for enzymes in the microbial and plant pantothenate biosynthesis pathway (an apparent duplicate of *aculy01g38350* was also uncovered, but the coding sequence was disrupted, and it lacked expression, suggesting it is a pseudogene) (*Figure 6*). PCR amplification linked both laterally acquired genes with either neighboring intron-containing genes (*aculy01g38350*) or conserved eukaryotic genes (*aculy04g02470* is located next to *aculy04g02480*, which encodes a Gtr1/RagA protein); in addition, an *aculy01g38350* transcript (Illumina contig 1934) had a polyA tail, suggestive of eukaryotic transcription (*Figure 6—figure supplement 1*). Pantothenate, or vitamin B5, is a life-essential compound, and whereas plants and bacteria are able to synthesize this compound de novo, animals rely on dietary uptake. Genes for pantothenate synthesis are present in tetranychid mites, and genomic and phylogenetic approaches have pointed to an ancient HGT event prior to speciation within the Tetranychidae family for both genes. Constrained tree tests rejected the topology where ketopantoate hydroxymethyltransferase of *A. lycopersici* was the sister lineage to the group of spider mite biosynthetic proteins, but not for pantoate β-alanine ligase, suggesting that *A. lycopersici* acquired the ketopantoate hydroxymethyltransferase gene from a different bacterial donor species (*Figure 6*, Approximately Unbiased tests, p-value cut-off of 0.01).
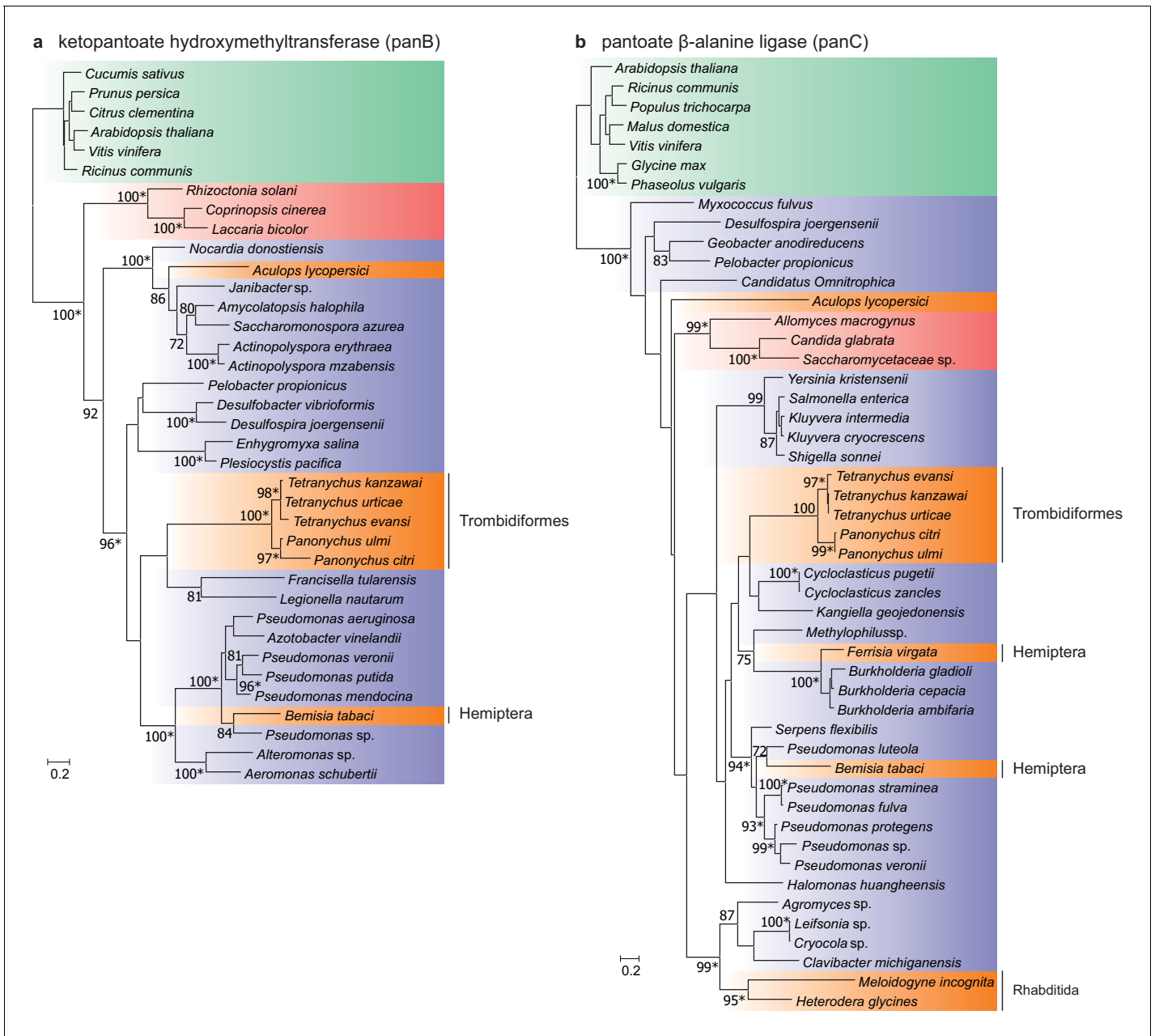
Figure 6. Maximum-likelihood phylogenetic inference for ketopantoate hydroxymethyltransferase and pantoate β-alanine ligase of *A. lycopersici*. (**a**) Ketopantoate hydroxymethyltransferase. (**b**) Pantoate β-alanine ligase. Branches are color coded depending on their position within the tree of life; plants: green, animals: orange, fungi: red and bacteria: blue. RAxML phylogenetic reconstructions are consistent with the evolutionary scenario of independent horizontal transfer events of the two pantothenate biosynthetic genes in the *A. lycopersici* lineage, tetranychid spider mites, and hemipterans. Only RAxML bootstrap support values higher than 70 are depicted and the scale bars represent 0.2 amino acid substitutions per site. Informative nodes were identical and well-supported in another maximum-likelihood analysis (IQ-TREE; an asterisk indicates nodes with ultrafast bootstrap values above or equal to 95 in the IQ-TREE analyses). Plant homologues were used to root both phylogenetic trees. The alignments used for phylogenetic inference can be found in *Supplementary file 7*.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Integration of the *ketopantoate hydroxymethyltransferase* (*aculy01g38350*, *panB*) and *pantoate β-alanine ligase* (*aculy04g02470*, *panC*) genes into the *A. lycopersici* genome.

In *T. urticae* the acquisition of pantothenate biosynthetic genes is accompanied by the horizontal gene transfer of two methylenetetrahydrofolate dehydrogenases (MTHFDs), enzymes of the folate pathway and connected to the pantothenate biosynthesis pathway (*Wybouw et al., 2018*). Although such a HGT was not detected in *A. lycopersici*, an expansion of MTHFDs was detected compared to other mite species (OG0000706 in *Supplementary file 1* — 'Table S7' Tab).

Secreted proteins

Small molecules or proteins produced in salivary glands are one mechanism by which arthropod herbivores can manipulate the defenses of their host plants. As *A. lycopersici* is able to potently suppress tomato defenses (*Glas et al., 2014*; *Schimmel et al., 2018*), we predicted its secretome, and found that 612 of the 10,263 annotated *A. lycopersici* proteins (6%) are putatively secreted (*Supplementary file 1* — 'Table S17' Tab). Only one of the more than 600 secreted *A. lycopersici* proteins (aculy02g17370, a glycosyl hydrolase, family 13, IPR013780) had a best BLASTp hit with a *T. urticae* protein that was previously identified in *T. urticae* saliva using an LC-MS/MS (*Jonckheere et al., 2016*). More than half (351) of these proteins were absent in orthogroups in non-herbivorous arthropod species, and are less than 350 amino acids in length. Only 15 of these 351 proteins belonged to an orthogroup with more than one member in *A. lycopersici* (OG0006384, OG0009325, OG0009954 and OG0010904). Among these, OG0009325 contains three short *A. lycopersici* proteins <90 amino acids in length (aculy01g11450, aculy01g12600, and aculy01g12690). Of note, the gene encoding the single *T. urticae* representative in this group, *tetur24g01070*, was previously found to be overexpressed in the *T. urticae* salivary gland region (*Jonckheere et al., 2016*). OG0006384, on the other hand, contains cysteine peptidases (Peptidase C1A, papain C-terminal domain; InterPro IPR000668), which are enzymes reported to have key roles in plant-pathogen/pest interactions (*Shindo and Van der Hoorn, 2008*), and for which two lineage-specific expansions are present in *A. lycopersici* (*Figure 4—figure supplement 20*).

Small RNA pathways

We also characterized components of small RNA pathways that might be of potential relevance for agricultural control methods. The *A. lycopersici* genome harbors highly conserved miRNA sequences, such as *let-7*, *miR-1*, and *miR-9a* (*Supplementary file 1* — 'Table S18' Tab). However, in contrast to *T. urticae*, a clear *A. lycopersici* homologue of Exportin-5, a dsRNA-binding protein mediating nuclear transport of pre-miRNAs (*Bohnsack et al., 2004*; *Kim, 2005*), is lacking, suggesting a deviating miRNA pathway in *A. lycopersici*. In line with the latter hypothesis, we could not identify an *A. lycopersici* homologue of *Staufen*, while this gene is present in *T. urticae* (*Grbić et al., 2011*; *Supplementary file 1* — 'Table S19' Tab) and other arachnids (OrthoDb v 9.1, group EOG091G07A0 and EOG090Z04UZ, respectively) and was shown to negatively modulate miRNA activity in the nematode *C. elegans* (*Ren et al., 2016*).

The *A. lycopersici* genome contains, in line with *T. urticae*, clear homologues of Dicer-1, Loquacious, Drosha and Pasha and an expansion of the AGO1 and PIWI/AGO3 subfamilies. Of note, we found one *A. lycopersici* protein (aculy02g00240) that was highly homologous to the *T. castaneum* Dicer-1 enzyme (bitscore of 294) and that contained both an RNA-binding domain (PAZ-domain, cl00301) and the RNAse III domain (cd00593) while two *A. lycopersici* proteins (aculy02g04810 and aculy02g19970) showed reciprocal BLASTp hits with *T. castaneum* Dicer-2 and Dicer-1, respectively, but were relatively short (about 500 amino acids (aa) compared to 1726 aa for aculy02g00240) and only contained the RNAse III domain. However, the genes encoding these proteins are located next to a sequencing gap in the current assembly and it could be that gene-models for these Dicer-like enzymes are not complete. Similar to *T. urticae*, we could not identify clear homologues of R2D2 and AGO2 (*Grbić et al., 2011*; *Supplementary file 1* — 'Table S19' Tab), suggesting that the siRNA pathway is either absent or non-canonical in both mite species (*Okamura et al., 2011*).

Further, important players in the PIWI-interacting RNA (piRNA) pathway (*Iwasaki et al., 2015*) were identified in the *A. lycopersici* genome (PIWI/AGO3, Zucchini, Armitage, Maelstrom and SoYb; *Supplementary file 1* — 'Table S19' Tab), while homologues of Armitage and Zucchini could not be identified in *T. urticae*, which is in line with the recently suggested non-canonical piRNA pathway in *T. urticae* (*Supplementary file 1* — 'Table S19' Tab, *Huang et al., 2014*; *Mondal et al., 2018b*).

Finally, RNA-dependent polymerases are known to be essential for the amplification of the RNA silencing effect (systemic RNAi) in *C. elegans* and some plants (*Tomoyasu et al., 2008*). Genes encoding these enzymes are absent in insect genomes while 1 to 5 have been reported in Acari genomes (*Grbić et al., 2011*; *Hoy et al., 2016*; *Joga et al., 2016*; *Mondal et al., 2018a*; *Zong et al., 2009*). Surprisingly, we could not identify RNA-dependent polymerase genes in the *A. lycopersici* genome (*Supplementary file 1* — 'Table S19' Tab), which might indicate that these genes have been lost since the divergence of Eriophyoidea from other acariform lineages. However, as systemic RNAi does seem to occur in some insect orders, for example, Coleoptera (*Joga et al., 2016*), we cannot exclude that systemic RNAi might also occur in *A. lycopersici*.

Discussion

Genome size varies enormously within the Acari. While tick genomes can be larger than 2 Gb (*Gulia-Nuss et al., 2016*), those of mite species belonging to the Acariformes are small (*Gregory and Young, 2020*). This is especially true for mites within the order Sarcoptiformes, including dust mites and scabies mites, for which genomes of lengths ~55-60 Mb have been reported (*Chan et al., 2015*; *Rider et al., 2015*). Eriophyoid mites like *A. lycopersici* have traditionally been placed in the order of the Trombidiformes, but recent work suggests they belong to the Sarcoptiformes, or a sister taxon (*Arribas et al., 2020*; *Bolton et al., 2017*; *Klimov et al., 2018*; *Xue et al., 2017*). Our work supports this conjecture, as within Acariformes, *A. lycopersici* fell in a well-supported clade with the house dust mite *D. pteronyssinus* (Sarcoptiformes), as opposed to *T. urticae* (Trombidiformes) (*Figure 2—figure supplement 3*).

Mirroring that of its closest sequenced relatives, the genome of *A. lycopersici* is tiny. At 32.5 Mb, it is the smallest reported to date for an arthropod and among the smallest metazoan genomes sequenced so far (*Slyusarev et al., 2020*). Its size is also consistent with cytological data that erio-phyoid mites have few chromosomes that are extremely small (*Helle and Wysoki, 1996*; *Helle and Wysoki, 1983*) and with several trends. In broad terms eukaryotic genome sizes correlate positively with larger cell (nuclei) sizes, and vary inversely with cell division times (*Elliott and Gregory, 2015*; and references therein). While little is known about the minimal cell sizes for *A. lycopersici*, the whole mite is smaller than many single eukaryotic cells and neuron somata sizes of less than 1 μm have been observed for another eriophyoid mite of similar size (*Whitmoyer et al., 1972*). *A. lycopersici* is also half the size (or less) of mites like *D. pteronyssinus*, and its minute physical stature and genome size are consistent with a recent analysis that revealed a positive correlation within Acari between organismal size and haploid DNA content (*Gregory and Young, 2020*). The *A. lycopersici* generation time, a potential (albeit imperfect) proxy for cell cycle progression, is also near the minimum reported for other mites, or for microinsects (*Danks, 2006*; *Kawai and Haque, 2004*; *Rice and Strong, 1962*). The force(s) that have led to the small physical and genome size of *A. lycopersici* are not known. However, russet mites can use their short stylets only to feed on plant epidermal cells (*Royalty and Perring, 1988*). This is in contrast to many other (larger) herbivores, including other herbivorous mites like *T. urticae* (*Bensoussan et al., 2016*), that can reach and consume the photosynthetically active, sugar-rich mesophyll cells (*Borsuk and Brodersen, 2019*; *Koroleva et al., 2000*) underneath the epidermis. The nutrient-poor diet of *A. lycopersici* may favor small physical size, and under some conditions, nutrient limitations have been proposed to select specifically for low DNA content (*Hessen et al., 2010a*). Regardless, the rapid generation time of *A. lycopersici* facilitates dense populations on its host (*Figure 1c,d*), and outcrossing by deposition of spermato-phores (*Al-Azzazy and Alhewairini, 2018*) in the environment may approximate panmixia, and hence high effective population sizes, and therefore more efficient selection against the accumula-tion of non-coding sequences associated with large eukaryotic genomes (*Lynch et al., 2011*). There-fore, a collection of life history features may underlie the streamlining observed in the *A. lycopersici* genome.

In addition to a very low content of repetitive sequences, a derived genomic organization under-pins the reduced *A. lycopersici* genome. As compared to the ~3 fold larger *T. urticae* genome (*Grbić et al., 2011*), the relative intergenic and intronic fractions are reduced, while compared to the ~2 fold larger *D. pteronyssinus* genome (*Waldron et al., 2017*), the intergenic fraction is nearly identical, while the genomic percent in introns is less. The latter reduction reflects massive intron loss in *A. lycopersici*, as 83.7% of genes were intronless, a value more than threefold higher than for

T. urticae or *D. pteronyssinus.* As observed in other intron-poor species (*Mourier and Jeffares, 2003*), we observed greater retention of 5′ introns in *A. lycopersici*, potentially a consequence of intron loss via 3′-biased intron removal by recombination with cDNAs following reverse transcription of spliced transcripts (also known as Reverse Transcriptase-Mediated Intron Loss, or RTMIL) (*Mourier and Jeffares, 2003*; *Roy and Gilbert, 2005*). Alternatively, or in concert, the pattern may reflect retention of 5′ introns rich in cis regulatory sequences (*Roy and Gilbert, 2005*), an explanation consistent with *A. lycopersici*'s relatively long median intron lengths as compared to other insects and mites with compact genomes (*Table 1*, *Figure 2—figure supplement 6*). Previously, comparisons of intron loss events among close relatives, where few mutational steps have occurred, have been important in establishing plausible mechanisms of intron loss (*Yenerall et al., 2011*; *Zhu and Niu, 2013*). Such analyses are challenging to perform for *A. lycopersici*, as the time of divergence from the most recent common ancestor with a sequenced genome is hundreds of millions of years. Nevertheless, for a set of *A. lycopersici* intron losses in highly conserved genes – for which confident assignment of intron positions could be made in multi-species protein alignments – the overwhelming majority of loss events were consistent with precise intron excisions (i.e. *Figure 2c*). This pattern is consistent with a major role for intron removal via RTMIL, which has also been suggested to be a frequent mechanism underlying intron loss events in the genomes of (comparatively) closely related *Drosophila* species (*Yenerall et al., 2011*). However, a more prominent role for precise (or nearly precise) genomic deletions of introns as a loss mechanism in *A. lycopersici* cannot be ruled out, especially as our analysis necessarily involved conserved genes for which imprecise intronic deletions would likely be highly detrimental. *A. lycopersici* also has a very rapid generation time, and as it is evolutionary distant from its closest sequenced relatives (*Figure 3*), many lineage-specific uncommon mutation events (such as genomic deletion of introns) have potentially been sampled. Currently, more closely related genomes are needed to distinguish between RTMIL or genomic deletions as the predominant driver of intron loss in *A. lycopersici*, as well as to assess contributions of other possible mechanisms – for instance, retrotransposition by target-primed reverse transcription of spliced transcripts (*Cordaux and Batzer, 2009*; *Wang et al., 2014*), with subsequent loss of source, intron-containing loci. Likewise, more closely related genomes will be critical to establish the timing of intron losses. As additional genomes in this lineage become available, eriophyoid mites promise to be an attractive system to investigate the dynamics of intron evolution.

Apart from the dearth of introns, the complement of coding genes in the *A. lycopersici* genome deviates from that of relatives with larger genomes, and seems to be associated with its reduced morphology and distinct life history (*Lindquist and Oldfield, 1996*). Compared to other arthropods, a mere handful of gene families were expanded, including one that encodes a troponin domain. While this result was unexpected, as troponin performs a conserved role in muscle contraction and is single or low copy number in most arthropods, in a transcriptome assembly of *Aceria tosichella*, a non-galling eriophyoid pest of wheat and other grasses, an expansion of troponin-encoding genes was also observed (*Gupta et al., 2019*). Possibly, this expansion may be related to the derived body musculature of eriophyoids, as their skeletal and peripheral musculature is very pronounced, with the latter enabling the maintenance of body turgidity (*Nuzzaci and Alberti, 1996*). Nevertheless, the dominant force in shaping the genic composition of *A. lycopersici* is loss, including for genes involved in highly conserved metazoan or arthropod cell processes (e.g. for the Golgi apparatus), as well as gene families and specific genes (or conserved domains) involved in many aspects of arthropod development and physiology. The latter include Hairy Orange domain proteins, nuclear receptors, and other transcription factors that have broadly conserved roles in animal development (*Holland, 2013*; *Iso et al., 2003*; *Pflugfelder et al., 2017*; *Sebé-Pedrós and Ruiz-Trillo, 2017*; *Shimeld et al., 2010*), and whose reduction (or simplification by domain loss) in *A. lycopersici* may be related to the eriophyoid body plan. For example, in contrast to other mites, *A. lycopersici* has no orthologue of the T-box gene *org-1*, which in *D. melanogaster* plays a pivotal role in diversification of circular visceral muscle (*Schaub and Frasch, 2013*). This musculature is reduced in the Eriophyoidea (*Nuzzaci and Alberti, 1996*; *Whitmoyer et al., 1972*) compared to other mites (*Alberti and Crooker, 1985*; *Coons, 1978*; *Mathieson and Lehane, 2002*), as it also is in studied microinsects (*Polilov, 2015*). Furthermore, in most chelicerates, the Hox gene *pb* is expressed in the pedipalps and in three to four pairs of legs (*Barnett and Thomas, 2013*; *Schwager et al., 2015*; *Telford and Thomas, 1998*). Whether the lack of *pb* in the *A. lycopersici* genome is related to the reduction in legs in Eriophyoidea is unknown; however, *pb* has also been lost in other ecdysozoan

animals such as Nematoda (*Aboobaker and Blaxter, 2003*) and Tardigrada (*Smith et al., 2016*; *Yoshida et al., 2017*), lineages that either lack legs (Nematoda) or in which leg formation has been suggested to be highly aberrant ('walking heads', *Maderspacher, 2016*) from the panarthropodan ancestor (Tardigrada, *Smith and Goldstein, 2017*). Further, in *D. melanogaster* mutants of both *dachs* and *four-jointed*, each of which is absent in *A. lycopersici*, have similar phenotypes including shortened legs (*Buckles et al., 2001*). *A. lycopersici*-specific losses in cell cycle regulatory genes like *unkempt* and *fat* are also candidates to underlie allometric changes in tissues and organs, a general feature of diminutive mites (like *A. lycopersici*) and microinsects (*Danks, 2006*; *Polilov, 2015*).

A remarkable feature of the genome of *T. urticae* is the presence of hundreds of genes acquired from fungal or bacterial sources, including microbe-derived UGTs (*Wybouw et al., 2018*). While a modest number of UGTs of putative bacterial origin are present in *A. lycopersici*, horizontally transferred genes were otherwise absent, except for two genes in the pathway for the synthesis of pantothenate, an essential B vitamin. Previous studies have shown that pantothenate biosynthetic genes have been laterally transferred into tetranychid mites, the silverleaf whitefly, and nematodes (*Chen et al., 2016*; *Craig et al., 2009*; *Wybouw et al., 2018*)(*Ren et al., 2020*). In *A. lycopersici*, the HGT event of ketopantoate hydroxymethyltransferase appears to be distinct from the transfer in the tetranychid mite lineage. The apparent independent HGT of pantothenate biosynthetic genes in Acariformes, coupled with acquisitions in insect lineages, is a strong signal of adaptive significance for de novo pantothenate biosynthesis in arthropod herbivores.

Finally, nowhere were reductions in *A. lycopersici* gene families more striking than in genes associated with host plant use. Recently, the importance of chemosensory receptors in host plant use and breadth has attracted intense interest (*Gloss et al., 2019*; *Ngoc et al., 2016*). *A. lycopersici* completely lacks the expansion of chemosensory receptors reported (to varying extents) in nearly all other arthropods, as only a handful of members are present for any of the characterized chemosensory receptor families. This finding is consistent with a reduced role for chemosensation in specialist herbivores, although it may also reflect a more general loss of sensory structures during miniaturization, as the number of sensilla (which include sites of chemosensation) are dramatically reduced in microinsects (*Polilov, 2015*), as well as in eriophyoid mites (*Figure 1a,b*; *Lindquist and Oldfield, 1996*). Next to chemosensory receptor genes, the detoxification gene complement of *A. lycopersici* is minimal compared to the generalist herbivore *T. urticae* (*Dermauw et al., 2013b*; *Grbić et al., 2011*), as well as to insect herbivores (*Rane et al., 2019*). This was particularly striking for CCEs and GSTs, for which lineage-specific expansions are absent, and for which most members are in highly conserved clades that likely perform more general (non-detoxification) roles. Several of the few notable lineage-specific expansions in *A. lycopersici* do involve subfamilies of the MFS. However, while some MFS genes are differentially regulated upon host shift or xenobiotic exposure in *T. urticae* (*Dermauw et al., 2013b*), MFS proteins have diverse roles, and additional work is needed to assess if MFS mini-expansions in *A. lycopersci* are associated with host use.

The minimal detoxification gene repertoire and the paucity of chemoreceptor genes in *A. lycopersici* are in line with ecological specialization theory that predicts that herbivores with a narrow host range only need a limited number of environmental response genes (*Berenbaum, 2002*; *Rane et al., 2019*). However, although *A. lycopersici* has a narrow host-range relative to the spider mite *T. urticae*, it can be found on related solanaceous plant species (*Perring and Farrar, 1986*), as well as on several hosts outside the nightshade family (*Perring and Royalty, 1996*; *Rice and Strong, 1962*). Hence, the extent to which this mite has specialized on these hosts is unclear. Nevertheless, the minimal detoxification and chemoreception repertoire gene sets support the idea that modification of the local environment by defense suppression may alter selection imposed by the environment, thereby reducing the requirement for environmental response genes (*Laland et al., 2016*). How eriophyoids manipulate their hosts is unknown, but likely involves orally delivered salivary metabolites (*De Lillo and Monfreda, 2004*), or alternatively secreted proteins, termed effectors. Currently, the molecular nature of herbivore effectors, and their mechanisms of action, are poorly understood (*Blaazer et al., 2018*; *Erb and Reymond, 2019*). However, proteins secreted by the larvae of several lepidopteran species have been shown to attenuate plant defenses, including by physical interaction with a component of the JA signal transduction pathway (*Chen et al., 2019*; *Musser et al., 2002*). Further, a salivary ferritin from the whitefly *Bemisia tabaci* suppresses oxidative signals in tomato, and blunts JA-mediate defense responses (*Su et al., 2019*), and expression of salivary products of unknown molecular function from spider mites in plants was recently demonstrated

to impair defense signaling downstream of the phytohormone salicylic acid (*Villarroel et al., 2016*), and may also act to suppress JA signaling (*Schimmel et al., 2017*). The divergent molecular nature of these effectors mirrors findings from plant-pathogen (*Toruño et al., 2016*) and plant-nematode (*Rehman et al., 2016*) systems, where secreted effectors can be highly species-specific, hindering identification based solely on sequence information. These findings highlight the need for functional studies to establish if secreted proteins (or metabolites) in *A. lycopersici* saliva underlie this mite's ability to potently suppress tomato defenses. More generally, as additional genomes of herbivores that induce or suppress plant defenses become available – and that vary in their magnitude and mechanisms of host suppression – the *A. lycopersici* genome will serve as a key reference for comparative studies to test hypotheses surrounding the evolution of gene families that respond to or modulate plant defenses.

Conclusion

At only 32.5 Mb, the *A. lycopersici* genome is the smallest sequenced arthropod genome to date. In contrast to its closest sequenced relatives, the majority of genes lack introns, few repetitive sequences are present, and many genes conserved in most animals are absent. Compared to its larger relatives, the simplification of *A. lycopersici*'s body plan, and that of eriophyoid mites more generally, is reminiscent of that observed in other microarthropods (*Maderspacher, 2016*). The compressed genome architecture of *A. lycopersici* is in line with genome streamlining concepts (*Hessen et al., 2010a*; *Hessen et al., 2010b*), some of which speculate that maintaining a high growth rate in nutritionally limited environments (in this study the plant epidermis) may be a driver for the evolution of compact genomes. Further, the extreme reduction of several environmental response gene families aligns with predictions that follow from ecological specialization theories (*Devictor et al., 2010*; *Futuyma and Moreno, 1988*; *Laland et al., 2016*) since the mite's suppression of plant defenses may allow for such families to minimize during the course of its evolution. Finally, this first eriophyoid genome provides a resource for methods of early detection of mite infestations using molecular markers, and its reduced complement of defense genes – a common source of pesticide resistance – may also reveal novel Achilles' heels for the control of *A. lycopersici*. But foremost, this genome is a milestone for accelerating our understanding of the evolutionary forces underpinning metazoan life at the limits of small physical and genome size.

Materials and methods

Collection of DNA for genomic sequencing

A. lycopersici individuals were reared in insect cages (BugDorm-44590DH, Bug Dorm Store, Mega-View Science, Taichung, Taiwan) in a walk-in growth chamber on tomato plants (*Solanum lycopersicum*, cv. Castlemart) that were between 3 and 6 weeks old. The climate room was set to day/night temperatures of 27°C/25°C, a 16/8 hr light/dark regime and 60% relative humidity. Harvesting of *A. lycopersici* mites was performed by detaching highly infested tomato leaflets and placing them in 1.5 mL Eppendorf tubes. Eppendorf tubes were filled with water and mites (adults, juveniles and eggs) were washed off by rinsing and briefly vortexing the tubes. The tubes were then centrifuged (13,000 rpm for 2 min), after which bulk tomato tissue was removed and water was pipetted away. Contamination from tomato tissue was limited to small amounts (less than ~5%) of material consisting primarily of tomato trichomes. Resulting 'pellets' of russet mites were frozen in liquid nitrogen and stored at −80°C until DNA was extracted.

DNA sequencing and genome assembly

DNA was extracted using a modified version of the CTAB method (*Doyle and Doyle, 1987*). Sixty µg of DNA dissolved in TE buffer was sent to Eurofins MWG Operon (Ebersberg, Germany) for sequencing. Sequencing reads were produced with the standard Roche/454 sequencing protocol on the GS FLX system running Data Analysis Software Modules version 2.3. Three different libraries were prepared and sequenced in accordance with the recommendations of Roche/454: random primed shotgun, 8 kb paired-end, and 20 kb paired-end. From the shotgun library the mean length was 503 bp, while for the 8 kb and 20 kb libraries the mean lengths were 366 bp and 359 bp, respectively. Sequencing reads were trimmed to remove adapters and low-quality bases, as well as

to split each paired-end read into a forward and reverse pair; this yielded a total of 1,854,028 shotgun reads, 1,076,303 reads from the 8 kb library, and 1,274,414 reads from the 20 kb library. Contigs were assembled by the in-house pipeline of Eurofins MWG Operon (Ebersberg, Germany) based on Newbler 2.6 (*Margulies et al., 2005*). Following scaffolding and filtering for plant (tomato), prokaryotic, and adaptor sequences, a reference for the nuclear genome was generated that consisted of seven scaffolds (scaffolds 1, 2, 3, 4, 5, 11, and 17) with a total length of 32.53 Mb (the Newbler 'peakDepth', or coverage, for the assembly was 38). An additional scaffold (scaffold 6) of length 13.5 kb consisted of the mitochondrial genome.

Genome size and completeness estimations

A k-mer size estimate of the *A. lycopersici* genome was performed using the genomic 454 sequence reads and Jellyfish 2.2.6 (*Marçais and Kingsford, 2011*). Following the recommendations of T. Nishiyama (http://koke.asrc.kanazawa-u.ac.jp/HOWTO/kmer-genomesize.html), genome size was estimated by running Jellyfish (*Marçais and Kingsford, 2011*) with the following settings '-t 24 iC -s 20M' for all odd k-mer values from 17 to 31, with averaging of the results provided from the eight different estimates. Completeness of the genome was also assessed using CEGMA 2.5 (*Parra et al., 2007*) as well as BUSCO v3 (*Simão et al., 2015*), as well as with an alignment of the *A. lycopersici* Illumina-based transcriptome assembly to the genomic scaffolds (see below, and Results section).

RNA collection, 454 cDNA sequencing, and transcriptome assembly

Mixed developmental stages (adults, juveniles, and eggs) were collected from tomato leaflets as was done for DNA preparation. Similar to DNA extraction, small amounts of tomato trichome contamination were evident, but at low levels. RNA was extracted using a Qiagen RNeasy kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Forty-five µg of RNA was provided to Eurofins MWG Operon for library preparation according to standard Roche protocols. Following poly(A) selection and strand-specific cDNA library preparation, the library was analyzed on a Shimadzu MultiNA microchip electrophoresis system (Shimadzu, Kyoto, Japan) to verify that the gel size selection was in the range of 500–800 bp. A total of 1,370,892 sequencing reads were collected using a Roche GS FLX system employing the Titanium series chemistry. After trimming of cDNA reads to remove low quality reads and adapter sequences, the remaining 1,370,005 reads were assembled using MIRA (*Chevreux et al., 2004*).

RNA collection, Illumina sequencing, and transcriptome assembly

RNA was extracted from eight *A. lycopersici* pools using the Qiagen RNeasy purification kit (Qiagen, Hilden, Germany) with the following adaptations: Step 3: 50 µl of RNEasy lysis buffer (RLT) + ß -mercaptoethanol were added to the mite pool in a 1.5 mL tube, followed by 1–2 min of cell lysis performed by twisting and turning a 1.5 mL-tube-pestle. Three hundred µl of RLT + -mercaptoethanol was then used to rinse the pestle; Step 11: RNA was eluted in 30 µl RNAse-free water and stored on ice. All samples were stored at −20˚C. Strand-specific paired-end RNA library preparation and sequencing were carried out by the Centro Nacional de Análisis Genómico (Barcelona, Spain) to yield a total of 86.6 million 101 bp read pairs.

To construct a transcriptome assembly from the Illumina RNA-seq reads, the reads were first aligned to the *A. lycopersici* reference genome sequence using STAR 2.5.2b (*Dobin et al., 2013*) with the following settings: twopassMode Basic, sjdbOverhang 100, and alignIntronMax 20000. Reads that did not align to the reference were subsequently aligned against the tomato genome release SL 2.50 (*Tomato Genome Consortium et al., 2012*) to filter out contamination from the host plant with the same settings used to align to the mite genome except for alignIntronMax, which remained unspecified. The reads that did not align to the tomato genome were pooled with the reads that aligned to the *A. lycopersici* genome and imported into CLC Genomics Workbench 9.0.1 (https://www.qiagenbioinformatics.com/), where they were trimmed using the default parameters (quality score limit 0.05 and a maximum of two ambiguous nucleotides) before being assembled with the default settings and a minimum contig length of 200. The resulting 13,428 transcript sequences were aligned back to the *A. lycopersici* genome assembly using BLAST 2.5.0+ (*Camacho et al., 2009*) to provide a measure of the genome completeness for transcribed regions. Of the 243 transcripts that did not align, 23 had no hits in any database, and 108, 84 and 20

appeared to be from bacterial, plant and fungal sources, respectively. Only eight had homology to arthropod sequences present in the NCBI NR, NT, Other Genomic, RefSeq Genomic, RefSeq RNA, Representative Genomes, and WGS databases (downloaded January 9, 2017).

Annotation of the *Aculops lycopersici* genome

A first-pass annotation was produced using EuGene (*Schiex et al., 2001*) specifically trained for the studied genome using the 454 transcript read data as a guide. As a consequence of the close proximity of adjacent genes (see Results and *Table 1*), we observed that transcript contigs often merged adjacent genes, creating apparent chimeric genes. To circumvent this issue, only junctions spanning introns as assessed from the aligned 454 data were kept after mapping. Besides using transcript data, protein homology to the invertebrate section from RefSeq, curated proteins from SWISSprot and the proteome from *T. urticae* were used.

Subsequently, the annotation was revised in several ways. The deep dataset of Illumina RNA-seq reads was aligned to the genome using the default settings of Bowtie 2.2.3 (*Langmead and Salzberg, 2012*)/TopHat 2.0.12 (*Kim et al., 2013*), as well as STAR 2.5.2b (*Dobin et al., 2013*) with the parameters described previously. Transcripts from the CLC transcriptome assembly were also located on the genome using BLAT 36 (*Kent, 2002*). Then, Cufflinks 2.2.1 (*Trapnell et al., 2013*) and TransDecoder (Release 20140704) (*Haas et al., 2013*) were used to identify additional ORFs of over 300 bp in length that had not been detected by EuGene. Resulting gene models were then added where supported by the strand-specific RNA-seq reads and/or transcript alignments. The compact nature of the *A. lycopersici* genome, coupled with the finding that most genes were intronless (*Table 1*), made it feasible to then manually inspect all gene models against the aligned Illumina RNA-seq read data. This inspection step was performed using the Integrative Genomics Viewer (*Robinson et al., 2011*), which allowed simultaneous display of gene models and RNA-seq read alignments. Manual adjustments to gene models, where required, were performed using Genome-View N29 (*Abeel et al., 2012*). Additionally, members of specific gene families were expertly annotated as described in the section 'Comparative analyses with specific gene families', with resulting adjustments also incorporated in the final annotation. GenomeTools 1.5.10 (*Gremme et al., 2013*) was used to sort, correct phase information, and validate the resulting GFF3.

Genome metric calculations

Coding gene numbers and the percentages of intronless genes were calculated with the 'stat -exon-numberdistri' command of the GenomeTools 1.5.6 package (*Gremme et al., 2013*) using the respective GFF3 annotation files as input (*Table 1*). Where multiple isoforms were present for a gene, only the longest isoform was used for this and subsequent analyses. Regions of the respective genomes were then classified as coding, intergenic or intronic by parsing the location of coding sequences (CDS) from the respective GFF3 annotation files; due to the unreliability of untranslated sequence prediction or their complete absence in some annotations, these regions were not considered. In instances where CDS sequences overlapped, their coordinates were merged so that no region of the genome would be counted multiple times. Regions of the genome between the start and end of the CDS sequences of adjacent genes were classified as intergenic, while regions of the genome within genes that did not fall into CDS coordinate blocks were classified as intronic (in instances where genes were located within the introns of other genes, the CDS sequences of the genes within the introns were classified as coding, with the remaining portion counted as intronic).

Transposable element annotation

The consensus of the repeated DNA (≥2 copies) in the genome was constructed by employing RepeatScout (v.1.0.5) (*Price et al., 2005*). The repeats that were ≥90% identical with a minimum overlap of 40 bp were assembled using CAP3 (*Huang and Madan, 1999*). Gene families were identified based on homology with cellular genes by employing tBLASTx 2.2.28+ (*Altschul et al., 1997*) searches against the Refseq mRNA database at NCBI and BLASTn 2.2.28+ (*Altschul et al., 1997*) searches against the annotated genes in the *A. lycopersici* genome. All candidate gene families were filtered upon manual verification. The remaining repeats were classified by REPCLASS (*Feschotte et al., 2009*) and RepeatMasker (*Smit et al., 2013*) protein searches (http://www.repeatmasker.org/cgi-bin/RepeatProteinMaskRequest). The repeats that were classified based on the

structure or TSD module of REPCLASS were manually verified. The criteria of requiring at least one defined end were used to classify a repeat as a TE. To identify if the elements had at least one defined end, the unclassified repeats (≥65 bp) were aligned with the respective copies with extended flanking sequences using MUSCLE (*Edgar, 2004*). Repeats were classified and full-length copies were extracted when possible. To identify low copy non-LTR retrotransposons, the non-LTR proteins from the related mite *T. urticae* were used as queries in homology-based tBLASTn 2.2.25+ (*Altschul et al., 1997*) searches against the *A. lycopersici* genome. To identify the genomic coverage, the curated repeat library was used to mask the genome using RepeatMasker (v 4.0.5) (*Smit et al., 2013*). The final RepeatMasker output was parsed using parseRM.pl (*Kapusta et al., 2017*; *Kapusta, 2017*) to identify the contribution of TEs (*Figure 2—figure supplement 1*, *Supplementary file 1* — 'Table S1' Tab). Last, a gene and TE density plot was constructed using karyoploteR version 1.14.0 (*Gel and Serra, 2017*) and the GFF3 annotation file of the *A. lycopersici* genome (*Table 1—source data 1*) and the RepeatMasker output (*Supplementary file 1* — 'Table S2' Tab), respectively.

Analysis of intronic features

The longest protein isoforms for the following organisms were extracted for orthologue identification: *A. lycopersici* (current genome), *Anopheles gambiae* AgamP4.7 (*Holt et al., 2002*), *Bombyx mori* ASM15162 (Ensembl release 37) (*Mita et al., 2004*), *Caenorhabditis elegans* Wormbase release WS261 (The *C. elegans* *The C. elegans Sequencing Consortium, 1998*), *Centruroides sculpuratus* CEXE 0.5.3 (*Schwager et al., 2017*), *Danio rerio* GRCz10 (Ensembl release 89) (*Howe et al., 2013*), *Daphnia pulex* PA42 3.0 (*Ye et al., 2017*), *Dermatophagoides pteronyssinus* (ASM190122v2) (*Waldron et al., 2017*), *Drosophila melanogaster* Flybase release 6.16 (*Adams et al., 2000*; *Gramates et al., 2017*), *Homo sapiens* GRCh38.p10 (Ensembl release 89) (*Lander et al., 2001*; *Venter et al., 2001*), *Ixodes scapularis* (IscaW1.5) (*Gulia-Nuss et al., 2016*), *Limulus polyphemus* 2.1.2 (*Simpson et al., 2017*), *Metaseiulus occidentalis* 1.0 (GNOMON release) (*Hoy et al., 2016*), *Parasteatoda tepidariorum* 1.0 (*Schwager et al., 2017*), *Pediculus humanus* PhumU2 (Ensembl release 36) (*Kirkness et al., 2010*), *Strigamia maritima* Smar1 (Ensembl release 36) (*Chipman et al., 2014*), *T. urticae* (ORCAE August 11, 2016 release) (*Grbić et al., 2011*), and *Tribolium castaneum* Tcas5.2 (Ensembl release 36) (*Richards et al., 2008*). The identification of orthologous protein sequences was performed with OrthoFinder 1.1.8 (*Emms and Kelly, 2015*) using BLAST 2.6.0+.

We found 147 single-copy orthologues across all species that we then aligned using MAFFT 7.305b (*Katoh and Standley, 2013*) with 'genafpair' and 'maxiterate 1000'; a concatenation of the alignments for the 147 orthologues was then generated prior to trimming with trimAl 1.4.rev15 (*Capella-Gutiérrez et al., 2009*) using the 'strictplus' option. The trimmed sequences were used for a phylogenetic reconstruction with RAxML 8.2.12 (*Stamatakis, 2014*) using the LG+G+F model as identified for phylogenetic reconstruction by ProtTest 3.4.2 (*Darriba et al., 2011*) according to the Akaike Information Criterion, and a total of 1000 rapid bootstrap replicates ('-f a -x 12345' option). Although the 'estimate proportion of invariable sites (+I)' was also recommended by ProtTest, the developer of RAxML v8, on page 59 of the RAxML v8.2.X manual, cautions against using this option, and this and all subsequent optimal models for reconstructions with RAxML were adjusted to adhere to this developer recommendation.

Orthologous protein clusters were selected for intron analysis on the basis of the following criteria: the cluster had to have at least one orthologue from *A. lycopersici*, orthologous protein sequences from at least 14 other species had to be present, and no species could have more than three orthologous proteins in the cluster; when multiple orthologues for a single species were present, only the longest one was retained. The sequences in these clusters were aligned using MAFFT 7.305b (*Katoh and Standley, 2013*) with the settings previously described. GNU Parallel (*Tang, 2011*) was used to align multiple clusters at once. Custom Python scripts using BioPython 1.70 (*Chapman and Chang, 2000*) and the BCBio GFF parser (*Chapman, 2016*) were used to parse and append intron position information to the FASTA sequence identifier line as required by Malin (*Csűrös, 2008*). The 2371 clusters that met the requisite criteria, along with the tree built from the 147 single-copy orthologues, were used in the Malin analysis (*Csűrös, 2008*). Intron positions for gain/loss analysis were selected from those that were considered unambiguous in *A. lycopersici* and at least 11 other species, with five amino acids present on either side of the intron position (a Malin criteria to reduce the possibility of incorrect inference resulting from misalignments).

To investigate the consequence of intron losses in *A. lycopersici* on predicted protein sequences, which can shed light on underlying mechanisms of loss (see Discussion), a subset of orthogroups was selected for which sequences for each of *A. lycopersici*, *D. pteronyssinus*, *T. urticae*, *M. occidentalis*, *B. mori* and *D. melanogaster* were present as single copies (1216 in total); apart from *A. lycopersici*, the five other species were selected because of their close phylogenetic position to *A. lycopersici* (*Figure 3*), and/or because they have high-quality genomes and annotations. The protein sequences for the six species for each of these orthogroups were aligned using MAFFT 7.407 with the settings previously described, and a table of intron sites for these orthogroups was generated in Malin using the following settings: Minimum nongap positions: 0 (On both sides); Minimum unambiguous characters at a site: 1; There must be at least one unambiguous character in the following clades: All unselected. From this table, intron positions that were present and had the same phase in all arthropods except *A. lycopersici* (indicating a high degree of conservation), and for which Malin identified a missing intron in a region of unambiguous alignment for *A. lycopersici* sequences, were manually examined across all protein sequence alignments to assess if intron loss events in the respective genes introduced gains or losses of residues in the encoded products. The classification results for these sites (100 in total among 80 orthogroups), are included in *Supplementary file 1* — 'Table S3' Tab; the sequence alignments and annotations of intron positions for each orthogroup are given in *Supplementary file 3*.

Gene family expansions and contractions

The OrthoFinder analysis (see section 'Analysis of intronic features') generated 86,686 orthologous groups (OGs) in total, of which 13,817 contained more than one protein (*Supplementary file 1* — 'Table S7' Tab). InterProScan 5.25–64.0 (*Quevillon et al., 2005*) was run to assign domains to each of the proteins in all 18 species, and the domain information was subsequently assigned to the OrthoFinder OGs using the KinFin software (*Laetsch and Blaxter, 2017*) and an associated Python script (functional_annotation_of_clusters.py with the options: '–p 0.3 and –x 0.3'). Two different strategies were used to identify contracted and/or expanded gene families in *A. lycopersici*. First, we used the CAFE software to detect contracted/expanded orthologous groups (orthogroups, OGs) among 18 metazoan species, while the second strategy was focused on OG expansions within the acariform mites, *A. lycopersici*, *D. pteronyssinus* and *T. urticae* using an arbitrary rule. OrthoFinder 1.1.8 (*Emms and Kelly, 2015*) with BLAST 2.6.0+ was used to identify OGs among the proteomes of 18 metazoan species (see 'Analysis of intronic features' for proteome versions that were used as input for OrthoFinder).

To maximize the probability of achieving convergence in the maximum likelihood analysis performed in CAFE, OGs were processed to remove OGs present in only a few species and were subsequently divided into OGs having <100 gene copies in any species ('small' OGs) and orthogroups having one or more species with ≥100 gene copies ('large' OGs); see 'Known Limitations' section in CAFE 4.0 Manual of March 14, 2017 and section 2.2.4 of the CAFE 4.0 tutorial online at https://iu.app.box.com/v/cafetutorial-pdf, and also *Casola and Koralewski, 2018*. We retained 6,496 OGs that occurred in no less than 10 out of 18 species consisting of 6,467 'small' OGs and 29 'large' OGs. Together with an ultrametric species tree the 'small' OG dataset was used as input in CAFE to estimate the birth/death parameter λ (the probability that a gene will be gained or lost) and to identify rapidly evolving OGs (*p*-value threshold of 0.05). The estimated λ (0.00055594301461) was then used to identify rapidly evolving OGs in a CAFE analysis with 'large' OGs and using the same p-value threshold and ultrametric species tree as in the CAFE analysis with 'small' OGs. The ultrametric species tree used as input in both CAFE analyses was obtained by using the species tree generated for the Malin intron analysis, subsequently rooting this tree using vertebrates as outgroup, and converting this rooted tree into an ultrametric tree using the *convert_to_ultrametric()* command in the *Tree* package of the ETE toolkit (ete 3.0.0b35) (*Huerta-Cepas et al., 2016*). Next, branch lengths of the ultrametric tree were scaled to time units using the software treePL (*Smith and O'Meara, 2012*) with the following options: 'smooth = 0.01, numsites = 41107 (number of sites in the alignment used for the Malin analysis), thorough, opt = 4, moredetailad, optad = 2, optcvad = 2, moredetailcvad' and using seven calibration timepoints: the divergence time between Eriophyoidea and Sarcoptiformes (352–410 MYA), Sarcoptiformes and Trombidiformes (410–421 MYA) and Mesostigmata and Ixodida (283–418 MYA) as derived from *Xue et al., 2017*, and the divergence time between *D. melanogaster* and *A. gambiae* (211–335 MYA), Scorpiones and Araneae (379–410 MYA), Mandibulata

and Chelicerata (560–642 MYA) and *H. sapiens* and *D. rerio* (425–446 MYA), as obtained from Time-Tree (*Kumar et al., 2017*) on February 20, 2019. The options used in treePL were determined following the 'Quick run' guidelines of the treePL wiki (*Smith, 2012*). The output of the two CAFE analyses ('small' and 'large' OGs) was summarized using a Python script (cafetutorial_report_analysis.py using the '-l' option and with a *p*-value cutoff set to 0.05) available at the CAFE tutorial website (https://iu.app.box.com/v/cafetutorial-files/folder/22161236877, accessed February 20, 2019). The tree with OG expansions and contractions was visualized in MEGA 6.0 (*Tamura et al., 2013*) and edited with Corel Draw software (Corel Draw, Inc); the list of rapidly evolving (expanding or contracting) OGs can be found in *Supplementary file 1* — 'Table S5' Tab. Rapidly contracting *A. lycopersici* gene families with more than ten members were analyzed for the percentage of *A. lycopersici* members showing orthology with the majority of chelicerate species (*Supplementary file 1* — 'Table S6' Tab). Orthology was determined based on the Orthofinder generated output in the 'Orthologues_Aculops_lycopersici' folder. One of the six rapidly contracted *A. lycopersici* families belonged to the CYP family and was excluded from the analysis, as only few orthology relationships has been observed within this family (*Feyereisen, 2011*).

Apart from gene families that we identified as expanded in the high-level CAFE analysis, we looked as well for more subtle expansions across acariform mites. Across all orthogroups identified by Orthofinder, we identified eleven orthogroups with (1) *A. lycopersici* having more than five members and (2) *A. lycopersici* having twofold more members than the average number in *T. urticae* and *D. pteronyssinus* (OG0000024, OG0000271, OG0000546, OG0000706, OG0004829, OG0006109, OG0006384, OG0007553, OG0007554, OG0008410, OG0008412). For two orthogroups (OG0007554, OG00084112), no InterPro domain could be assigned, while OG0000271, OG0000706, OG0004829, OG0006384, OG0007553, and OG0008410 contained proteins with a DnaJ domain (IPR011701), Formate-tetrahydrofolate ligase domain (IPR000559), Acyltransferase 3 domain (IPR002656), a Peptidase C1A domain (IPR000668), Chromo domain (IPR023780) and a Lipase/vitellogenin domain (IPR013818), respectively. The proteins of the three remaining orthogroups (OG0000024, OG0000546, and OG0006109) belonged to the Major facilitator superfamily (MFS, IPR011701 or IPR024989).

Gene ontology enrichment analysis of absent conserved genes and identification of orthogroups containing *Drosophila* essential genes

For *D. melanogaster* proteins belonging to orthogroups with (1) members in all included arthropods except *A. lycopersici* and (2) a maximum of two *D. melanogaster* members (343 *D. melanogaster* proteins in total, *Supplementary file 1* — 'Table S8' Tab), we performed an Over-Representation analysis (ORA) using the WEB-based GEne SeT AnaLysis Toolkit (*Liao et al., 2019*). An ORA was performed for each GO category (Biological Process, Molecular Function and Cellular Component) using default settings (and 'genome protein coding' as reference set) and a Benjamini-Hochberg multiple testing correction (false discovery rate, FDR, of 0.05). In addition, we also identified those orthogroups that contain purported *D. melanogaster* essential genes, using the list of 427 essential genes provided in the respective study's first supplementary data table (*Aromolaran et al., 2020*).

Comparative analyses with specific gene families

We specifically analyzed genes and gene families associated with herbivory in other animals, as well as those associated with physiological or developmental process related to *A. lycopersici*'s life history or derived morphology (GSTs, CCEs, CYPs, ABC transporters, MFS proteins, proteases, chemosensory receptors, and transcription factors, including Hox genes). We also characterized genes involved in processes including circadian rhythm, small RNA pathways, and potential regulation of plant defense responses (secreted proteins).

Characterization of detoxification and feeding associated gene families
Glutathione-S-transferases

The *A. lycopersici* genome and proteome were mined for glutathione-S-transferases (GSTs) by tBLASTn and BLASTp searches, respectively, using cytosolic and microsomal *T. urticae* GST protein sequences as query (*Grbić et al., 2011*) and an E-value threshold of E^{-5}. In total, four *A. lycopersici* cytosolic GSTs were identified. *A. lycopersici* cytosolic GSTs were aligned with those of *T. urticae* (31

GSTs) (*Grbić et al., 2011*), *D. melanogaster* (36 GSTs, the atypical GST CG4623/Gdap1 was not included as it is very divergent from other *D. melanogaster* GSTs) (*Shi et al., 2012*) and *M. occidentalis* (13 GSTs) (*Wu and Hoy, 2016*) using the online version of MAFFT v7.356b (*Katoh and Standley, 2013*) with 1000 iterations with the options 'E-INS-i' and 'reorder' (see *Supplementary file 7*). Model selection was performed with ProtTest 3.4 (*Darriba et al., 2011*), and according to the Akaike information criterion LG+I+G+F was optimal for the phylogenetic reconstruction. A maximum likelihood analysis was performed using RAxML v8 HPC2-XSEDE (*Stamatakis, 2014*) on the CIPRES Science Gateway (*Miller et al., 2010*) with 1000 rapid bootstrapping replicates ('-f a -x 12345' option). The resulting tree was midpoint rooted, visualized using MEGA 6.0 (*Tamura et al., 2013*) and edited with Corel Draw software (Corel Draw Inc).

Carboxyl/cholinesterases

Putative carboxyl/cholinesterase (CCE) genes were identified in *A. lycopersici* using tBLASTn and BLASTp searches (E-value threshold of E^{-5}) with *T. urticae* CCE sequences (*Grbić et al., 2011*) as query. Putative *A. lycopersici* CCEs were aligned with those of *T. urticae* (*Grbić et al., 2011*), *M. occidentalis* (*Wu and Hoy, 2016*), a selection (8) of conserved CCEs from the horseshoe crab *Limulus polyphemus* (*Wei et al., 2020*), a selection (10) of *D. melanogaster* CCEs belonging to different CCE clades (*Claudianos et al., 2006*), and AChE1/AChE2 of *B. mori* and *D. pulex* using the online version of MAFFT v7.380 (*Katoh et al., 2019*) with 1000 iterations and the options 'L-INS-i' and 'reorder' (see *Supplementary file 7*). Maximum likelihood phylogenetic analysis was performed as in *Wei et al., 2020* using RAxML v8 HPC2-XSEDE (*Stamatakis, 2014*) on the CIPRES Science Gateway (*Miller et al., 2010*) and the automatic protein model assignment algorithm using maximum likelihood criterion and 500 rapid bootstrap replicates ('-f a -x 12345' option). The resulting tree was midpoint rooted and visualized using MEGA 6.0 (*Tamura et al., 2013*) and edited with Adobe Illustrator software (Adobe Inc).

Cytochrome P450 monooxygenases and diflavin reductases

The *A. lycopersici* genome and proteome was mined for cytochrome P450 monooxygenase (CYP) genes by tBLASTn and BLASTp searches using *T. urticae* CYP protein sequences as query (*Grbić et al., 2011*) and an E-value threshold of E^{-5}. All CYP gene models with predicted proteins that included the canonical heme-binding sequence were verified manually for the presence of the other key features of P450 enzymes (*Feyereisen, 2012*) and gene models were corrected when necessary. New *A. lycopersici* CYP gene models were created using GenomeView (*Abeel et al., 2012*). All CYP sequences were named according to the CYP nomenclature by Dr. D. R. Nelson (University of Tennessee, USA). Pseudogenes (*CYP18C2P* and *CYP3120A4P)* were distinguished from putative full length CYP coding sequences by a long in frame non-P450 insertion (*CYP18C2P*) and by a stop codon and frameshift (*CYP3120A4P*), both anomalies confirmed by their respective transcripts. All *A. lycopersici* CYP protein sequences (full-length and pseudogenes) were aligned with CYP protein sequences from *T. urticae, M. occidentalis*, a set of *D. melanogaster* marker P450 sequences and the CYP18 protein sequence of the house dust mite *D. pteronyssinus* (Dpte.g6170.1) using MAFFT v7.380 (*Katoh et al., 2019*) with 1000 iterations and the options 'E-INS-i' and 'reorder' (see *Supplementary file 7*). Model selection was done with ProtTest 3.4 (*Darriba et al., 2011*) and according to the Akaike information criterion LG+I+G+F was optimal for phylogenetic reconstruction. A maximum likelihood analysis was performed using RAxML v8 HPC2-XSEDE (*Stamatakis, 2014*) on the CIPRES Science Gateway (*Miller et al., 2010*) with 1000 rapid bootstrapping replicates ('-f a -x 12345' option). The resulting tree was midpoint rooted and visualized using MEGA 6.0 (*Tamura et al., 2013*).

ABC transporters

Putative *A. lycopersici* ATP-binding cassette (ABC) genes were identified by BLASTp and tBLASTn searches (E-value threshold of E^{-5}) against the *A. lycopersici* proteome and genome, respectively, and using *T. urticae* ABC protein sequences (*Dermauw et al., 2013a*) as query. *A. lycopersici* ABC pseudogenes and incomplete genes [aculy01g37790, aculy01g37820 (pseudogenes), and aculy01g27210 (incomplete gene)] were separated from putative full-length ABC coding sequences. Putative *M. occidentalis* ABC genes were identified by a BLASTp search against the *M. occidentalis* proteome using *T. urticae* and *D. melanogaster* ABC protein sequences as query (*Dermauw et al.,*

2013a). The nucleotide-binding domain (NBD) sequences of *A. lycopersici, T. urticae, M. occidentalis*, and *D. melanogaster* ABC protein sequences were extracted using the ScanProsite facility (*de Castro et al., 2006*) and the Prosite profile PS50893. The NBDs of four putative *M. occidentalis* ABC proteins (GNOMON-2147495233, GNOMON-2147494257, GNOMON-2147494305 and GNO-MON-2147512403) had best BLASTp hits with bacterial ABC sequences and were excluded from further analysis. N-terminal NBDs of *A. lycopersici* (44), *T. urticae* (103), *M. occidentalis* (55), and *D. melanogaster* (56) ABC proteins were aligned using the online version of MAFFT v7.380 (*Katoh et al., 2019*) with 1000 iterations and the options 'G-INS-i' and 'reorder' (see *Supplementary file 7*). Model selection was performed with ProtTest 3.4 (*Darriba et al., 2011*) and according to the Akaike information criterion LG+G+F was optimal for phylogenetic reconstruction. Next, a maximum likelihood analysis was performed using RAxML v8 HPC2-XSEDE (*Stamatakis, 2014*) on the CIPRES Science Gateway (*Miller et al., 2010*) with 1000 rapid bootstrapping replicates ('-f a -x 12345' option). The resulting tree was midpoint rooted and visualized using MEGA 6.0 (*Tamura et al., 2013*) and edited with Adobe Illustrator software (Adobe Inc).

Major facilitator superfamily proteins

A. lycopersici members of two orthogroups (OG0000024 and OG0006109) that have an MFS Inter-Pro domain (IPR011701 or IPR024989), and were expanded in *A. lycopersici* (see Results), were used as query in tBLASTn and BLASTp searches (with an E-value threshold of E^{-5}) against the *A. lycopersici* genome and proteome, respectively. Next, the *A. lycopersici* queries and resulting hits were used as query in tBLASTn and BLASTp searches (with an E-value threshold of E^{-5}) against the genome and proteome of *T. urticae* and in a BLASTp search (using an E-value threshold of E^{-5}) against the proteomes of *M. occidentalis* and *D. melanogaster* (for genes with multiple isoforms, only the longest protein isoform was retained). Incomplete MFS genes (less than 250 amino acids long) were separated from full-length MFS genes. Full-length MFS proteins (*A. lycopersici*: 27, *T. urticae*: 23, *M. occidentalis*: 60 and *D. melanogaster*: 18) were aligned using MAFFT v7.356b (*Katoh and Standley, 2013*) with 1000 iterations with the options 'E-INS-i' and 'reorder' (see *Supplementary file 7*). Model selection was done with ProtTest 3.4 (*Darriba et al., 2011*) and according to the Akaike information criterion LG+I+G+F was optimal for the phylogenetic reconstruction of mite and *D. melanogaster* MFS proteins. A maximum likelihood analysis was performed using RAxML v8 HPC2-XSEDE (*Stamatakis, 2014*) on the CIPRES Science Gateway (*Miller et al., 2010*) with 1000 rapid bootstrapping replicates ('-f a -x 12345' option). The resulting tree was midpoint rooted, visualized using MEGA 6.0 (*Tamura et al., 2013*) and edited with Corel Draw software (Corel Draw, Inc).

C1A cysteine proteases

OG0006384, one of the few expanded OGs in *A. lycopersici* contained proteins with a 'Peptidase C1A, papain C-terminal' domain (InterPro domain IPR000668) (see Results). Subsequently the complete proteome of *A. lycopersici, M. occidentalis* and *D. melanogaster* was mined for IPR000668 domain containing proteins/C1A peptidases. *T. urticae* C1A peptidases were previously annotated (*Grbić et al., 2011*). Thirty-nine, 16, 28, and 57 C1A peptidase genes were found in *A. lycopersici, M. occidentalis, D. melanogaster*, and *T. urticae*, respectively. Protein sequences from 32, 13, 27, and 52 *A. lycopersici, M. occidentalis, D. melanogaster* and *T. urticae* C1A peptidase genes were larger than 250 aa, respectively, and aligned using MAFFT version 7 (*Katoh and Standley, 2013*) with 1000 iterations with the options 'E-INS-i' and 'reorder' (see *Supplementary file 7*). Model selection was performed with ProtTest 3.4 (*Darriba et al., 2011*), and according to the Akaike information criterion VT+G was optimal for phylogenetic reconstruction. A maximum likelihood analysis was performed using RAxML v8 HPC2-XSEDE (*Stamatakis, 2014*) on the CIPRES Science Gateway (*Miller et al., 2010*) with 1000 rapid bootstrapping replicates ('-f a -x 12345' option) and the VT+G model. The resulting tree was midpoint rooted, visualized using MEGA 6.0 (*Tamura et al., 2013*) and edited with Corel Draw software (Corel Draw Inc).

Gustatory receptors

Potential gustatory receptor (GR) genes were identified with BLASTp using query gustatory receptor sequences from *D. melanogaster* (*Robertson et al., 2003*), *D. pulex* (*Peñalva-Arana et al., 2009*), *M. occidentalis* (*Hoy et al., 2016*), and *T. urticae* (*Ngoc et al., 2016*), as well as odorant receptor

sequences from *D. melanogaster* (*Robertson et al., 2003*). Further, searches were performed with query sequences against the *A. lycopersici* genome using tBLASTn from the BLAST 2.6.0+ (*Camacho et al., 2009*) suite allowing an E-value of up to 1 (*Ngoc et al., 2016*). Where required, existing gene models were modified or new models were added using GenomeView N29 (*Abeel et al., 2012*). InterProScan 5.25–64.0 (*Quevillon et al., 2005*) was used to validate one of the two existing genes (*aculy03g00430*) with the '7tm Chemosensory Receptor' InterPro domain, while strong BLAST support against *T. urticae* GR genes (best hit E-value $<E^{-22}$) was observed for *aculy03g06080*. The two putative GR genes were then aligned back to the genome with tBLASTn to identify additional sequences, but none were identified. Protein sequences of GR genes from *A. lycopersici*, *T. urticae*, *M. occidentalis*, and *D. melanogaster*, were aligned using version 7 of MAFFT (*Katoh and Standley, 2013*) with the 'E-INS-i' option selected on the web service hosted by the Computational Biology Research Consortium (https://mafft.cbrc.jp/alignment/server/) (see *Supplementary file 7*). Model selection was performed by ProtTest 3.4.2 (*Darriba et al., 2011*), with the JTT+I+G+F model selected as the best according to the Akaike information criterion for phylogenetic reconstruction. The CIPRES Science Gateway (*Miller et al., 2010*) 'RAxML-HPC on XSEDE' tool (*Stamatakis, 2014*) was used to construct a phylogenetic tree using 1000 rapid bootstrap replicates ('-f a -x 12345' option), which was subsequently visualized using MEGA7 (*Kumar et al., 2016*, p. 7) and edited in Adobe Illustrator CC 2017 (Adobe Software, Inc).

Degenerin/epithelian Na+ channels

Candidate *A. lycopersici* degenerin/epithelial Na+ Channels (ENaC) genes were identified by aligning *D. melanogaster* (*Zelle et al., 2013*) and *T. urticae* ENaCs (*Ngoc et al., 2016*) against the *A. lycopersici* genome using tBLASTn, allowing an E-value of up to 1. Gene model adjustment or creation was performed with GenomeView N29. The presence of the Pfam PF00858 domain ('Amiloride-sensitive sodium channel') identified using InterProScan 5.25–64 was used as an additional criteria to identify ENaCs genes, see *Ngoc et al., 2016*. An additional round of genomic searches using tBLASTn with the four *A. lycopersici* ENaCs revealed no additional members. *A. lycopersici*, *T. urticae*, *M. occidentalis* and *D. melanogaster* ENaC protein sequences were aligned using MAFFT version 7 (*Katoh and Standley, 2013*) with the 'E-INS-i' option selected on the web service hosted by the Computational Biology Research Consortium (https://mafft.cbrc.jp/alignment/server/) (see *Supplementary file 7*). WAG+I+G+F was identified as the best model for phylogenetic reconstruction according to the Akaike information criterion by ProtTest 3.4.2 (*Darriba et al., 2011*). The 'RAxML-HPC on XSEDE' tool (*Stamatakis, 2014*) hosted by the CIPRES Science Gateway (*Miller et al., 2010*) was used to construct a phylogenetic tree with 1000 rapid bootstrap replicates ('-f a -x 12345' option). MEGA7 (*Kumar et al., 2016*) was used to visualize the resulting tree, which was subsequently edited in Adobe Illustrator CC 2017 (Adobe Software, Inc).

Ionotropic receptors

Putative ionotropic receptor (IR) and related genes were identified by aligning *D. melanogaster* (*Gramates et al., 2017*) and *T. urticae* IRs (*Ngoc et al., 2016*) along with ionotropic glutamate receptors (iGluR) and glutamate ionotropic receptor NMDA type (GRIN) sequences from *T. urticae* (*Ngoc et al., 2016*) to the *A. lycopersici* reference genome using tBLASTn with an E-value of up to one allowed. Where appropriate, GenomeView N29 was used to manually adjust or create new gene models based on the alignments. BLAST hits with E-values $<E^{-5}$, in combination with detection of the Pfam domains PF00060, PF01094 and/or PF10613 were used to classify *A. lycopersici* genes as members of the IR/iGluR/GRIN group (*Ngoc et al., 2016*). Iterative tBLASTn searches with the 10 identified sequences to the *A. lycopersici* genome identified no additional candidates. IR/iGluR/GRIN protein sequences from *A. lycopersici*, *T. urticae*, *M. occidentalis*, and *D. melanogaster* were aligned using the Computational Biology Research Consortium's MAFFT version 7 (*Katoh and Standley, 2013*) web service (https://mafft.cbrc.jp/alignment/server/) with the 'E-INS-i' option selected excepting *M. occidentalis* non-IR sequences, which were not provided in the supplementary files of *Hoy et al., 2016* (see *Supplementary file 7*). ProtTest 3.4.2 (*Darriba et al., 2011*) identified LG+I+G+F as the best model for phylogenetic construction according to the Akaike information criterion. A phylogenetic tree was constructed using the 'RAxML-HPC on XSEDE' tool (*Stamatakis, 2014*) hosted by the CIPRES Science Gateway (*Miller et al., 2010*) with 1000 rapid bootstrap replicates ('-f a -x 12345' option). Visualization of the tree was performed using MEGA7

(*Kumar et al., 2016*, p. 7), with further edits carried out in Adobe Illustrator CC 2017 (Adobe Software, Inc).

Transient receptor potential channels

The transient receptor potential (TRP) channel sequences for *D. melanogaster*, *M. musculus*, *M. occidentalis*, and *T. urticae* identified by *Peng et al., 2015* were downloaded from Ensembl (for *D. melanogaster*, *M. musculus*, and *M. occidentalis*) or ORCAE (*T. urticae*) using the IDs provided in that study. These protein sequences were aligned to the *A. lycopersici* genome sequence using tBLASTn 2.6.0+ to identify candidate TRP genes using an E-value of E^{-10}, as in *Peng et al., 2015*. Where appropriate, gene models were manually updated with GenomeView N29 using a combination of the BLAST alignments and transcriptome data. TRP channel protein sequences for *A. lycopersici*, *T. urticae*, *M. occidentalis*, and *D. melanogaster* were aligned using MAFFT version 7 (*Katoh and Standley, 2013*) with the 'E-INS-i' option selected using the web service hosted by the Computational Biology Research Consortium (https://mafft.cbrc.jp/alignment/server/) (see *Supplementary file 7*). The LG+I+G+F model was identified as optimal for phylogenetic construction by ProtTest 3.4.2 (*Darriba et al., 2011*) according to the Akaike information criterion. A phylogenetic tree was generated using the 'RAxML-HPC on XSEDE' tool (*Stamatakis, 2014*) hosted on the CIPRES Science Gateway (*Miller et al., 2010*), with 1000 rapid bootstrap replicates ('-f a -x 12345' option) and members of the Shaker family set as an outgroup after *Peng et al., 2015* to anchor the tree. MEGA7 (*Kumar et al., 2016*) was used to visualize the tree, with subsequent editing performed in Adobe Illustrator CC 2017 (Adobe Software, Inc).

Characterization of transcription factors

Pfam domains that were assigned by InterProScan to each of the proteins in the 18 metazoan species included in the Orthofinder analysis (see 'Gene family expansions and contractions' in Results) were mined for all Pfam transcription factor (TF) domains as defined in 'Table 1' of *Huang et al., 2012* and two PFAM domains [BTB (PF00651) and BACK (PF07707)] that have been implicated in transcriptional regulation (*Stogios and Privé, 2004*). Results were summarized using the *dplyr* (*Wickham et al., 2017*) and *stringr* packages (*Wickham, 2017*) within the R framework (*R Development Core Team, 2018*). Additionally, we characterized a subset of transcription factor families in greater depth [the nuclear receptor (NR), T-box, Hairy Orange, and Hox families].

Analysis of nuclear receptors

A reciprocal BLASTp analysis (using an E-value threshold of E^{-10}) was performed using the *A. lycopersici*, *D. pteronyssinus*, and *T. urticae* proteomes and using the *T. urticae* nuclear receptor (NR) protein sequences as queries (*Grbić et al., 2011*) to identify putative *A. lycopersici* and *D. pteronyssinus* NRs. A tBLASTn search (using an E-value threshold of E^{-10}) using *T. urticae* NR protein sequences as queries (*Grbić et al., 2011*) was also performed against the *A. lycopersici* genome but only overlap with existing *A. lycopersici* NR gene models was found. LBDs of *A. lycopersici* NRs were considered present if searching with PfamScan (https://www.ebi.ac.uk/Tools/pfa/pfamscan/) or Conserved domain (CD)- search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) yielded either a PF00104 (LBD of hormone nuclear receptor) or cl11397 (The ligand binding domain of nuclear receptors, a family of ligand-activated transcription regulators) domain, respectively. Those *A. lycopersici* NRs that, in contrast to their orthologues in arthropods, were not predicted with a LBD were aligned with their orthologues in *D. melanogaster* (*Thomson et al., 2009*), *D. pteronyssinus* and *T. urticae* using the online version of MAFFT v7.380 (*Katoh et al., 2019*) with 1000 iterations and the options 'E-INS-i' and 'reorder'.

T-box transcriptional regulators

All *A. lycopersici* T-box proteins identified in our PFAM transcription factor domain analysis were first used in BLASTp and tBLASTn searches (E-value threshold E^{-10}) against the *A. lycopersici* predicted proteome and genome, respectively, and no additional T-box gene models were identified. T-box proteins of *D. pteronyssinus* and *M. occidentalis* were identified in their proteomes by a BLASTp search (E-value threshold E^{-10}) using the conserved T-box domain amino acids 198–385 of *D. melanogaster* org-1 (FBpp0311870) as query, while those of *T. urticae* and *D. melanogaster* were derived from the PFAM analysis. *A. lycopersici* T-box proteins were aligned with those of *T. urticae*, *D.

pteronyssinus, *M. occidentalis*, and *D. melanogaster* using the online version of MAFFT v7.380 (*Katoh et al., 2019*) with 1000 iterations and the options 'E-INS-i' and 'reorder' (see *Supplementary file 7*). Model selection was performed with ProtTest 3.4 (*Darriba et al., 2011*) and according to the Akaike information criterion LG+I+G+F was optimal for phylogenetic reconstruction. Next, a maximum likelihood analysis was performed using RAxML v8 HPC2-XSEDE (*Stamatakis, 2014*) on the CIPRES Science Gateway (*Miller et al., 2010*) with 1000 rapid bootstrapping replicates ('-f a -x 12345' option). The resulting tree was midpoint rooted, visualized using MEGA 6.0 (*Tamura et al., 2013*) and edited with Corel Draw software (Corel Draw, Inc).

A. lycopersici Hairy Orange domain proteins

A. lycopersici and *D. pteronyssinus* orthologues of *T. urticae* Hairy Orange domain (PF07527) proteins were identified by a BLASTp search (E-value E^{-5}) against the *A. lycopersici* (this study) and *D. pteronyssinus* proteome (*Waldron et al., 2017*) using *T. urticae* Hairy Orange domain proteins as query. The resulting *A. lycopersici* hits were aligned with their counterparts in *D. melanogaster* (*Dearden, 2015*), *D. pteronyssinus*, and *T. urticae* using the online version of MAFFT v7.380 (*Katoh et al., 2019*) with 1000 iterations and the options 'E-INS-i' and 'reorder'.

A. lycopersici Sox proteins

The high mobility group (HMG)-box domain (Pfam domain PF00505) of *D. melanogaster* Sox proteins (*Janssen et al., 2018*) was used as query in a BLASTp search against the *A. lycopersici*, *D. pteronyssinus*, *T. urticae*, and *M. occidentalis* proteomes. For each species, those BLASTp hits that had an E-value lower than the lowest E-value of BLASTp hits with the species orthologue of *Drosophila* capicua (a HMG-box domain protein used as outgroup in phylogenetic analysis of Sox proteins [*Janssen et al., 2018*]; aculy02g30040, g444.t1, tetur21g00740 and rna18440 in *A. lycopersici*, *D. pteronyssinus*, *T. urticae*, and *M. occidentalis*, respectively) were retained as putative Sox proteins. Almost all Acari Sox proteins contained the highly conserved RPMNAFMVW motif, characteristic of Sox proteins (*Bonatto Paese et al., 2018*); the one exception was aculy04g11170, which has a minor conservative substitution (Ala to Ser) in this motif. A tBLASTn search, using the HMG-box domain of *A. lycopersici* BLASTp hits with *D. melanogaster* Sox proteins as query, was performed to identify non-annotated *A. lycopersici* proteins; yielding one additional *A. lycopersici* Sox protein (aculy02g08510), for which a pseudogene model was created using GenomeView (*Abeel et al., 2012*). *D. pteronyssinus*, *T. urticae*, and *M. occidentalis* Sox and capicua proteins were aligned with the HMG domain of *D. melanogaster* and *P. tepidariorum* Sox proteins (*Janssen et al., 2018*) using MAFFT v7.380 (*Katoh et al., 2019*) with 1000 iterations and the options 'E-INS-i' and 'reorder'. Next, the alignment was trimmed (see *Supplementary file 7*) to contain the HMG domain only and a phylogenetic analysis of Sox HMG-box domains was performed, similar to the analysis described in *Zhong et al., 2011*. Model selection was performed with ProtTest 3.4 (*Darriba et al., 2011*) and according to the Akaike information criterion LG+I+G was optimal for phylogenetic reconstruction. A Bayesian inference was performed using MrBayes 3.2.7a (*Huelsenbeck and Ronquist, 2001*) on XSEDE on the CIPRES Science Gateway (*Miller et al., 2010*). The Monte Carlo Markov Chain search was run with four chains over 1000000 generations with trees sampled every 1000 generations. The first 250 trees were discarded as 'burn-in'. The remaining trees were used to calculate Bayesian posterior probabilities. The resulting tree was converted into a newick format using a Perl script named AfterPhylo.pl (*Zhu, 2014*), rooted with capicua proteins, visualized using MEGA 6.0 (*Tamura et al., 2013*) and edited with Corel Draw software (Corel Draw, Inc).

Hox genes

Hox protein sequences of the oribatid mite *A. longisetosus* (*Sharma et al., 2014*), the spider mite *T. urticae* (*Grbić et al., 2011*), the deer tick *I. scapularis* (*Pace et al., 2016*) and the red flour beetle *T. castaneum* (*Pace et al., 2016*) were aligned using MAFFT v7.38 (*Katoh et al., 2019*) with 1000 iterations and the options 'L-INS-i' and 'reorder'. The 57 amino acid Homeobox domains (Pfam domain PF00046) were extracted from this alignment and used as query in a tBLASTn search (using an E-value threshold of E^{-10}) against the *A. lycopersici* genome to identify Hox (and by extension, also Homeobox) genes that were not automatically predicted. In one case a tBLASTn hit did show no overlap with an existing gene model and a new *A. lycopersici* Homeobox gene model (aculy01g39110) was created using GenomeView (*Abeel et al., 2012*).

To identify putative *A. lycopersici* orthologues of Hox proteins, a reciprocal BLASTp analysis (using an E-value threshold of E^{-10}) was performed against the *A. lycopersici* proteome (including proteins encoded by newly created gene models) using full-length *T. urticae*, *I. scapularis* and *T. castaneum* Hox protein sequences and their available proteomes (*T. urticae* version 11 August 2016, *T. castaneum* version 5.2.36 and *Ixodes scapularis* Wikel colony version 1.5). Finally, to verify the results of our reciprocal BLASTp analysis, we performed an additional BLASTp search (using an E-value threshold of E^{-10}) with the partial but well-studied *A. longisetosus* Hox protein sequences (*Barnett and Thomas, 2013*; *Sharma et al., 2014*; *Telford and Thomas, 1998*) as query. Using a similar approach (reciprocal BLASTp analysis with *I. scapularis* Hox proteins/*Ixodes scapularis* Wikel colony version 1.5 proteome and a BLASTp search with *A. longisetosus* Hox protein sequences), we also identified Hox protein sequences in *D. pteronyssinus*, version 2 (*Waldron et al., 2017*).

Annotation of clock genes

Clock genes of *A. lycopersici* were identified by a tBLASTn search and reciprocal best BLASTp hit analysis (E-value threshold of E^{-10}) against the *A. lycopersici* genome and proteome, respectively, with *T. urticae* clock proteins (*Hoy et al., 2016*) as query.

Prediction of the *A. lycopersici* secretome

Signal peptides of *A. lycopersici* proteins were predicted with SignalP 5.0 and using default settings (*Almagro Armenteros et al., 2019*). Transmembrane domains were predicted using the Phobius server (*Käll et al., 2007*) at http://phobius.sbc.su.se/ and protein subcellular localization was predicted using WoLF PSORT (organism type: 'Animal') at https://wolfpsort.hgc.jp/. *A. lycopersici* proteins that, according to Phobius, did not have transmembrane regions outside the 60 amino acid N-terminal region, were predicted with a signal peptide by SignalP 5.0 and were predicted to be extracellular according to Wolf PSORT, were considered as putatively secreted proteins. Putatively secreted *A. lycopersici* proteins were used as query in a BLASTp search (with E-value threshold of E^{-10} and maximum target sequences set at 1) against the *T. urticae* proteome. Subsequently, *T. urticae* best BLASTp hits were mined for their presence in an LC-MS/MS analysis of *T. urticae* saliva (*Jonckheere et al., 2016*).

miRNA identification

Mature miRNA sequences for all available arthropod species were downloaded from Release 21 of miRbase (*Kozomara and Griffiths-Jones, 2014*). miRNA sequences were aligned using STAR 2.5.2b (*Dobin et al., 2013*) to the genome of *A. lycopersici* with the following parameters '`-alignIntronMax 0 –alignEndsType EndToEnd –outFilterMismatchNmax 2 –outFilterMultimapNmax 100`'; this ensured that all miRNA sequences that aligned had no more than two mismatches; alignments with insertions or deletions relative to the reference were removed from further consideration, and the resulting alignment file was sorted by position and indexed using SAMtools 1.3.1 (*Li et al., 2009*). Where miRNAs from different species aligned to the same position, they were denoted as being members of the same clusters (*Supplementary file 1* — 'Table S18' Tab).

Identification of genes in small RNA pathways

A tBLASTn search (with an E-value threshold of E^{-5}) using *T. castaneum* (*Prentice et al., 2015*; *Rodrigues et al., 2017*), *C. elegans* (*Hoy et al., 2016*), and *D. melanogaster* (*Iwasaki et al., 2015*) small RNA pathway-related protein sequences as query, was performed against the *A. lycopersici* genome to identify putative *A. lycopersici* small RNA pathway related genes that were not automatically predicted by the gene prediction software. As all tBLASTn hits showed overlap with existing gene models, no new gene models needed to be created. Next, a reciprocal best BLASTp hit analysis (with an E-value threshold of E^{-5}) was performed against the *A. lycopersici* and *T. urticae* proteome using *T. castaneum* (*Prentice et al., 2015*; *Rodrigues et al., 2017*), *C. elegans* (*Hoy et al., 2016*) and *D. melanogaster* (*Iwasaki et al., 2015*) small RNA pathway-related protein sequences and their available proteomes (*T. castaneum* version 5.2.36, *C. elegans* version WS262 and *D. melanogaster* FB2020_02 release) to identify putative small RNA pathway-related genes in *A. lycopersici* and *T. urticae*.

Genomic HGT screen and phylogenetic validation

We performed a genomic HGT screen as previously described in *Wybouw et al., 2018*. Briefly, the *A. lycopersici* proteome was aligned with metazoan and non-metazoan proteome databases and the bitscores of the best BLASTp hits were recorded. For each protein query, the *h*-index metric was calculated by subtracting the best metazoan bitscore from the best non-metazoan bitscore. An *A. lycopersici* gene was designated as a horizontally transferred gene candidate when it exhibited a best non-metazoan bitscore $\geq$75 and an *h*-index $\geq$30. In our screen, we also performed a tBLASTn-search against the tomato russet mite scaffolds using all identified horizontally transferred *T. urticae* genes as queries. Maximum-likelihood phylogenies were subsequently constructed for all *A. lycopersici* horizontally transferred gene candidates, except for a putative UGT pseudogene that was located on scaffold 5 between coordinates 140,638 and 140,871. All complete *A. lycopersici* UGT genes were sent to the UGT Nomenclature Committee to obtain unique UGT gene names (https://prime.vetmed.wsu.edu/resources/udp-glucuronsyltransferase-homepage). For the final phylogenetic reconstruction of the pantothenate biosynthetic genes, homologues of *aculy01g38350* (ketopantoate hydroxymethyltransferase, *panB*) and *aculy04g02470* (pantoate β-alanine ligase, *panC*) were identified by BLASTn and tBLASTn searches (E^{-10} cut-off) against the nonredundant nucleotide and protein NCBI databases, respectively, and were grouped based on their position in the tree of life (fungi, animals, bacteria, plants, and other). Proteins were selected per group based on manual inspection of the alignments and were combined with homologues as identified by *Wybouw et al., 2018*. In addition, we also added a panC homologue of the mealybug *Ferrisia virgata* to the final set of proteins (*Husnik and McCutcheon, 2016*). For the phylogenetic analysis of UGTs, we added UGT protein sequences from the annotated genome assembly of the house dust mite *D. pteronyssinus* (*Waldron et al., 2017*) to our UGT phylogenetic reconstruction. Applying an E-value of E^{-10} as the cut-off for the alignments, 27 *D. pteronyssinus* sequences were identified by reciprocal BLASTp-searches between the *D. pteronyssinus* proteome and the 87 *T. urticae* and *A. lycopersici* UGT sequences. Protein sequences were aligned using the online version of MAFFT v7.380 (*Katoh et al., 2019*) (available at https://mafft.cbrc.jp/alignment/software/) with 1000 iterations and the options 'E-INS-i' and 'reorder' (see *Supplementary file 7*). Protein models were selected based on the Akaike Information Criterion using ProtTest 3.4 (*Darriba et al., 2011*) (panB: LG+G, panC: LG+G, and UGT: LG+G+F). Maximum likelihood analyses were performed using RAxML v8 HPC2-XSEDE (*Stamatakis, 2014*) on the CIPRES Science Gateway (*Miller et al., 2010*) with 1000 rapid bootstrap replicates ('-f a -x 12345' option). An additional maximum likelihood tree reconstruction with ultra-fast bootstrapping (1000 replicates) was performed for the pantothenate biosynthetic proteins using IQ-TREE version 1.6.12 (*Hoang et al., 2018*; *Nguyen et al., 2015*). ModelFinder identified LG+I+G4 as the best protein model based on the Bayesian Information Criterion (*Kalyaanamoorthy et al., 2017*). Constrained tree tests for alternative topologies whereby *A. lycopersici* is the sister lineage to the spider mite pantothenate biosynthetic proteins were performed using the approximately unbiased test of IQ-TREE version 1.6.12 (10,000 RELL replicates) (*Shimodaira, 2002*). The random number seed was set at 12345. Last, the physical location of the *aculy01g38350* and a*culy04g02470* genes in the *A. lycopersici* genome was examined by PCR amplification. *A. lycopersici* mites were collected by soaking infested tomato leaves overnight in 40 mL of 70% ethanol. Mites in ethanol were centrifuged at 2000 rpm for 1 min, ethanol was removed, and pelleted mites were ground using liquid nitrogen. One mL of CTAB buffer with 2% beta-mercaptoethanol and 1% proteinase K was added to the ground mites, followed by incubation in a warm water bath at 56℃. Next, samples were washed with 1 ml of choloroform:isoamyl alcohol (21:1) and DNA was precipitated with isopropanol on ice for 1 hr. Primer sequences that successfully amplified genomic regions are listed in *Supplementary file 1* — 'Table S20' Tab. PCRs were performed using the recommended protocol for Phusion High Fidelity polymerase (Thermo Scientific, The Netherlands) and 1 µL of extracted DNA (50 ng/microL) and 0.2 µM of each primer. PCR conditions for fragment 1 and 3 were 98℃ for 30 min, followed by 35 cycles of denaturation at 98℃ for 10 s, annealing at 55℃ for 30 s, and extension at 72℃ for 1 min (fragment 1) or 45 s (fragment 3) followed by a final extension step at 72℃ for 5 min. PCR conditions for fragment two were as follows: 98℃ for 30 s, 5 cycles of 98℃ for 10 s, 65℃ for 10 s, 72℃ for 60 s, five cycles of 98℃ for 10 s, 60℃ for 10 s, 72℃ for 60 s, and 20 cycles of 98℃ for 10 s, 60℃ for 10 s, 72℃ for 60 s, followed by a final extension step at 72℃ for 3 min. Resulting amplicons were Sanger sequenced by Eurofins (Leiden, The Netherlands) using PCR (with 'PCR'

suffix) and sequencing (with 'seq' suffix) primers as indicated in *Supplementary file 1* — 'Table S20' Tab.

Acknowledgements

Additional information

Competing interests

Merijn R Kant: Reviewing editor, *eLife*. The other authors declare that no competing interests exist.

Funding

Author contributions

Robert Greenhalgh, Data curation, Formal analysis, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing; Wannes Dermauw, Conceptualization, Formal analysis, Investigation, Visualization, Methodology, Writing - original draft, Project administration, Writing - review and editing; Joris J Glas, Resources, Formal analysis, Investigation, Methodology; Stephane Rombauts, Data curation, Formal analysis, Methodology; Nicky Wybouw, Jainy Thomas, Ellen J

Pritham, René Feyereisen, Formal analysis, Investigation, Methodology; Juan M Alba, Resources, Investigation, Methodology; Saioa Legarrea, Resources, Investigation; Yves Van de Peer, Thomas Van Leeuwen, Conceptualization; Richard M Clark, Conceptualization, Supervision, Methodology, Writing - original draft, Project administration, Writing - review and editing; Merijn R Kant, Conceptualization, Resources, Funding acquisition, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs
Robert Greenhalgh http://orcid.org/0000-0003-2816-3154
Wannes Dermauw https://orcid.org/0000-0003-4612-8969
Joris J Glas https://orcid.org/0000-0002-8080-4564
Stephane Rombauts https://orcid.org/0000-0002-3985-4981
Nicky Wybouw https://orcid.org/0000-0001-7874-9765
Juan M Alba http://orcid.org/0000-0003-4822-9827
Saioa Legarrea http://orcid.org/0000-0002-9127-2794
René Feyereisen http://orcid.org/0000-0002-9560-571X
Yves Van de Peer http://orcid.org/0000-0003-4327-3730
Thomas Van Leeuwen https://orcid.org/0000-0003-4651-830X
Richard M Clark https://orcid.org/0000-0002-1470-301X
Merijn R Kant https://orcid.org/0000-0003-2524-8195

Decision letter and Author response
Decision letter https://doi.org/10.7554/eLife.56689.sa1
Author response https://doi.org/10.7554/eLife.56689.sa2

Additional files

Supplementary files
- Supplementary file 1. Supplementary Tables S1-20 as Tabs in a. xlsx file.
- Supplementary file 2. 2371 orthologous protein clusters used as input for Malin.
- Supplementary file 3. Sequence alignments and annotations of intron positions for *A. lycopersici*, *D. pteronyssinus*, *T. urticae*, *M. occidentalis*, *B. mori*, and *D. melanogaster* members of 80 orthogroups.
- Supplementary file 4. Small and large orthogroups used as input for CAFE analysis.
- Supplementary file 5. Ultrametric tree used as input for CAFE analysis.
- Supplementary file 6. Homeodomain regions of Hox protein sequences of *A. lycopersici*, *D. pteronyssinus*, *T. urticae*, *A. longisetosus*, *I. scapularis*, and *T. castaneum*.
- Supplementary file 7. Protein alignments used for phylogenetic tree construction in *Figures 2*, *4* and *6*, and the respective figure supplements.
- Transparent reporting form

Data availability
The genomic and 454 transcriptomic datasets generated by this project are available under BioProject accessions PRJNA588358 and PRJNA588365, respectively; the Illumina transcriptome data are available under BioProject accession PRJNA588358. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession WNKI00000000. The version described in this paper is version WNKI01000000. Additional datasets are hosted by the Online Resource for Community Annotation of Eukaryotes (ORCAE) at https://bioinformatics.psb.ugent.be/orcae/, where the annotation can be viewed and de novo transcriptomes (Illumina and 454) can be downloaded.

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Greenhalgh R, Dermauw W, Glas JJ, Rombauts S, Wybouw N, Thomas J, Alba JM, Pritham EJ, Legarrea S, Feyereisen R, Van de Peer Y, Van Leeuwen T, Clark RM, Kant MR	2020	Aculops lycopersici genome sequencing and assembly and Illumina transcriptome sequencing	http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA588358	NCBI BioProject, PRJNA588358
Greenhalgh R, Dermauw W, Glas JJ, Rombauts S, Wybouw N, Thomas J, Alba JM, Pritham EJ, Legarrea S, Feyereisen R, Van de Peer Y, Van Leeuwen T, Clark RM, Kant MR	2020	Aculops lycopersici Transcriptome or gene expression	http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA588365	NCBI BioProject, PRJNA588365
Greenhalgh R, Dermauw W, Glas JJ, Rombauts S, Wybouw N, Thomas J, Alba JM, Pritham EJ, Legarrea S, Feyereisen R, Van de Peer Y, Van Leeuwen T, Clark RM, Kant MR	2020	Aculops lycopersici, whole genome shotgun sequencing project	https://www.ncbi.nlm.nih.gov/nuccore/WNKI00000000.1/	NCBI Nucleotide, WNKI00000000

References

Abeel T, Van Parys T, Saeys Y, Galagan J, Van de Peer Y. 2012. GenomeView: a next-generation genome browser. *Nucleic Acids Research* **40**:e12. DOI: https://doi.org/10.1093/nar/gkr995, PMID: 22102585

Aboobaker AA, Blaxter ML. 2003. Hox gene loss during dynamic evolution of the nematode cluster. *Current Biology* **13**:37–40. DOI: https://doi.org/10.1016/S0960-9822(02)01399-4, PMID: 12526742

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**:2185–2195. DOI: https://doi.org/10.1126/science.287.5461.2185, PMID: 10731132

Ahn SJ, Dermauw W, Wybouw N, Heckel DG, Van Leeuwen T. 2014. Bacterial origin of a diverse family of UDP-glycosyltransferase genes in the *Tetranychus urticae* genome. *Insect Biochemistry and Molecular Biology* **50**: 43–57. DOI: https://doi.org/10.1016/j.ibmb.2014.04.003, PMID: 24727020

Al-Azzazy MM, Alhewairini SS. 2018. Relationship between temperature and developmental rate of tomato russet mite *Aculops lycopersici* (Massee) (Acari: eriophyideae) on tomato. *Journal of Food, Agriculture and Environment* **16**:18–23. DOI: https://doi.org/10.1234/4.2018.5531

Alba JM, Schimmel BC, Glas JJ, Ataide LM, Pappas ML, Villarroel CA, Schuurink RC, Sabelis MW, Kant MR. 2015. Spider mites suppress tomato defenses downstream of jasmonate and salicylate independently of hormonal crosstalk. *New Phytologist* **205**:828–840. DOI: https://doi.org/10.1111/nph.13075, PMID: 25297722

Alberti G, Crooker AR. 1985. 1.1.2 Internal anatomy. In: Helle W, Sabelis M. W (Eds). *Spider Mites: Their Biology, Natural Enemies, and Control, Volume 1A, World Crop Pests*. Amsterdam, The Netherlands: Elsevier. p. 29–58.

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* **37**: 420–423. DOI: https://doi.org/10.1038/s41587-019-0036-z, PMID: 30778233

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389–3402. DOI: https://doi.org/10.1093/nar/25.17.3389, PMID: 9254694

Anderson LD. 1954. The tomato russet mite in the united States1. *Journal of Economic Entomology* **47**:1001–1005. DOI: https://doi.org/10.1093/jee/47.6.1001

Arkhipova IR. 2018. Neutral theory, transposable elements, and eukaryotic genome evolution. *Molecular Biology and Evolution* **35**:1332–1337. DOI: https://doi.org/10.1093/molbev/msy083, PMID: 29688526

Aromolaran O, Beder T, Oswald M, Oyelade J, Adebiyi E, Koenig R. 2020. Essential gene prediction in *Drosophila melanogaster* using machine learning approaches based on sequence and functional features.

Computational and Structural Biotechnology Journal **18**:612–621. DOI: https://doi.org/10.1016/j.csbj.2020.02.022, PMID: 32257045

Arribas P, Andújar C, Moraza ML, Linard B, Emerson BC, Vogler AP. 2020. Mitochondrial metagenomics reveals the ancient origin and phylodiversity of soil mites and provides a phylogeny of the acari. *Molecular Biology and Evolution* **37**:683–694. DOI: https://doi.org/10.1093/molbev/msz255, PMID: 31670799

Bailey SF, Keifer HH. 1943. The tomato russet mite, *Phyllocoptes destructor* Keifer: Its Present Status. *Journal of Economic Entomology* **36**:706–712. DOI: https://doi.org/10.1093/jee/36.5.706

Barnett AA, Thomas RH. 2013. Posterior hox gene reduction in an arthropod: *Ultrabithorax* and *Abdominal-B* are expressed in a single segment in the mite *Archegozetes longisetosus*. *EvoDevo* **4**:23. DOI: https://doi.org/10.1186/2041-9139-4-23, PMID: 23991696

Ben-Shahar Y. 2011. Sensory functions for degenerin/epithelial sodium channels (DEG/ENaC). *Advances in Genetics* **76**:1–26. DOI: https://doi.org/10.1016/B978-0-12-386481-9.00001-8, PMID: 22099690

Bensoussan N, Santamaria ME, Zhurov V, Diaz I, Grbić M, Grbić V. 2016. Plant-Herbivore interaction: dissection of the cellular pattern of *Tetranychus urticae* Feeding on the Host Plant. *Frontiers in Plant Science* **7**:1105. DOI: https://doi.org/10.3389/fpls.2016.01105, PMID: 27512397

Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. 2009. Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in *Drosophila*. *Cell* **136**:149–162. DOI: https://doi.org/10.1016/j.cell.2008.12.001

Berenbaum MR. 2002. Postgenomic chemical ecology: from genetic code to ecological interactions. *Journal of Chemical Ecology* **28**:873–896. DOI: https://doi.org/10.1023/a:1015260931034, PMID: 12049229

Blaazer CJH, Villacis-Perez EA, Chafi R, Van Leeuwen T, Kant MR, Schimmel BCJ. 2018. Why do herbivorous mites suppress plant defenses? *Frontiers in Plant Science* **9**:1057. DOI: https://doi.org/10.3389/fpls.2018.01057, PMID: 30105039

Bodofsky S, Koitz F, Wightman B. 2017. Conserved and exapted functions of nuclear receptors in animal development. *Nuclear Receptor Research* **4**:101305. DOI: https://doi.org/10.11131/2017/101305, PMID: 29333434

Bohnsack MT, Czaplinski K, Gorlich D. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**:185–191. DOI: https://doi.org/10.1261/rna.5167604, PMID: 14730017

Bolton SJ, Chetverikov PE, Klompen H. 2017. Morphological support for a clade comprising two vermiform mite lineages: Eriophyoidea (Acariformes) and Nematalycidae (Acariformes) . *Systematic and Applied Acarology* **22**:1096–1131. DOI: https://doi.org/10.11158/saa.22.8.2

Bonatto Paese CL, Leite DJ, Schönauer A, McGregor AP, Russell S. 2018. Duplication and expression of sox genes in spiders. *BMC Evolutionary Biology* **18**:205. DOI: https://doi.org/10.1186/s12862-018-1337-4, PMID: 30587109

Bonneton F, Laudet V. 2012. Evolution of nuclear receptors in insects. *Insect Endocrinology* **1**:219–252. DOI: https://doi.org/10.1016/B978-0-12-384749-2.10006-8

Borsuk AM, Brodersen CR. 2019. The spatial distribution of chlorophyll in leaves. *Plant Physiology* **180**:1406–1417. DOI: https://doi.org/10.1104/pp.19.00094, PMID: 30944156

Buckles GR, Rauskolb C, Villano JL, Katz FN. 2001. *four-jointed* interacts with *Dachs*, *abelson* and *enabled* and feeds back onto the *notch* pathway to affect growth and segmentation in the *Drosophila* leg. *Development* **128**:3533–3542. PMID: 11566858

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421. DOI: https://doi.org/10.1186/1471-2105-10-421, PMID: 20003500

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973. DOI: https://doi.org/10.1093/bioinformatics/btp348, PMID: 19505945

Casola C, Koralewski TE. 2018. Pinaceae show elevated rates of gene turnover that are robust to incomplete gene annotation. *The Plant Journal* **95**:862–876. DOI: https://doi.org/10.1111/tpj.13994

Chan TF, Ji KM, Yim AK, Liu XY, Zhou JW, Li RQ, Yang KY, Li J, Li M, Law PT, Wu YL, Cai ZL, Qin H, Bao Y, Leung RK, Ng PK, Zou J, Zhong XJ, Ran PX, Zhong NS, et al. 2015. The draft genome, Transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. *Journal of Allergy and Clinical Immunology* **135**:539–548. DOI: https://doi.org/10.1016/j.jaci.2014.09.031, PMID: 25445830

Chapman B. 2016. bcbio-gff. *GitHub*. v0.6.4. https://github.com/chapmanb/bcbb/tree/master/gff

Chapman B, Chang J. 2000. Biopython: Python tools for computational biology. *ACM Sigbio Newsletter* **20**:15–19. DOI: https://doi.org/10.1145/360262.360268

Charoensawan V, Wilson D, Teichmann SA. 2010. Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Research* **38**:7364–7377. DOI: https://doi.org/10.1093/nar/gkq617, PMID: 20675356

Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, Stensmyr MC, Zheng Y, Liu W, Sun H, Xu Y, Luo Y, Kruse A, Yang X, Kontsedalov S, Lebedev G, Fisher TW, Nelson DR, Hunter WB, Brown JK, et al. 2016. The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biology* **14**:110. DOI: https://doi.org/10.1186/s12915-016-0321-y, PMID: 27974049

Chen CY, Liu YQ, Song WM, Chen DY, Chen FY, Chen XY, Chen ZW, Ge SX, Wang CZ, Zhan S, Chen XY, Mao YB. 2019. An effector from cotton bollworm oral secretion impairs host plant defense signaling. *PNAS* **116**:14331–14338. DOI: https://doi.org/10.1073/pnas.1905471116, PMID: 31221756

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WE, Wetter T, Suhai S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research* **14**:1147–1159. DOI: https://doi.org/10.1101/gr.1917404, PMID: 15140833

Chipman AD, Ferrier DE, Brena C, Qu J, Hughes DS, Schröder R, Torres-Oliva M, Znassi N, Jiang H, Almeida FC, Alonso CR, Apostolou Z, Aqrawi P, Arthur W, Barna JC, Blankenburg KP, Brites D, Capella-Gutiérrez S, Coyle M, Dearden PK, et al. 2014. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. *PLOS Biology* **12**:e1002005. DOI: https://doi.org/10.1371/journal.pbio.1002005, PMID: 25423365

Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, Berenbaum MR, Feyereisen R, Oakeshott JG. 2006. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Molecular Biology* **15**:615–636. DOI: https://doi.org/10.1111/j.1365-2583.2006.00672.x, PMID: 17069637

Coons LB. 1978. Fine structure of the digestive system of *Macrocheles muscaedomesticae* (scopoli) (acarina: Mesostigmata). *International Journal of Insect Morphology and Embryology* **7**:137–153. DOI: https://doi.org/10.1016/0020-7322(78)90013-2

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics* **10**:691–703. DOI: https://doi.org/10.1038/nrg2640, PMID: 19763152

Craig JP, Bekal S, Niblack T, Domier L, Lambert KN. 2009. Evidence for horizontally transferred genes involved in the biosynthesis of vitamin B1, B5, and B7 in *Heterodera glycines*. *Journal of Nematology* **41**:281–290. PMID: 22736827

Csurös M. 2008. Malin: maximum likelihood analysis of intron evolution in eukaryotes. *Bioinformatics* **24**:1538–1539. DOI: https://doi.org/10.1093/bioinformatics/btn226, PMID: 18474506

Damann N, Voets T, Nilius B. 2008. TRPs in our senses. *Current Biology : CB* **18**:R880–R889. DOI: https://doi.org/10.1016/j.cub.2008.07.063, PMID: 18812089

Danks HV. 2006. Short life cycles in insects and mites. *The Canadian Entomologist* **138**:407–463. DOI: https://doi.org/10.4039/n06-803

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**:1164–1165. DOI: https://doi.org/10.1093/bioinformatics/btr088, PMID: 21335321

de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research* **34**:W362–W365. DOI: https://doi.org/10.1093/nar/gkl124, PMID: 16845026

de la Paz Celorio-Mancera M, Wheat CW, Vogel H, Söderlind L, Janz N, Nylin S. 2013. Mechanisms of macroevolution: polyphagous plasticity in butterfly larvae revealed by RNA-Seq. *Molecular Ecology* **22**:4884–4895. DOI: https://doi.org/10.1111/mec.12440, PMID: 23952264

de Lillo E, Pozzebon A, Valenzano D, Duso C. 2018. An intimate relationship between eriophyoid mites and their host plants - A review. *Frontiers in Plant Science* **9**:1786. DOI: https://doi.org/10.3389/fpls.2018.01786, PMID: 30564261

De Lillo E, Monfreda R. 2004. "Salivary secretions" of eriophyoids (Acari: Eriophyoidea): first results of an experimental model. *Experimental & Applied Acarology* **34**:291–306. DOI: https://doi.org/10.1007/s10493-004-0267-6, PMID: 15651526

de Lillo E, Skoracka A. 2010. What's "cool" on eriophyoid mites? *Experimental & Applied Acarology* **51**:3–30. DOI: https://doi.org/10.1007/s10493-009-9297-4, PMID: 19760102

Dearden PK. 2015. Origin and evolution of the enhancer of split complex. *BMC Genomics* **16**:712. DOI: https://doi.org/10.1186/s12864-015-1926-1, PMID: 26384649

Dermauw W, Osborne EJ, Clark RM, Grbić M, Tirry L, Van Leeuwen T. 2013a. A burst of ABC genes in the genome of the polyphagous spider mite *Tetranychus urticae*. *BMC Genomics* **14**:317. DOI: https://doi.org/10.1186/1471-2164-14-317, PMID: 23663308

Dermauw W, Wybouw N, Rombauts S, Menten B, Vontas J, Grbic M, Clark RM, Feyereisen R, Van Leeuwen T. 2013b. A link between host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus urticae*. *PNAS* **110**:E113–E122. DOI: https://doi.org/10.1073/pnas.1213214110, PMID: 23248300

Dermauw W, Van Leeuwen T. 2014. The ABC gene family in arthropods: Comparative genomics and role in insecticide transport and resistance. *Insect Biochemistry and Molecular Biology* **45**:89–110. DOI: https://doi.org/10.1016/j.ibmb.2013.11.001, PMID: 24291285

Després L, David JP, Gallet C. 2007. The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology & Evolution* **22**:298–307. DOI: https://doi.org/10.1016/j.tree.2007.02.010, PMID: 17324485

Devictor V, Clavel J, Julliard R, Lavergne S, Mouillot D, Thuiller W, Venail P, Villéger S, Mouquet N. 2010. Defining and measuring ecological specialization. *Journal of Applied Ecology* **47**:15–25. DOI: https://doi.org/10.1111/j.1365-2664.2009.01744.x

Di Z, Yu Y, Wu Y, Hao P, He Y, Zhao H, Li Y, Zhao G, Li X, Li W, Cao Z. 2015. Genome-wide analysis of homeobox genes from *Mesobuthus martensii* reveals hox gene duplication in scorpions. *Insect Biochemistry and Molecular Biology* **61**:25–33. DOI: https://doi.org/10.1016/j.ibmb.2015.04.002, PMID: 25910680

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. DOI: https://doi.org/10.1093/bioinformatics/bts635, PMID: 23104886

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**:11–15.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**:1792–1797. DOI: https://doi.org/10.1093/nar/gkh340, PMID: 15034147

Elliott TA, Gregory TR. 2015. What's in a genome? the C-value enigma and the evolution of eukaryotic genome content. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**:20140331. DOI: https://doi.org/10.1098/rstb.2014.0331, PMID: 26323762

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**:157. DOI: https://doi.org/10.1186/s13059-015-0721-2, PMID: 26243257

Enjin A, Zaharieva EE, Frank DD, Mansourian S, Suh GS, Gallio M, Stensmyr MC. 2016. Humidity sensing in *Drosophila*. *Current Biology* **26**:1352–1358. DOI: https://doi.org/10.1016/j.cub.2016.03.049, PMID: 27161501

Erb M, Reymond P. 2019. Molecular interactions between plants and insect herbivores. *Annual Review of Plant Biology* **70**:527–557. DOI: https://doi.org/10.1146/annurev-arplant-050718-095910, PMID: 30786233

Fahrbach SE, Smagghe G, Velarde RA. 2012. Insect nuclear receptors. *Annual Review of Entomology* **57**:83–106. DOI: https://doi.org/10.1146/annurev-ento-120710-100607, PMID: 22017307

Feschotte C, Keswani U, Ranganathan N, Guibotsy ML, Levine D. 2009. Exploring repetitive DNA landscapes using REPCLASS, a tool that automates the classification of transposable elements in eukaryotic genomes. *Genome Biology and Evolution* **1**:205–220. DOI: https://doi.org/10.1093/gbe/evp023, PMID: 20333191

Feyereisen R. 2011. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. *Biochimica Et Biophysica Acta (BBA) - Proteins and Proteomics* **1814**:19–28. DOI: https://doi.org/10.1016/j.bbapap.2010.06.012, PMID: 20601227

Feyereisen R. 2012. Insect CYP genes and P450 EnzymesInsect. *Molecular Biology and Biochemistry* **2012**:236–316. DOI: https://doi.org/10.1016/B978-0-12-384747-8.10008-X

Futuyma DJ, Moreno G. 1988. The evolution of ecological specialization. *Annual Review of Ecology and Systematics* **19**:207–233. DOI: https://doi.org/10.1146/annurev.es.19.110188.001231

García-Bellido A, de Celis JF. 2009. The complex tale of the *achaete-scute* complex: a paradigmatic case in the analysis of gene organization and function during development. *Genetics* **182**:631–639. DOI: https://doi.org/10.1534/genetics.109.104083, PMID: 19622761

Gel B, Serra E. 2017. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**:3088–3090. DOI: https://doi.org/10.1093/bioinformatics/btx346, PMID: 28575171

Gerson U, Weintraub PG. 2012. Mites (Acari) as a factor in greenhouse management. *Annual Review of Entomology* **57**:229–247. DOI: https://doi.org/10.1146/annurev-ento-120710-100639, PMID: 21910634

Glas JJ, Alba JM, Simoni S, Villarroel CA, Stoops M, Schimmel BC, Schuurink RC, Sabelis MW, Kant MR. 2014. Defense suppression benefits herbivores that have a monopoly on their feeding site but can backfire within natural communities. *BMC Biology* **12**:98. DOI: https://doi.org/10.1186/s12915-014-0098-9, PMID: 25403155

Gloss AD, Abbot P, Whiteman NK. 2019. How interactions with plant chemicals shape insect genomes. *Current Opinion in Insect Science* **36**:149–156. DOI: https://doi.org/10.1016/j.cois.2019.09.005, PMID: 31698152

Govind G, Mittapalli O, Griebel T, Allmann S, Böcker S, Baldwin IT. 2010. Unbiased transcriptional comparisons of generalist and specialist herbivores feeding on progressively defenseless *Nicotiana attenuata* plants. *PLOS ONE* **5**:e8735. DOI: https://doi.org/10.1371/journal.pone.0008735, PMID: 20090945

Gramates LS, Marygold SJ, Santos GD, Urbano JM, Antonazzo G, Matthews BB, Rey AJ, Tabone CJ, Crosby MA, Emmert DB, Falls K, Goodman JL, Hu Y, Ponting L, Schroeder AJ, Strelets VB, Thurmond J, Zhou P, The FlyBase Consortium. 2017. FlyBase at 25: looking to the future. *Nucleic Acids Research* **45**:D663–D671. DOI: https://doi.org/10.1093/nar/gkw1016, PMID: 27799470

Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, Osborne EJ, Dermauw W, Ngoc PC, Ortego F, Hernández-Crespo P, Diaz I, Martinez M, Navajas M, Sucena É, Magalhães S, Nagy L, Pace RM, Djuranović S, Smagghe G, et al. 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* **479**:487–492. DOI: https://doi.org/10.1038/nature10640, PMID: 22113690

Gregory TR, Young MR. 2020. Small genomes in most mites (but not ticks). *International Journal of Acarology* **46**:1–8. DOI: https://doi.org/10.1080/01647954.2019.1684561

Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **10**:645–656. DOI: https://doi.org/10.1109/TCBB.2013.68, PMID: 24091398

Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, Sattelle DB, de la Fuente J, Ribeiro JM, Megy K, Thimmapuram J, Miller JR, Walenz BP, Koren S, Hostetler JB, Thiagarajan M, Joardar VS, Hannick LI, Bidwell S, Hammond MP, et al. 2016. Genomic insights into the *Ixodes scapularis* tick vector of lyme disease. *Nature Communications* **7**:10507. DOI: https://doi.org/10.1038/ncomms10507, PMID: 26856261

Gupta AK, Scully ED, Palmer NA, Geib SM, Sarath G, Hein GL, Tatineni S. 2019. Wheat streak mosaic virus alters the transcriptome of its vector, wheat curl mite (*Aceria tosichella* keifer), to enhance mite development and population expansion. *Journal of General Virology* **100**:889–910. DOI: https://doi.org/10.1099/jgv.0.001256, PMID: 31017568

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protocols* **8**:1494–1512. DOI: https://doi.org/10.1038/nprot.2013.084, PMID: 23845962

Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution* **30**:1987–1997. DOI: https://doi.org/10.1093/molbev/mst100, PMID: 23709260

Heckel DG. 2014. Insect Detoxification and Sequestration StrategiesInsect-Plant Interactions. In: Voelckel C, Jander G (Eds). *Annual Plant Reviews*. Wiley Blackwell. p. 77–114. DOI: https://doi.org/10.1002/9781118829783.ch3

Helle W, Wysoki M. 1983. The chromosomes and sex-determination of some actinotrichid taxa (Acari), with special reference to Eriophyidae. *International Journal of Acarology* **9**:67–71. DOI: https://doi.org/10.1080/01647958308683315

Helle W, Wysoki M. 1996. 1.3.2 Arrhenotokous parthenogenesis. In: Lindquist E. E, Sabelis M. W, Bruin J (Eds). *Eriophyoid Mites – Their Biology, Natural Enemies and Control, World Crop Pests*. Elsevier. p. 169–172. DOI: https://doi.org/10.1016/S1572-4379(96)80008-X

Hessen DO, Jeyasingh PD, Neiman M, Weider LJ. 2010a. Genome streamlining and the elemental costs of growth. *Trends in Ecology & Evolution* **25**:75–80. DOI: https://doi.org/10.1016/j.tree.2009.08.004, PMID: 19796842

Hessen DO, Jeyasingh PD, Neiman M, Weider LJ. 2010b. Genome streamlining in prokaryotes versus eukaryotes. *Trends in Ecology & Evolution* **25**:320–321. DOI: https://doi.org/10.1016/j.tree.2010.03.003

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**:518–522. DOI: https://doi.org/10.1093/molbev/msx281, PMID: 29077904

Holland PW. 2013. Evolution of homeobox genes. *Wiley Interdisciplinary Reviews: Developmental Biology* **2**:31–45. DOI: https://doi.org/10.1002/wdev.78, PMID: 23799629

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, et al. 2002. The genome sequence of the malaria mosquito *anopheles gambiae*. *Science* **298**:129–149. DOI: https://doi.org/10.1126/science.1076181, PMID: 12364791

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**:498–503. DOI: https://doi.org/10.1038/nature12111, PMID: 23594743

Howe GA, Jander G. 2008. Plant immunity to insect herbivores. *Annual Review of Plant Biology* **59**:41–66. DOI: https://doi.org/10.1146/annurev.arplant.59.032607.092825, PMID: 18031220

Hoy MA. 2004. Four-Legged Mites (Eriophyoidea or Tetrapodili). In: Capinera J. L (Ed). *Encyclopedia of Entomology*. Springer. p. 95–262. DOI: https://doi.org/10.1007/978-1-4020-6359-6_3883

Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, Pomerantz AF, Simão FA, Thomas J, Jiggins FM, Murphy TD, Pritham EJ, Robertson HM, Zdobnov EM, Gibbs RA, Richards S. 2016. Genome sequencing of the phytoseiid predatory mite *Metaseiulus occidentalis* Reveals Completely Atomized *Hox* Genes and Superdynamic Intron Evolution. *Genome Biology and Evolution* **8**:1762–1775. DOI: https://doi.org/10.1093/gbe/evw048, PMID: 26951779

Huang S, Gao Y, Liu J, Peng X, Niu X, Fei Z, Cao S, Liu Y. 2012. Genome-wide analysis of WRKY transcription factors in *Solanum lycopersicum*. *Molecular Genetics and Genomics* **287**:495–513. DOI: https://doi.org/10.1007/s00438-012-0696-6, PMID: 22570076

Huang H, Li Y, Szulwach KE, Zhang G, Jin P, Chen D. 2014. AGO3 slicer activity regulates mitochondria-nuage localization of armitage and piRNA amplification. *Journal of Cell Biology* **206**:217–230. DOI: https://doi.org/10.1083/jcb.201401002, PMID: 25049272

Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Research* **9**:868–877. DOI: https://doi.org/10.1101/gr.9.9.868, PMID: 10508846

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754–755. DOI: https://doi.org/10.1093/bioinformatics/17.8.754, PMID: 11524383

Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution* **33**:1635–1638. DOI: https://doi.org/10.1093/molbev/msw046, PMID: 26921390

Hughes CL, Kaufman TC. 2002. Hox genes and the evolution of the arthropod body plan. *Evolution and Development* **4**:459–499. DOI: https://doi.org/10.1046/j.1525-142X.2002.02034.x, PMID: 12492146

Husnik F, McCutcheon JP. 2016. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *PNAS* **113**:E5416–E5424. DOI: https://doi.org/10.1073/pnas.1603910113, PMID: 27573819

Hwang DS, Lee BY, Kim HS, Lee MC, Kyung DH, Om AS, Rhee JS, Lee JS. 2014. Genome-wide identification of nuclear receptor (NR) superfamily genes in the copepod *Tigriopus japonicus*. *BMC Genomics* **15**:993. DOI: https://doi.org/10.1186/1471-2164-15-993, PMID: 25407996

Iso T, Kedes L, Hamamori Y. 2003. HES and HERP families: multiple effectors of the notch signaling pathway. *Journal of Cellular Physiology* **194**:237–255. DOI: https://doi.org/10.1002/jcp.10208, PMID: 12548545

Iwasaki YW, Siomi MC, Siomi H. 2015. PIWI-Interacting RNA: its biogenesis and functions. *Annual Review of Biochemistry* **84**:405–433. DOI: https://doi.org/10.1146/annurev-biochem-060614-034258, PMID: 25747396

Janssen R, Andersson E, Betnér E, Bijl S, Fowler W, Höök L, Leyhr J, Mannelqvist A, Panara V, Smith K, Tiemann S. 2018. Embryonic expression patterns and phylogenetic analysis of panarthropod sox genes: insight into nervous system development, segmentation and gonadogenesis. *BMC Evolutionary Biology* **18**:88. DOI: https://doi.org/10.1186/s12862-018-1196-z, PMID: 29884143

Jeppson LR, Keifer H, Baker EW. 1975. *Mites Injurious to Economic Plants*. Univ of California Press.

Joga MR, Zotti MJ, Smagghe G, Christiaens O. 2016. RNAi efficiency, systemic properties, and novel delivery methods for pest insect control: what we know so far. *Frontiers in Physiology* **7**:553. DOI: https://doi.org/10.3389/fphys.2016.00553, PMID: 27909411

Jonckheere W, Dermauw W, Zhurov V, Wybouw N, Van den Bulcke J, Villarroel CA, Greenhalgh R, Grbić M, Schuurink RC, Tirry L, Baggerman G, Clark RM, Kant MR, Vanholme B, Menschaert G, Van Leeuwen T. 2016. The salivary protein repertoire of the polyphagous spider mite *Tetranychus urticae*: A Quest for Effectors. *Molecular & Cellular Proteomics* **15**:3594–3613. DOI: https://doi.org/10.1074/mcp.M116.058081, PMID: 27703040

Joseph RM, Carlson JR. 2015. *Drosophila* chemoreceptors: a molecular interface between the chemical world and the brain. *Trends in Genetics* **31**:683–695. DOI: https://doi.org/10.1016/j.tig.2015.09.005, PMID: 26477743

Käll L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction–the phobius web server. *Nucleic Acids Research* **35**:W429–W432. DOI: https://doi.org/10.1093/nar/gkm256, PMID: 17483518

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**:587–589. DOI: https://doi.org/10.1038/nmeth.4285, PMID: 28481363

Kant MR, Sabelis MW, Haring MA, Schuurink RC. 2008. Intraspecific variation in a generalist herbivore accounts for differential induction and impact of host plant defences. *Proceedings of the Royal Society B: Biological Sciences* **275**:443–452. DOI: https://doi.org/10.1098/rspb.2007.1277, PMID: 18055390

Kant MR, Jonckheere W, Knegt B, Lemos F, Liu J, Schimmel BC, Villarroel CA, Ataide LM, Dermauw W, Glas JJ, Egas M, Janssen A, Van Leeuwen T, Schuurink RC, Sabelis MW, Alba JM. 2015. Mechanisms and ecological consequences of plant defence induction and suppression in herbivore communities. *Annals of Botany* **115**:1015–1051. DOI: https://doi.org/10.1093/aob/mcv054, PMID: 26019168

Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. *PNAS* **114**:E1460–E1469. DOI: https://doi.org/10.1073/pnas.1616702114, PMID: 28179571

Kapusta A. 2017. Parsing-RepeatMasker-Outputs, parseRM. *GitHub*. v5.8.2. https://github.com/4ureliek/Parsing-RepeatMasker-Outputs

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**:1160–1166. DOI: https://doi.org/10.1093/bib/bbx108, PMID: 28968734

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**:772–780. DOI: https://doi.org/10.1093/molbev/mst010, PMID: 23329690

Kawai A, Haque MM. 2004. Population dynamics of tomato russet mite, *aculops lycopersici* (Massee) and its natural enemy, *homeopronematus anconai* (Baker). *Japan Agricultural Research Quarterly: JARQ* **38**:161–166. DOI: https://doi.org/10.6090/jarq.38.161

Keifer HH. 1946. A review of north american economic eriophyid Mites1. *Journal of Economic Entomology* **39**:563–570. DOI: https://doi.org/10.1093/jee/39.5.563

Kenny NJ, Chan KW, Nong W, Qu Z, Maeso I, Yip HY, Chan TF, Kwan HS, Holland PW, Chu KH, Hui JH. 2016. Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity* **116**:190–199. DOI: https://doi.org/10.1038/hdy.2015.89, PMID: 26419336

Kent WJ. 2002. BLAT–the BLAST-like alignment tool. *Genome Research* **12**:656–664. DOI: https://doi.org/10.1101/gr.229202, PMID: 11932250

Kim VN. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Reviews Molecular Cell Biology* **6**:376–385. DOI: https://doi.org/10.1038/nrm1644, PMID: 15852042

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**:R36. DOI: https://doi.org/10.1186/gb-2013-14-4-r36, PMID: 23618408

Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JM, Ribeiro JM, Rozas J, et al. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *PNAS* **107**:12168–12173. DOI: https://doi.org/10.1073/pnas.1003379107, PMID: 20566863

Klimov PB, OConnor BM, Chetverikov PE, Bolton SJ, Pepato AR, Mortazavi AL, Tolstikov AV, Bauchan GR, Ochoa R. 2018. Comprehensive phylogeny of acariform mites (Acariformes) provides insights on the origin of the four-legged mites (Eriophyoidea), a long branch. *Molecular Phylogenetics and Evolution* **119**:105–117. DOI: https://doi.org/10.1016/j.ympev.2017.10.017, PMID: 29074461

Koroleva OA, Tomos AD, Farrar J, Roberts P, Pollock CJ. 2000. Tissue distribution of primary metabolism between epidermal, mesophyll and parenchymatous bundle sheath cells in barley leaves. *Functional Plant Biology* **27**:747–755. DOI: https://doi.org/10.1071/PP99156

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* **42**:D68–D73. DOI: https://doi.org/10.1093/nar/gkt1181, PMID: 24275495

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**:1870–1874. DOI: https://doi.org/10.1093/molbev/msw054, PMID: 27004904

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution* **34**:1812–1819. DOI: https://doi.org/10.1093/molbev/msx116, PMID: 28387841

Laetsch DR, Blaxter ML. 2017. KinFin: software for Taxon-Aware analysis of clustered protein sequences. *G3: Genes, Genomes, Genetics* **7**:3349–3357. DOI: https://doi.org/10.1534/g3.117.300233, PMID: 28866640

Laland K, Matthews B, Feldman MW. 2016. An introduction to niche construction theory. *Evolutionary Ecology* **30**:191–202. DOI: https://doi.org/10.1007/s10682-016-9821-z, PMID: 27429507

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921. DOI: https://doi.org/10.1038/35057062, PMID: 11237011

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nature Methods* **9**:357–359. DOI: https://doi.org/10.1038/nmeth.1923, PMID: 22388286

Lee C, Bae K, Edery I. 1999. PER and TIM inhibit the DNA binding activity of a *Drosophila* CLOCK-CYC/dBMAL1 heterodimer without disrupting formation of the heterodimer: a basis for circadian transcription. *Molecular and Cellular Biology* **19**:5316–5325. DOI: https://doi.org/10.1128/MCB.19.8.5316, PMID: 10409723

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. DOI: https://doi.org/10.1093/bioinformatics/btp352, PMID: 19505943

Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research* **47**:W199–W205. DOI: https://doi.org/10.1093/nar/gkz401, PMID: 31114916

Lindquist EE, Oldfield GN. 1996. Chapter 1.5 Evolution and phylogeny 1.5.1 Evolution of eriophyoid mites in relation to their host plants. In: Lindquist E. E, Sabelis M. W, Bruin J (Eds). *Eriophyoid Mites: Their Biology, Natural Enemies and Control, World Crop Pests*. Elsevier. p. 277–300. DOI: https://doi.org/10.1016/S1572-4379(96)80018-2

Litoff EJ, Garriott TE, Ginjupalli GK, Butler L, Gay C, Scott K, Baldwin WS. 2014. Annotation of the *daphnia magna* nuclear receptors: comparison to *daphnia pulex*. *Gene* **552**:116–125. DOI: https://doi.org/10.1016/j.gene.2014.09.024, PMID: 25239664

Lynch M, Bobay LM, Catania F, Gout JF, Rho M. 2011. The repatterning of eukaryotic genomes by random genetic drift. *Annual Review of Genomics and Human Genetics* **12**:347–366. DOI: https://doi.org/10.1146/annurev-genom-082410-101412, PMID: 21756106

Maderspacher F. 2016. Zoology: the walking heads. *Current Biology* **26**:R194–R197. DOI: https://doi.org/10.1016/j.cub.2016.01.060, PMID: 26954437

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**:764–770. DOI: https://doi.org/10.1093/bioinformatics/btr011, PMID: 21217122

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380. DOI: https://doi.org/10.1038/nature03959, PMID: 16056220

Massee AM. 1937. An eriophyid mite injurious to tomato. *Bulletin of Entomological Research* **28**:403. DOI: https://doi.org/10.1017/S0007485300038864

Mathieson BR, Lehane MJ. 2002. Ultrastructure of the alimentary canal of the sheep scab mite, *Psoroptes ovis* (Acari: Psoroptidae). *Veterinary Parasitology* **104**:151–166. DOI: https://doi.org/10.1016/S0304-4017(01)00620-3, PMID: 11809334

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), 2010. IEEE 1–8. DOI: https://doi.org/10.1109/GCE.2010.5676129

Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-Okuda K, Yamamoto K, Ajimura M, Ravikumar G, Shimomura M, Nagamura Y, Shin-I T, Abe H, Shimada T, Morishita S, et al. 2004. The genome sequence of silkworm, *bombyx mori*. *DNA Research* **11**:27–35. DOI: https://doi.org/10.1093/dnares/11.1.27, PMID: 15141943

Moerkens R, Vanlommel W, Reybroeck E, Wittemans L, De Clercq P, Van Leeuwen T, De Vis R. 2018. Binomial sampling plan for tomato russet mite (*Aculopslycopersici* (Tryon) (Acari: Eriophyidae) in protected tomato crops . *Journal of Applied Entomology* **142**:820–827. DOI: https://doi.org/10.1111/jen.12529

Mondal M, Klimov P, Flynt AS. 2018a. Rewired RNAi-mediated genome surveillance in house dust mites. *PLOS Genetics* **14**:e1007183. DOI: https://doi.org/10.1371/journal.pgen.1007183, PMID: 29377900

Mondal M, Mansfield K, Flynt A. 2018b. siRNAs and piRNAs collaborate for transposon control in the two-spotted spider mite. *RNA* **24**:899–907. DOI: https://doi.org/10.1261/rna.065839.118, PMID: 29678924

Mourier T, Jeffares DC. 2003. Eukaryotic intron loss. *Science* **300**:1393. DOI: https://doi.org/10.1126/science.1080559, PMID: 12775832

Musser RO, Hum-Musser SM, Eichenseer H, Peiffer M, Ervin G, Murphy JB, Felton GW. 2002. Herbivory: caterpillar saliva beats plant defences. *Nature* **416**:599–600. DOI: https://doi.org/10.1038/416599a, PMID: 11948341

Navia D, Ochoa R, Welbourn C, Ferragut F. 2010. Adventive eriophyoid mites: a global review of their impact, pathways, prevention and challenges. *Experimental and Applied Acarology* **51**:225–255. DOI: https://doi.org/10.1007/s10493-009-9327-2, PMID: 19844795

Navia D, de Mendonça RS, Skoracka A, Szydło W, Knihinicki D, Hein GL, da Silva Pereira PR, Truol G, Lau D. 2013. Wheat curl mite, *Aceria Tosichella*, and transmitted viruses: an expanding pest complex affecting cereal crops. *Experimental and Applied Acarology* **59**:95–143. DOI: https://doi.org/10.1007/s10493-012-9633-y, PMID: 23179064

Ngoc PC, Greenhalgh R, Dermauw W, Rombauts S, Bajda S, Zhurov V, Grbić M, Van de Peer Y, Van Leeuwen T, Rouzé P, Clark RM. 2016. Complex evolutionary dynamics of massively expanded chemosensory receptor families in an extreme generalist chelicerate herbivore. *Genome Biology and Evolution* **8**:3323–3339. DOI: https://doi.org/10.1093/gbe/evw249, PMID: 27797949

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**:268–274. DOI: https://doi.org/10.1093/molbev/msu300, PMID: 25371430

Nuzzaci G, Alberti G. 1996. Chapter 1.2 Internal anatomy and physiology. In: Lindquist E. E, Sabelis M. W, Bruin J (Eds). *Eriophyoid Mites – Their Biology, Natural Enemies and Control, World Crop Pests*. Elsevier. p. 101–150. DOI: https://doi.org/10.1016/S1572-4379(96)80006-6

Okamura K, Robine N, Liu Y, Liu Q, Lai EC. 2011. R2D2 organizes small regulatory RNA pathways in *Drosophila*. *Molecular and Cellular Biology* **31**:884–896. DOI: https://doi.org/10.1128/MCB.01141-10, PMID: 21135122

Oldfield GN. 1996. 1.4.3 Diversity and host plant specificity. In: Lindquist E. E, Sabelis M. W, Bruin J (Eds). *Eriophyoid Mites – Their Biology, Natural Enemies and Control, World Crop Pests*. Elsevier. p. 199–216. DOI: https://doi.org/10.1016/S1572-4379(96)80011-X

Oldfield GN, Michalska K. 1996. 1.4.2 Spermatophore Deposition, Mating Behavior and Population Mating Structure. In: Lindquist E. E, Sabelis M. W, Bruin J (Eds). *Eriophyoid Mites: Their Biology, Natural Enemies and Control, World Crop Pests*. Elsevier. p. 185–198. DOI: https://doi.org/10.1016/S1572-4379(96)80010-8

Pace RM, Grbić M, Nagy LM. 2016. Composition and genomic organization of arthropod hox clusters. *EvoDevo* **7**:11. DOI: https://doi.org/10.1186/s13227-016-0048-4, PMID: 27168931

Pao SS, Paulsen IT, Saier MH. 1998. Major facilitator superfamily. *Microbiology and Molecular Biology Reviews* **62**:1–34. DOI: https://doi.org/10.1128/MMBR.62.1.1-34.1998, PMID: 9529885

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**:1061–1067. DOI: https://doi.org/10.1093/bioinformatics/btm071, PMID: 17332020

Peñalva-Arana DC, Lynch M, Robertson HM. 2009. The chemoreceptor genes of the waterflea *daphnia pulex*: many Grs but no Ors. *BMC Evolutionary Biology* **9**:79. DOI: https://doi.org/10.1186/1471-2148-9-79, PMID: 19383158

Peng G, Shi X, Kadowaki T. 2015. Evolution of TRP channels inferred by their classification in diverse animal species. *Molecular Phylogenetics and Evolution* **84**:145–157. DOI: https://doi.org/10.1016/j.ympev.2014.06.016, PMID: 24981559

Perring TM. 1996. 3.2.7 Vegetables. In: Lindquist E. E, Sabelis M. W, Bruin J (Eds). *Eriophyoid Mites: Their Biology, Natural Enemies and Control, World Crop Pests*. Elsevier. p. 593–610. DOI: https://doi.org/10.1016/S1572-4379(96)80038-8

Perring TM, Farrar CA. 1986. *Historical Perspective and Current World Status of the Tomato Russet Mite (Acari: Eriophyidae)*: Entomological Soc America.

Perring TM, Royalty RN. 1996. Nature of Damage and its Assessment. In: Lindquist E. E, Sabelis M. W, Bruin J (Ed). *Eriophyoid Mites – Their Biology, Natural Enemies and Control, World Crop Pests*. Elsevier. p. 493–513. DOI: https://doi.org/10.1016/S1572-4379(96)80031-5

Peschel N, Helfrich-Förster C. 2011. Setting the clock–by nature: circadian rhythm in the fruitfly *Drosophila melanogaster*. *FEBS Letters* **585**:1435–1442. DOI: https://doi.org/10.1016/j.febslet.2011.02.028, PMID: 21354415

Pflugfelder GO, Eichinger F, Shen J. 2017. T-Box genes in *Drosophila* limb development. *Current Topics in Developmental Biology* **122**:313–354. DOI: https://doi.org/10.1016/bs.ctdb.2016.08.003, PMID: 28057269

Polilov AA. 2015. Small is beautiful: features of the smallest insects and limits to miniaturization. *Annual Review of Entomology* **60**:103–121. DOI: https://doi.org/10.1146/annurev-ento-010814-020924, PMID: 25341106

Prentice K, Pertry I, Christiaens O, Bauters L, Bailey A, Niblett C, Ghislain M, Gheysen G, Smagghe G. 2015. Transcriptome analysis and systemic RNAi response in the african sweetpotato weevil (*Cylas puncticollis*, Coleoptera, brentidae). *PLOS ONE* **10**:e0115336. DOI: https://doi.org/10.1371/journal.pone.0115336, PMID: 25590333

Price AL, Jones NC, Pevzner PA. 2005. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21 Suppl 1**:i351–i358. DOI: https://doi.org/10.1093/bioinformatics/bti1018, PMID: 15961478

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Research* **33**:W116–W120. DOI: https://doi.org/10.1093/nar/gki442, PMID: 15980438

R Development Core Team. 2018. R: A language and environment for statistical computing . Vienna, Austria, R Foundation for Statistical Computing. http://www.r-project.org

Rane RV, Ghodke AB, Hoffmann AA, Edwards OR, Walsh TK, Oakeshott JG. 2019. Detoxifying enzyme complements and host use phenotypes in 160 insect species. *Current Opinion in Insect Science* **31**:131–138. DOI: https://doi.org/10.1016/j.cois.2018.12.008, PMID: 31109666

Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**:1079–1083. DOI: https://doi.org/10.1038/nature08742, PMID: 20147900

Rehman S, Gupta VK, Goyal AK. 2016. Identification and functional analysis of secreted effectors from phytoparasitic Nematodes. *BMC Microbiology* **16**:48. DOI: https://doi.org/10.1186/s12866-016-0632-8, PMID: 27001199

Ren Z, Veksler-Lublinsky I, Morrissey D, Ambros V. 2016. Staufen negatively modulates MicroRNA activity in *Caenorhabditis elegans*. *G3: Genes, Genomes, Genetics* **6**:1227–1237. DOI: https://doi.org/10.1534/g3.116.027300, PMID: 26921297

Ren FR, Sun X, Wang TY, Yao YL, Huang YZ, Zhang X, Luan JB. 2020. Biotin provisioning by horizontally transferred genes from Bacteria confers animal fitness benefits. *The ISME Journal* **14**:2542–2553. DOI: https://doi.org/10.1038/s41396-020-0704-5, PMID: 32572143

Rice RE, Strong FE. 1962. Bionomics of the tomato russet mite, *vasates lycopersici* (Massee)1. *Annals of the Entomological Society of America* **55**:431–435. DOI: https://doi.org/10.1093/aesa/55.4.431

Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, Gibbs R, Beeman RW, Brown SJ, Bucher G, Friedrich M, Grimmelikhuijzen CJ, Klingler M, Lorenzen M, Richards S, Roth S, Schröder R, Tautz D, Zdobnov EM, Muzny D, et al. 2008. The genome of the model beetle and pest *tribolium castaneum*. *Nature* **452**:949–955. DOI: https://doi.org/10.1038/nature06784, PMID: 18362917

Rider SD, Morgan MS, Arlian LG. 2015. Draft genome of the scabies mite. *Parasites & Vectors* **8**:585. DOI: https://doi.org/10.1186/s13071-015-1198-2, PMID: 26555130

Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *PNAS* **100 Suppl 2**:14537–14542. DOI: https://doi.org/10.1073/pnas.2335847100, PMID: 14608037

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature Biotechnology* **29**:24–26. DOI: https://doi.org/10.1038/nbt.1754, PMID: 21221095

Rodrigues TB, Dhandapani RK, Duan JJ, Palli SR. 2017. RNA interference in the Asian Longhorned beetle: identification of key RNAi genes and reference genes for RT-qPCR. *Scientific Reports* **7**:8913. DOI: https://doi.org/10.1038/s41598-017-08813-1, PMID: 28827780

Roy SW, Gilbert W. 2005. The pattern of intron loss. *PNAS* **102**:713–718. DOI: https://doi.org/10.1073/pnas.0408274102, PMID: 15642949

Royalty RN, Perring TM. 1988. Morphological analysis of damage to tomato leaflets by tomato russet mite (Acari: eriophyidae). *Journal of Economic Entomology* **81**:816–820. DOI: https://doi.org/10.1093/jee/81.3.816

Rytz R, Croset V, Benton R. 2013. Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in *Drosophila* and beyond. *Insect Biochemistry and Molecular Biology* **43**:888–897. DOI: https://doi.org/10.1016/j.ibmb.2013.02.007, PMID: 23459169

Sabelis MW, Bruin J. 1996. 1.5.3 Evolutionary Ecology: Life history patterns, food plant choice and dispersal. In: Lindquist E. E, Sabelis M. W, Bruin J (Eds). *Eriophyoid Mites – Their Biology, Natural Enemies and Control, World Crop Pests*. Elsevier. p. 329–366. DOI: https://doi.org/10.1016/S1572-4379(96)80020-0

Schaub C, Frasch M. 2013. Org-1 is required for the diversification of circular visceral muscle founder cells and normal midgut morphogenesis. *Developmental Biology* **376**:245–259. DOI: https://doi.org/10.1016/j.ydbio.2013.01.022, PMID: 23380635

Schiex T, Moisan A, Rouzé P. 2001. Eugène: An Eukaryotic Gene Finder That Combines Several Sources of Evidence. In: Gascuel O, Sagot M. F (Eds). *Computational Biology. JOBIM 2000. Lecture Notes in Computer Science*. **2066** Berlin, Heidelberg: Springer. p. 111–125. DOI: https://doi.org/10.1007/3-540-45727-5_10

Schimmel BCJ, Ataide LMS, Chafi R, Villarroel CA, Alba JM, Schuurink RC, Kant MR. 2017. Overcompensation of herbivore reproduction through hyper-suppression of plant defenses in response to competition. *New Phytologist* **214**:1688–1701. DOI: https://doi.org/10.1111/nph.14543, PMID: 28386959

Schimmel BCJ, Alba JM, Wybouw N, Glas JJ, Meijer TT, Schuurink RC, Kant MR. 2018. Distinct signatures of host defense suppression by Plant-Feeding mites. *International Journal of Molecular Sciences* **19**:3265. DOI: https://doi.org/10.3390/ijms19103265, PMID: 30347842

Schwager EE, Schönauer A, Leite DJ, Sharma PP, McGregor AP. 2015. Chelicerata. In: Wanninger A (Ed). *Evolutionary Developmental Biology of Invertebrates*. Springer. p. 99–139. DOI: https://doi.org/10.1007/978-3-7091-1862-7

Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-Oda Y, Esposito L, Bechsgaard J, Bilde T, Buffry AD, Chao H, Dinh H, Doddapaneni H, Dugan S, Eibner C, Extavour CG, Funch P, Garb J, Gonzalez LB, et al. 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biology* **15**:62. DOI: https://doi.org/10.1186/s12915-017-0399-x, PMID: 28756775

Sebé-Pedrós A, Ruiz-Trillo I. 2017. Chapter 1 - Evolution and Classification of the T-Box Transcription Factor Family. In: Sebé-Pedrós A (Ed). *Current Topics in Developmental Biology*. Elsevier. p. 1–26. DOI: https://doi.org/10.1016/bs.ctdb.2016.06.004

Sharma PP, Schwager EE, Extavour CG, Wheeler WC. 2014. Hox gene duplications correlate with posterior heteronomy in scorpions. *Proceedings of the Royal Society B: Biological Sciences* **281**:20140661. DOI: https://doi.org/10.1098/rspb.2014.0661, PMID: 25122224

Shi H, Pei L, Gu S, Zhu S, Wang Y, Zhang Y, Li B. 2012. Glutathione S-transferase (GST) genes in the red flour beetle, *Tribolium castaneum*, and comparative analysis with five additional insects. *Genomics* **100**:327–335. DOI: https://doi.org/10.1016/j.ygeno.2012.07.010, PMID: 22824654

Shimeld SM, Degnan B, Luke GN. 2010. Evolutionary genomics of the fox genes: origin of gene families and the ancestry of gene clusters. *Genomics* **95**:256–260. DOI: https://doi.org/10.1016/j.ygeno.2009.08.002, PMID: 19679177

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51**:492–508. DOI: https://doi.org/10.1080/10635150290069913, PMID: 12079646

Shindo T, Van der Hoorn RA. 2008. Papain-like cysteine proteases: key players at molecular battlefields employed by both plants and their invaders. *Molecular Plant Pathology* **9**:119–125. DOI: https://doi.org/10.1111/j.1364-3703.2007.00439.x, PMID: 18705889

Silbering AF, Benton R. 2010. Ionotropic and metabotropic mechanisms in chemoreception: 'chance or design'? *EMBO Reports* **11**:173–179. DOI: https://doi.org/10.1038/embor.2010.8, PMID: 20111052

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212. DOI: https://doi.org/10.1093/bioinformatics/btv351, PMID: 26059717

Simon MA, Xu A, Ishikawa HO, Irvine KD. 2010. Modulation of fat:dachsous binding by the cadherin domain kinase four-jointed. *Current Biology* **20**:811–817. DOI: https://doi.org/10.1016/j.cub.2010.04.016, PMID: 20434335

Simpson SD, Ramsdell JS, Watson III WH, Chabot CC. 2017. The Draft Genome and Transcriptome of the Atlantic Horseshoe Crab, *Limulus polyphemus* . *International Journal of Genomics* **2017**:1–14. DOI: https://doi.org/10.1155/2017/7636513

Skinner MK, Rawls A, Wilson-Rawls J, Roalson EH. 2010. Basic helix-loop-helix transcription factor gene family phylogenetics and nomenclature. *Differentiation* **80**:1–8. DOI: https://doi.org/10.1016/j.diff.2010.02.003, PMID: 20219281

Slyusarev GS, Starunov VV, Bondarenko AS, Zorina NA, Bondarenko NI. 2020. Extreme genome and nervous system streamlining in the invertebrate parasite *Intoshia variabili*. *Current Biology* **30**:1292–1298. DOI: https://doi.org/10.1016/j.cub.2020.01.061, PMID: 32084405

Smit A, Hubley R, Green P. 2013. RepeatMasker Open. 4.0. ISB. http://www.repeatmasker.org/

Smith S. 2012. wiki Quick run treePL. *GitHub*. 0.8. https://github.com/blackrim/treePL/wiki/Quick-run

Smith FW, Boothby TC, Giovannini I, Rebecchi L, Jockusch EL, Goldstein B. 2016. The compact body plan of tardigrades evolved by the loss of a large body region. *Current Biology : CB* **26**:224–229. DOI: https://doi.org/10.1016/j.cub.2015.11.059, PMID: 26776737

Smith FW, Goldstein B. 2017. Segmentation in tardigrada and diversification of segmental patterns in panarthropoda. *Arthropod Structure & Development* **46**:328–340. DOI: https://doi.org/10.1016/j.asd.2016.10.005, PMID: 27725256

Smith SA, O'Meara BC. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**:2689–2690. DOI: https://doi.org/10.1093/bioinformatics/bts492, PMID: 22908216

Snoeck S, Wybouw N, Van Leeuwen T, Dermauw W. 2018. Transcriptomic plasticity in the arthropod generalist *Tetranychus urticae* Upon Long-Term Acclimation to Different Host Plants. *G3: Genes, Genomes, Genetics* **8**:3865–3879. DOI: https://doi.org/10.1534/g3.118.200585, PMID: 30333191

Snoeck S, Pavlidi N, Pipini D, Vontas J, Dermauw W, Van Leeuwen T. 2019. Substrate specificity and promiscuity of horizontally transferred UDP-glycosyltransferases in the generalist herbivore *Tetranychus urticae*. *Insect Biochemistry and Molecular Biology* **109**:116–127. DOI: https://doi.org/10.1016/j.ibmb.2019.04.010, PMID: 30978500

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313. DOI: https://doi.org/10.1093/bioinformatics/btu033, PMID: 24451623

Stogios PJ, Privé GG. 2004. The BACK domain in BTB-kelch proteins. *Trends in Biochemical Sciences* **29**:634–637. DOI: https://doi.org/10.1016/j.tibs.2004.10.003, PMID: 15544948

Su Q, Peng Z, Tong H, Xie W, Wang S, Wu Q, Zhang J, Li C, Zhang Y. 2019. A salivary ferritin in the whitefly suppresses plant defenses and facilitates host exploitation. *Journal of Experimental Botany* **70**:3343–3355. DOI: https://doi.org/10.1093/jxb/erz152, PMID: 30949671

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**:2725–2729. DOI: https://doi.org/10.1093/molbev/mst197, PMID: 24132122

Tang O. 2011. *GNU Parallel: The Command-Line Power Tool, ; Login*: The USENIX Magazine.

Telford MJ, Thomas RH. 1998. Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *PNAS* **95**:10671–10675. DOI: https://doi.org/10.1073/pnas.95.18.10671, PMID: 9724762

The C. elegans Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**:2012–2018. DOI: https://doi.org/10.1126/science.282.5396.2012, PMID: 9851916

Thomson SA, Baldwin WS, Wang YH, Kwon G, Leblanc GA. 2009. Annotation, phylogenetics, and expression of the nuclear receptors in *Daphnia pulex*. *BMC Genomics* **10**:500. DOI: https://doi.org/10.1186/1471-2164-10-500, PMID: 19863811

Tomato Genome Consortium, Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isobe S, Kaneko T, Nakamura Y, Shibata D, Aoki K, Egholm M, Knight J, Bogden R, Li Changbao SY, Xu X, Pan S, Cheng S, Liu X, Ren Y, Wang J, et al. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**:635–641. DOI: https://doi.org/10.1038/nature11119, PMID: 22660326

Tomoyasu Y, Miller SC, Tomita S, Schoppmeier M, Grossmann D, Bucher G. 2008. Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in tribolium. *Genome Biology* **9**:R10. DOI: https://doi.org/10.1186/gb-2008-9-1-r10, PMID: 18201385

Toruño TY, Stergiopoulos I, Coaker G. 2016. Plant-Pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. *Annual Review of Phytopathology* **54**:419–441. DOI: https://doi.org/10.1146/annurev-phyto-080615-100204, PMID: 27359369

Touhara K, Vosshall LB. 2009. Sensing odorants and pheromones with chemosensory receptors. *Annual Review of Physiology* **71**:307–332. DOI: https://doi.org/10.1146/annurev.physiol.010908.163209, PMID: 19575682

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology* **31**:46–53. DOI: https://doi.org/10.1038/nbt.2450, PMID: 23222703

van Houten YM, Glas JJ, Hoogerbrugge H, Rothe J, Bolckmans KJ, Simoni S, van Arkel J, Alba JM, Kant MR, Sabelis MW. 2013. Herbivory-associated degradation of tomato trichomes and its impact on biological control of *Aculops lycopersici*. *Experimental and Applied Acarology* **60**:127–138. DOI: https://doi.org/10.1007/s10493-012-9638-6, PMID: 23238958

Van Leeuwen T, Witters J, Nauen R, Duso C, Tirry L. 2010. The control of eriophyoid mites: state of the art and future challenges. *Experimental and Applied Acarology* **51**:205–224. DOI: https://doi.org/10.1007/s10493-009-9312-9, PMID: 19768561

Van Leeuwen T, Dermauw W. 2016. The molecular evolution of xenobiotic metabolism and resistance in chelicerate mites. *Annual Review of Entomology* **61**:475–498. DOI: https://doi.org/10.1146/annurev-ento-010715-023907, PMID: 26982444

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, et al. 2001. The sequence of the human genome. *Science* **291**:1304–1351. DOI: https://doi.org/10.1126/science.1058040, PMID: 11181995

Villarroel CA, Jonckheere W, Alba JM, Glas JJ, Dermauw W, Haring MA, Van Leeuwen T, Schuurink RC, Kant MR. 2016. Salivary proteins of spider mites suppress defenses in *Nicotiana benthamiana* and promote mite reproduction. *The Plant Journal* **86**:119–131. DOI: https://doi.org/10.1111/tpj.13152, PMID: 26946468

Waldron R, McGowan J, Gordon N, McCarthy C, Mitchell EB, Doyle S, Fitzpatrick DA. 2017. Draft genome sequence of *Dermatophagoides pteronyssinus*, the European House Dust Mite. *Genome Announcements* **5**:e00789-17. DOI: https://doi.org/10.1128/genomeA.00789-17, PMID: 28798186

Wang H, Devos KM, Bennetzen JL. 2014. Recurrent loss of specific introns during angiosperm evolution. *PLOS Genetics* **10**:e1004843. DOI: https://doi.org/10.1371/journal.pgen.1004843, PMID: 25474210

Wei P, Demaeght P, De Schutter K, Grigoraki L, Labropoulou V, Riga M, Vontas J, Nauen R, Dermauw W, Van Leeuwen T. 2020. Overexpression of an alternative allele of carboxyl/choline esterase 4 (CCE04) of *tetranychus urticae* is associated with high levels of resistance to the keto-enol acaricide spirodiclofen. *Pest Management Science* **76**:1142–1153. DOI: https://doi.org/10.1002/ps.5627, PMID: 31583806

Whiteman NK, Pierce NE. 2008. Delicious poison: genetics of *Drosophila* host plant preference. *Trends in Ecology & Evolution* **23**:473–478. DOI: https://doi.org/10.1016/j.tree.2008.05.010, PMID: 18657878

Whitmoyer RE, Nault LR, Bradfute OE. 1972. Fine structure of *Aceria tulipae* (Acarina: Eriophyidae)1. *Annals of the Entomological Society of America* **65**:201–215. DOI: https://doi.org/10.1093/aesa/65.1.201

Wickham H. 2017. stringr:Simple, Consistent Wrappers for Common String Operations. *Github*. f70c4ba. https://github.com/tidyverse/stringr

Wickham H, Francois R, Henry L, Müller K. 2017. Dplyr. 1.0.2. A Grammar of Data Manipulation. https://dplyr.tidyverse.org/

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: https://doi.org/10.1371/journal.pone.0160009, PMID: 27467523

Wybouw N, Zhurov V, Martel C, Bruinsma KA, Hendrickx F, Grbić V, Van Leeuwen T. 2015. Adaptation of a polyphagous herbivore to a novel host plant extensively shapes the transcriptome of herbivore and host. *Molecular Ecology* **24**:4647–4663. DOI: https://doi.org/10.1111/mec.13330, PMID: 26211543

Wybouw N, Pauchet Y, Heckel DG, Van Leeuwen T. 2016. Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome Biology and Evolution* **8**:1785–1801. DOI: https://doi.org/10.1093/gbe/evw119, PMID: 27307274

Wybouw N, Van Leeuwen T, Dermauw W. 2018. A massive incorporation of microbial genes into the genome of *Tetranychus Urticae*, a polyphagous arthropod herbivore. *Insect Molecular Biology* **27**:333–351. DOI: https://doi.org/10.1111/imb.12374, PMID: 29377385

Xue XF, Dong Y, Deng W, Hong XY, Shao R. 2017. The phylogenetic position of eriophyoid mites (superfamily eriophyoidea) in acariformes inferred from the sequences of mitochondrial genomes and nuclear small subunit (18S) rRNA gene. *Molecular Phylogenetics and Evolution* **109**:271–282. DOI: https://doi.org/10.1016/j.ympev.2017.01.009, PMID: 28119107

Yan N. 2015. Structural biology of the major facilitator superfamily transporters. *Annual Review of Biophysics* **44**:257–283. DOI: https://doi.org/10.1146/annurev-biophys-060414-033901, PMID: 26098515

Ye Z, Xu S, Spitze K, Asselman J, Jiang X, Ackerman MS, Lopez J, Harker B, Raborn RT, Thomas WK, Ramsdell J, Pfrender ME, Lynch M. 2017. A new reference genome assembly for the microcrustacean *Daphnia pulex*. *G3: Genes, Genomes, Genetics* **7**:1405–1416. DOI: https://doi.org/10.1534/g3.116.038638, PMID: 28235826

Yenerall P, Krupa B, Zhou L. 2011. Mechanisms of intron gain and loss in *Drosophila*. *BMC Evolutionary Biology* **11**:364. DOI: https://doi.org/10.1186/1471-2148-11-364, PMID: 22182367

Yoshida Y, Koutsovoulos G, Laetsch DR, Stevens L, Kumar S, Horikawa DD, Ishino K, Komine S, Kunieda T, Tomita M, Blaxter M, Arakawa K. 2017. Comparative genomics of the tardigrades *Hypsibius dujardini* and

Ramazzottius varieornatus. PLOS Biology **15**:e2002266. DOI: https://doi.org/10.1371/journal.pbio.2002266, PMID: 28749982

Zelle KM, Lu B, Pyfrom SC, Ben-Shahar Y. 2013. The genetic architecture of degenerin/epithelial sodium channels in *Drosophila*. *G3: Genes, Genomes, Genetics* **3**:441–450. DOI: https://doi.org/10.1534/g3.112.005272, PMID: 23449991

Zhang Z-Q. 2011. *Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness*. Magnolia Press.

Zhong L, Wang D, Gan X, Yang T, He S. 2011. Parallel expansions of sox transcription factor group B predating the diversifications of the arthropods and jawed vertebrates. *PLOS ONE* **6**:e16570. DOI: https://doi.org/10.1371/journal.pone.0016570, PMID: 21305035

Zhu Q. 2014. AfterPhylo, a Perl script for manipulating trees after phylogenetic reconstruction. *GitHub*. v0.9.1. https://github.com/qiyunzhu/AfterPhylo

Zhu T, Niu DK. 2013. Mechanisms of intron loss and gain in the fission yeast schizosaccharomyces. *PLOS ONE* **8**:e61683. DOI: https://doi.org/10.1371/journal.pone.0061683, PMID: 23613904

Zong J, Yao X, Yin J, Zhang D, Ma H. 2009. Evolution of the RNA-dependent RNA polymerase (RdRP) genes: Duplications and possible losses before and after the divergence of major eukaryotic groups. *Gene* **447**:29–39. DOI: https://doi.org/10.1016/j.gene.2009.07.004, PMID: 19616606