

The genome of *Tetranychus urticae* reveals herbivorous pest adaptations

Miodrag Grbić^{1,2*}, Thomas Van Leeuwen^{3*}, Richard M. Clark^{4*}, Stephane Rombauts^{5,6}, Pierre Rouzé^{5,6}, Vojislava Grbić^{1,2}, Edward J. Osborne⁴, Wannes Dermauw³, Phuong Cao Thi Ngoc^{5,6}, Félix Ortego⁷, Pedro Hernández-Crespo⁷, Isabel Diaz⁸, Manuel Martínez⁸, Maria Navajas⁹, Élio Sucena^{10,11}, Sara Magalhães¹², Lisa Nagy¹³, Ryan M. Pace¹³, Sergej Djuranović¹⁴, Guy Smagghe³, Masatoshi Iga³, Olivier Christiaens³, Jan A. Veenstra¹⁵, John Ewer¹⁶, Rodrigo Mancilla Villalobos¹⁶, Jeffrey L. Hutter¹⁷, Stephen D. Hudson¹⁷, Marisela Velez¹⁸, Soojin V. Yi¹⁹, Jia Zeng¹⁹, Andre Pires-daSilva²⁰, Fernando Roch²¹, Marc Cazaux¹, Marie Navarro¹, Vladimir Zhurov¹, Gustavo Acevedo¹, Anica Bjelica¹, Jeffrey A. Fawcett^{5,6,†}, Eric Bonnet^{5,6,†}, Cindy Martens^{5,6}, Guy Baele^{5,6}, Lothar Wissler²², Aminael Sanchez-Rodriguez²³, Luc Tirry³, Catherine Blais²⁴, Kristof Demeester²⁵, Stefan R. Henz²⁶, T. Ryan Gregory²⁷, Johannes Mathieu²⁸, Lou Verdon²⁹, Laurent Farinelli³⁰, Jeremy Schmutz^{31,32}, Erika Lindquist³², René Feyereisen³³ & Yves Van de Peer^{5,6}

The spider mite *Tetranychus urticae* is a cosmopolitan agricultural pest with an extensive host plant range and an extreme record of pesticide resistance. Here we present the completely sequenced and annotated spider mite genome, representing the first complete chelicerate genome. At 90 megabases *T. urticae* has the smallest sequenced arthropod genome. Compared with other arthropods, the spider mite genome shows unique changes in the hormonal environment and organization of the Hox complex, and also reveals evolutionary innovation of silk production. We find strong signatures of polyphagy and detoxification in gene families associated with feeding on different hosts and in new gene families acquired by lateral gene transfer. Deep transcriptome analysis of mites feeding on different plants shows how this pest responds to a changing host environment. The *T. urticae* genome thus offers new insights into arthropod evolution and plant–herbivore interactions, and provides unique opportunities for developing novel plant protection strategies.

Mites belong to the Chelicerata, the second largest group of terrestrial animals. Chelicerates represent a basal branch of arthropods. Subsequent to their origin in the Cambrian period, arthropods radiated into two lineages: the Chelicerata and the Mandibulata (comprising the Myriapoda and the Pancrustacea (which includes both crustaceans and insects))^{1,2}. Extant lineages of chelicerates include Pycnogonida, Xiphosura (horseshoe crabs) and Arachnida (a large group comprising scorpions, spiders and the Acari (ticks and mites))^{3,4} (Supplementary Fig. 1.1). Within the Acari, *T. urticae* belongs to the Acariformes with the earliest fossils dating from the Lower Devonian period (410 million years ago). The Acari represent the most diverse chelicerate clade, with over 40,000 described species that exhibit tremendous variations in lifestyle, ranging from parasitic to predatory to plant-feeding. Some mites are of major concern to

human health and include allergy-causing dust mites, scabies mites and mite vectors of scrub typhus⁵.

The two-spotted spider mite, *Tetranychus urticae*, is a cosmopolitan agricultural pest⁶ belonging to an assemblage of web-spinning mites. The name ‘spider’ highlights their ability to produce silk-like webbing used to establish a colonial micro-habitat, protect against abiotic agents, shelter from predators, communicate via pheromones and provide a vehicle for dispersion⁷.

Tetranychus urticae represents one of the most polyphagous arthropod herbivores, feeding on more than 1,100 plant species belonging to more than 140 different plant families including species known to produce toxic compounds. It is a major pest in greenhouse production and field crops, destroying annual and perennial crops such as tomatoes, peppers, cucumbers, strawberries, maize, soy,

¹Department of Biology, The University of Western Ontario, London N6A 5B7, Canada. ²Instituto de Ciencias de la Vid y el Vino (CSIC, UR, Gobierno de La Rioja), 26006 Logroño, Spain. ³Department of Crop Protection, Faculty of Bioscience Engineering, Ghent University, B-9000 Ghent, Belgium. ⁴Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA. ⁵Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium. ⁶Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium. ⁷Department of Environmental Biology, Centro de Investigaciones Biológicas, CSIC, 28040 Madrid, Spain. ⁸Centro de Biotecnología y Genómica de Plantas, UPM-INIA, 28223 Madrid, Spain. ⁹INRA, UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Campus international de Baillarguet, 34988 Montpellier-sur-Lez, France. ¹⁰Instituto Gulbenkian de Ciência, 2781-901 Oeiras, Portugal. ¹¹Universidade de Lisboa, Faculdade de Ciências, Departamento de Biologia Animal, 1749-016 Lisbon, Portugal. ¹²Universidade de Lisboa, Faculdade de Ciências, Centro de Biologia Ambiental, 1749-016 Lisbon, Portugal. ¹³Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721, USA. ¹⁴Johns Hopkins University School of Medicine, Department of Molecular Biology & Genetics, Baltimore, Maryland 21205, USA. ¹⁵Institut de Neurosciences Cognitives et Intégratives d'Aquitaine Université de Bordeaux 1, 33405 Talence, France. ¹⁶Centro Interdisciplinario de Neurociencia de Valparaíso, Facultad de Ciencias, Universidad de Valparaíso, Valparaíso 2360102, Chile. ¹⁷Department of Physics and Astronomy, The University of Western Ontario, N6A 5B7 London, Canada. ¹⁸Instituto de Catálisis y Petroleoquímica CSIC, Madrid, Spain; IMDEA Nanociencias, Facultad de Ciencias, Universidad Autónoma de Madrid, 28050 Madrid, Spain. ¹⁹School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. ²⁰Department of Biology, University of Texas at Arlington, Arlington, Texas 76019, USA. ²¹Université de Toulouse, UPS, Centre de Biologie du Développement, Université Paul Sabatier, 31062 Toulouse, France; Centre National de la Recherche Scientifique, UMR 5547, Centre de Biologie du Développement, 31062 Toulouse, France. ²²Westfälische Wilhelms University, Institute for Evolution and Biodiversity, Evolutionary Bioinformatics Group, Hüfferstrasse 1, D-48149 Münster, Germany. ²³CMPG, Department of Microbial and Molecular Systems, K.U. Leuven, B-3001 Leuven, Belgium. ²⁴UPMC Univ Paris 06, UMR CNRS 7622, Equipe Biogenèse des signaux hormonaux, Case 29, 75005 Paris, France. ²⁵Research Group EnVOC, Department of Sustainable Organic Chemistry and Technology, Faculty of Bioscience Engineering, Ghent University, B-9000 Ghent, Belgium. ²⁶Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany. ²⁷Department of Integrative Biology, University of Guelph, N1G 2W1 Guelph, Canada. ²⁸Boyce Thompson Institute for Plant Research, Ithaca, New York 14853, USA. ²⁹Southern Crop Protection and Food Research Centre, Agriculture and Agri-Food Canada, N5V 4T3 London, Canada. ³⁰Fasteris SA, CH-1228 Plan-les-Ouates, Switzerland. ³¹HudsonAlpha Institute for Biotechnology Huntsville, Alabama 35806, USA. ³²DOE Joint Genome Institute, Walnut Creek, California 94598, USA. ³³UMR 1301, INRA, CNRS and Université de Nice Sophia Antipolis, 06903 Sophia Antipolis, France. †Present addresses: Institut Curie, 26 rue d'Ulm, Paris 75248, France; INSERM, U900, Paris 75248, France; Mines ParisTech, Fontainebleau 77300, France (E.B.); Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan (J.A.F.).

*These authors contributed equally to this work.

apples, grapes and citrus. The recent introduction of the related species *Tetranychus evansi* to Europe and Africa from South America demonstrates the invasive nature of these pests in global agriculture⁸. Computer modelling suggests that with intensifying global warming, the detrimental effects of spider mites in agriculture will markedly increase⁹ due to accelerated development at high temperatures.

Tetranychus urticae is known for its ability to develop rapid resistance to pesticides. Among arthropods it has the highest incidence of pesticide resistance¹⁰. Chemical control often causes a broad cross-resistance within and between pesticide classes, resulting in resistance to novel pesticides within 2–4 years. Many aspects of the biology of the spider mite, including rapid development, high fecundity and haplo-diploid sex determination, seem to facilitate rapid evolution of pesticide resistance. Control of multi-resistant mites has become increasingly difficult and the genetic basis of such resistance remains poorly understood¹¹.

As the first completed chelicerate genome, the comparison of the *T. urticae* genome with the genomes of insects and the crustacean *Daphnia pulex* expands the arthropod genetic toolkit. At the same time, the very compact *T. urticae* genome has unique attributes among arthropod genomes with remarkable instances of gene gains and losses. The completion of the *T. urticae* genome sequence opens new avenues for understanding the fundamentals of plant–herbivore interactions, developing novel pest-management strategies and producing new biomaterials on the nanometre scale.

The small genome of *T. urticae*

The *T. urticae* genome (strain London) was sequenced (Sanger) to 8.05× coverage and assembled into 640 scaffolds covering 89.6 megabases (Mb) (Supplementary Notes 1, 2.1 and 2.2). 70,778 Sanger expressed sequence tag (EST) sequences from embryos, larvae, nymphs and adults were generated, and further complemented with RNA-seq data on matching samples (Supplementary Note 2.3). We identified 18,414 protein-coding gene models, of which 84% (15,397) are supported by EST (8,243), protein homology (11,433) and/or RNA-seq data (14,545) (Supplementary Note 2.4 and Supplementary Fig. 2.4.1). From alignments of ~43-million paired-end Illumina reads from a second *T. urticae* strain (Montpellier) to the London sequence, 542,600 single nucleotide polymorphisms and small indels were predicted (Supplementary Note 2.5). The complete genome annotation of *T. urticae* is available at the BOGAS website¹². With an estimated genome size of about 90 Mb, the *T. urticae* genome is the smallest arthropod genome sequenced so far. The genomes of other chelicerates are much larger (565–7,100 Mb), with the unfinished genome of the tick *Ixodes scapularis* estimated at 2,100 Mb¹³. Multiple characteristics of the *T. urticae* genome correlate with its compact size: small transposable element content and microsatellite density, increased gene density and holocentric chromosomes (see Supplementary Note 3.1 for chromosomal features).

Transposable elements totalled 9.09 Mb (Supplementary Note 3.2), putting *T. urticae* together with *D. pulex* and *Apis mellifera* as arthropods with 10% or less of their genomes comprised of transposable elements. Long terminal repeat (LTR) retrotransposons, and in particular Gypsy-like elements, were the most abundant type of transposable elements. L1-like Long interspersed elements (LINEs), Tc1/Mariner-like DNA transposons, and Maverick (Polinton) elements were also detected (see Supplementary Table 3.2.1). Deep sequencing of small RNAs (~19–30 nucleotides) across developmental stages (Supplementary Note 4.1) identified 226,829 unique RNAs that mapped to 676,266 different loci in the genome. The number of unique small RNA counts per size category shows a peak at 21 and 26 nucleotides. These two peaks include short interfering RNAs and Piwi-interacting RNAs, respectively, similar to what is observed in *Drosophila melanogaster*¹⁴. Their alignments to the genome indicate that both probably silence diverse transposable elements. Included among ~21-nucleotide small RNAs are 52

predicted microRNAs (miRNAs). On the basis of the identity of their seed regions (nucleotides 2–7 of the miRNA sequence), the *T. urticae* miRNAs can be grouped into 43 families (Supplementary Note 4). Half of the predicted miRNAs were not conserved when compared to annotated miRNAs and available genomes of other arthropods¹⁵, suggesting that they might be *T. urticae*- or lineage-specific (Supplementary Tables 4.3.1–4.3.4).

The microsatellite density in the *T. urticae* genome is among the lowest observed for arthropods (Supplementary Note 3.3 and Supplementary Fig. 3.3.1), consistent with the expectation that repeat content of genomes typically scales with genome size. The *T. urticae* microsatellite classes have a distinct profile: mono-nucleotide repeats are virtually non-existent, and di-nucleotide repeats, normally the most abundant type of microsatellites, are found significantly less often than tri-nucleotides, as in *Tribolium castaneum*¹⁶. The gene density is twice as high compared to *D. melanogaster*, with 205 versus 92 genes per Mb, respectively. The mean number of exons per gene was low and similar to that found in *D. melanogaster* (~3.8 exons per gene). The size distribution of introns was typically skewed with a mean intron size of 400 bp and a median of 96 bp (see Supplementary Note 3.4, Supplementary Fig. 2.4.3 and Supplementary Table 2.4.1). The holocentric nature of *T. urticae* chromosomes¹⁷ (the absence of centromeres and the diffuse nature of the kinetochores) is correlated with a lack of large tracts of gene-poor heterochromatin. The uniformly distributed gene density (Supplementary Note 3.1.1 and Supplementary Fig. 3.2.1) contrasts with the human body louse (*Pediculus humanus*, Phthiraptera, a hemimetabolous insect with a small genome), where 95% of the genes are concentrated in only 55 Mb of the 110-Mb genome¹⁸.

Comparative genomics

As the first completely sequenced and annotated chelicerate genome, the *T. urticae* genome expands the set of arthropod genomes beyond Pancrustacea and provides an important out-group for comparative genomics. Comparison of the coding gene repertoire of *T. urticae* with the arthropods *T. castaneum*, *D. melanogaster*, *Nasonia vitripennis* and *D. pulex*, the chordate *Homo sapiens*, and the cnidarian *Nematostella vectensis* (Fig. 1) resulted in 2,667 shared gene families (Supplementary Note 5.1). Almost 3,000 gene families are common to the arthropods sampled, whereas 5,038 gene families (8,329 genes) are unique to *T. urticae* (Supplementary Fig. 5.1.1). Of those, 622 gene families (1,398 genes) have homologues in species other than those listed above, most of which belong to other arthropods. Homologues of 74 gene families (93 genes) were found in the unfinished genomes of tick¹³ and/or *Varroa destructor*¹⁹ and are probably chelicerate, rather than specific to *T. urticae*. Therefore, 4,416 gene families (6,609 genes) were found to be unique to *T. urticae*. A gene gain/loss analysis (Fig. 1 and

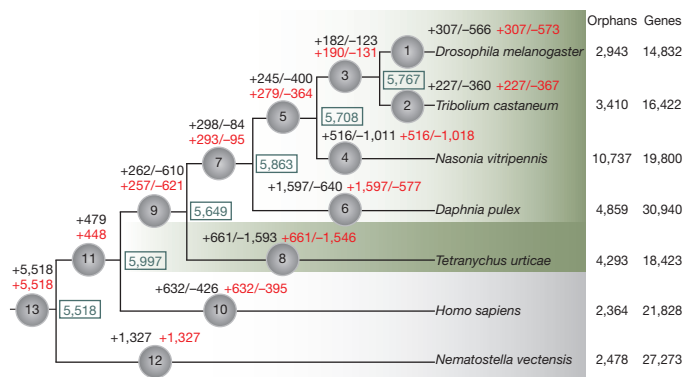


Figure 1 | Gene family history. At each time point (grey circles), the number of gains (+) and losses (–) of gene families is indicated as inferred by DOLLOP (black) and CAFÉ (red) programs. The inferred ancestral number of gene families, according to DOLLOP, is shown in green boxes.

Supplementary Note 5.2) of these genomes showed a gain of about 700 new gene families in the lineage leading to *T. urticae*, plus almost 4,300 genes that are single copy (orphans). More than 1,000 gene families, still present in other arthropods, were lost in *T. urticae*. The 58 gene families that are significantly (z -score >2) expanded in *T. urticae* compared to the other arthropods are shown in Supplementary Note 5.2 and Supplementary Fig. 5.2.1.

Feeding and detoxification

Tetranychus urticae is one of the most striking examples of polyphagy among herbivores and it has an unmatched ability to develop resistance to pesticides^{6,10}. We discovered that known gene families implicated in digestion, detoxification and transport of xenobiotics had a unique spider mite composition, and were often expanded when compared to insects (Supplementary Note 6.1). This included a three-fold proliferation of cysteine peptidase genes, particularly C1A papain and the C13 legumain genes (Supplementary Table 6.1.11), consistent with proteolytic digestion based mostly on cysteine peptidase activity²⁰. Eighty-six cytochrome P450 (CYP) genes were detected in the *T. urticae* genome, a total number similar to insects but with an expansion of *T. urticae*-specific intronless genes of the CYP2 clan (Supplementary Table 6.1.2). The carboxyl/cholinesterases (CCEs) gene family contained 71 genes, with a single acetylcholinesterase gene (*Ace1*) but two new clades at the root of the neurodevelopmental class of CCEs, representing 34 and 22 CCEs, respectively

(Supplementary Table 6.1.6). A notable case of expansion was found within the family of 32 glutathione *S*-transferases (GSTs) that include a group of 12 Mu-class GSTs that were, until now, believed to be vertebrate-specific (Supplementary Table 6.1.3). Finally, we discovered 39 multidrug resistance proteins belonging to the ATP-binding cassette (ABC) transporters (class C). The repertoire from this class of ABC transporters far exceeds the number (9–14) found in crustaceans, insects, vertebrates and nematodes (Supplementary Table 6.1.8). Few of the genes involved in detoxification had close insect homologues, and only four of the CYP genes could clearly be assigned as orthologues of insect and crustacean CYP genes.

The involvement of these gene families and their spider-mite-specific expansion in host plant adaptation is markedly illustrated by RNA-seq transcriptome profiling of spider mite feeding on its preferred host, bean (*Phaseolus vulgaris*), and on hosts to which the London strain is not adapted: *Arabidopsis thaliana* and tomato (*Solanum lycopersicum*) (Fig. 2) (Supplementary Notes 6.2). We found 24% of all genes to be differentially expressed upon host transfer (Fig. 2a–c); relative to bean, more genes were differentially expressed on tomato than on *A. thaliana* (Supplementary Note 6.2.4 and Supplementary Fig. 6.2.1), but responses were nonetheless correlated (Fig. 2b, c). Genes in the detoxification and peptidase families exhibited the most profound changes (Fig. 2a–c), with expression of nearly half of P450 genes affected by the host plant, including 19 of 39 genes in the intronless CYP392 family and the CYP389 family. These subfamilies

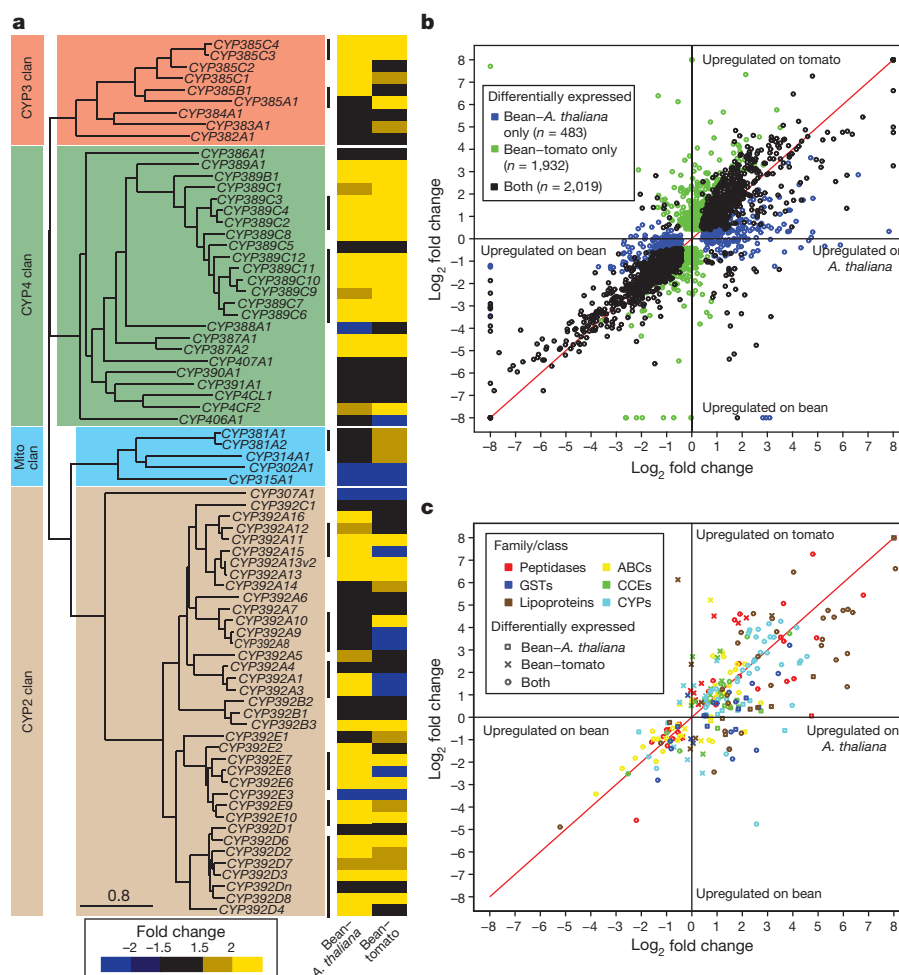


Figure 2 | Gene expression changes when mites are shifted from *P. vulgaris* (bean) to *A. thaliana* or to *S. lycopersicum* (tomato). **a**, A phylogeny of the cytochrome P450 (CYP) genes and heat map of the response of CYP genes to host transfer. Two-thirds of the genes that are tandemly duplicated or that form clusters (indicated by black vertical lines) are co-regulated. **b**, Global changes in

gene expression after host shift. **c**, Fold changes of important gene family members in digestion and detoxification are colour coded. The analysis of differential expression (**b** and **c**) is with a 5% false discovery rate as assessed with RNA-seq data collected in biological triplicate (fold changes between mean values are plotted).

are spider-mite-specific P450 expansions that define lineage-specific expansions²¹. This finding is unprecedented. In humans, only up to one-third of P450 genes are metabolizing xenobiotics²², and in *D. melanogaster* only one-third of the CYP genes are inducible by xenobiotics²³. The proportion of P450 genes responding to the chemical environment is much greater in the spider mite. Similar patterns were also found within other families (Fig. 2c). For GSTs and CCEs, the expression of Mu and Delta GSTs and the two spider-mite-specific CCE clades were most affected and about one-third of cysteine peptidases, the C1A papains and C13 legumains, were overexpressed after transfer to tomato. More than two-thirds of the CYP and GST genes affected by the host plant are present in clusters of (multiple) tandem duplicated genes. Co-regulation of the majority of tandem duplicates strongly indicates that the ancestral gene was already plant-responsive before duplication, and that a role in plant adaptation may have favoured duplicate retention.

Although these data indicate that spider-mite-specific expansion of known gene families contributes to the ability of spider mites to overcome host defences, many genes differentially regulated upon host transfer lack homology to genes of known function. Notably, among those with the most extreme expression fold-changes are genes that encode putative secreted proteins or lipid-binding proteins. Understanding extracellular binding and transport of small ligands is therefore likely to be important in further dissecting spider mite-plant interactions.

Lateral gene transfer

Our search for genes related to detoxification and digestion also revealed the existence and surprising expansion of intradiol ring cleavage dioxygenases, genes previously unreported from metazoan genomes but characteristic for bacteria and fungi²⁴. We annotated 16 functional genes in this family in *T. urticae*, whereas bacterial genomes usually carry only 1 to 7. They have an average sequence similarity of 43% with the homologue of *Streptomyces avermitilis* and share the conserved 2 His 2 Tyr non-haem iron(III) binding site. These dioxygenases might have evolved to metabolize aromatic compounds found in plant allelochemicals. Other clear instances of lateral gene transfers include (1) the presence of a cobalamin-independent methionine synthase (*MetE*) gene with four predicted introns and up to 58% sequence identity to the *MetE* gene of soil Bacilli (this sequence has not previously been reported in any animal species); (2) two very similar levanase-encoding genes of probable bacterial origin that encode secreted exo-fructosidases upregulated upon feeding on tomato; and (3) a cyanate lyase-encoding gene that might be involved in feeding on cyanogenic plants (Supplementary Table 6.3.1).

We detected two clusters of carotenoid biosynthesis genes in *T. urticae* representing homologues of genes from zygomycete fungi and aphids. The latter are the only animal carotenoid biosynthesis genes known so far, thought to be derived from fungal genes by lateral gene transfer²⁵. The unique intron–exon structure of the spider mite and aphid genes and their clustering in phylogenetic analyses is strong evidence that the genes from fungi were transferred only once to arthropods (Fig. 3). The sequence and orientation of the two spider mite clusters indicate that they are the result of an ancient transfer followed by duplications, rearrangements and divergence. They also suggest that a second, more recent transfer occurred between a spider mite and an aphid ancestor, although the sequence of the two transfers remains speculative. Carotenoids are known to have a role in diapause induction in spider mites²⁶ and our findings indicate that they can also synthesize them.

Ponasterone A as moulting hormone

Ecdysteroid control of moulting is one of the defining features of arthropods. We detected gene orthologues coding for ecdysteroid biosynthesis enzymes (Supplementary Table 7.1.1)²¹. Surprisingly,

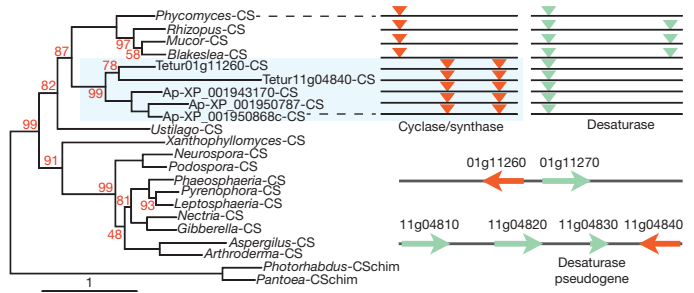


Figure 3 | Maximum likelihood phylogeny of the fungal and arthropod carotenoid cyclase/synthase (CS) fusion proteins. The out-group comprises chimaeric assemblies (CSchim) of the closest bacterial sequences of cyclases and synthases. The *T. urticae* and *Acyrtosiphon pisum* sequences form a monophyletic group closely related to the zygomycete sequences. Evidence for a single lateral gene transfer event is also shown by the common intron positions in the cyclase/synthase (orange) and desaturase (green) genes (upper right panel). Two clusters of carotenoid biosynthesis genes are found in *T. urticae*: a tail-to-tail arrangement on scaffold 1 as seen in zygomycetes and aphids, and a more complex head-to-head (re)arrangement on scaffold 11 (bottom right).

the *T. urticae* genome lacks two P450 genes, *CYP306A1* and *CYP18A1*, encoding, respectively, the biosynthetic C25 hydroxylase and a C26 hydroxylase/oxidase involved in hormone inactivation. The absence of *CYP306A1* indicates that the spider mite uses the ecdysteroid 25-deoxy-20-hydroxyecdysone (ponasterone A) as the moulting hormone, instead of the typical arthropod 20E. This was confirmed by biochemical analysis of spider mite extracts by HPLC–enzyme immunoassay and liquid chromatography/mass spectrometry that identified ponasterone A (Supplementary Note 7). *CYP306A1* and *CYP18A1* form a head-to-head cluster in all insect and crustacean genomes studied so far, therefore their absence from the *T. urticae* genome indicates that they were lost together, affecting both biosynthesis and inactivation pathways of the spider mite moulting hormone. Ponasterone A has been previously identified in some decapod crustaceans, albeit always coincident with 20E (ref. 27), and it is a high potency ligand of all known ecdysteroid receptors.

Reduced Hox cluster

Hox genes are a conserved set of homeobox-containing transcription factors typically found clustered within the genome and used to establish region-specific identity during early development. The body plan of mites consists of an anterior prosoma and posterior opisthosoma and is further distinguished by an extremely reduced body plan presumably achieved through the fusion of segments (Supplementary Note 8 and Fig. 4b). The ancestral arthropod is predicted to have a Hox cluster with 10 genes²⁸. The *T. urticae* genome contains 8 of the canonical 10 genes. The *ftz* gene is present in duplicate, in two closely linked copies; orthologues of *Hox3* and *abdominal A* (*abdA*) were not found (Fig. 4a). This is unusual among chelicerates: all 10 canonical Hox genes are present in the wandering spider²⁹. The absence of *abdA* in *T. urticae* correlates with the spider mite's reduced opisthosomal segmentation. Consistent with the absence of *abdA* and a reduced opisthosoma, only two opisthosomal stripes of the segment polarity gene *engrailed* (typically expressed in each arthropod segment) are detected in the developing embryo (Fig. 4c), in contrast to five *engrailed* stripes detected in the opisthosoma of the wandering spider³⁰. Although numerous examples correlate morphological variation in arthropods with changes in Hox gene expression, this is the first example that correlates morphological evolution with the loss of a Hox gene within a fully sequenced Hox cluster.

Nanometre dimensions of *T. urticae* silk

Silk production in spider mites (Fig. 5a, b and Supplementary Note 9) represents a *de novo* evolution of silk-spinning relative to silk production in spiders⁷. Spiders typically spin silk from a complex glandular

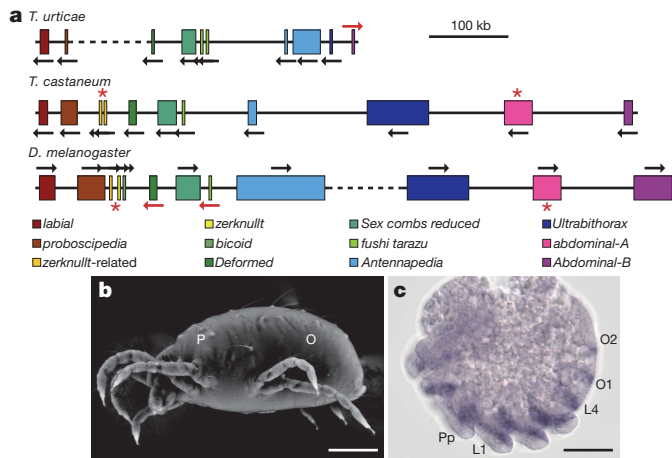


Figure 4 | Comparative organization of Hox clusters and expression pattern of the *T. urticae* engrailed gene. **a**, *T. urticae*, *T. castaneum* and *D. melanogaster* Hox clusters. Gene sizes and intergenic distances are shown to scale. Dashed lines represent breaks in the cluster >1 Mb. In *T. urticae*, *fushi tarazu* and *Antennapedia* are present in duplicate whereas *abdominal-A* and *Hox3/zerknullt* are missing (red asterisk). **b**, Variable pressure scanning electron microscopy (SEM) image of adult *T. urticae* with two main body regions indicated: P, prosoma; O, opisthosoma. **c**, *T. urticae* engrailed (*en*) expression pattern. *en* transcripts are detected in five prosomal stripes that correspond to future pedipalpal (Pp), four walking leg (L1–L4) and two opisthosomal (O1 and O2) segments. Scale bars: **b**, 0.125 mm; **c**, 40 μ m.

abdominal spinneret, whereas *T. urticae* uses paired silk glands connected to the mouth appendages (pedipalps)³¹. Seventeen fibroin genes were uncovered in the genome of *T. urticae* (Supplementary Table 9.1.1) encoding fibroins of unusually high (27–39%) serine content. We performed mechanical testing on fibres deposited by adult and larval mites with an atomic force microscope. This technique measures the Young's modulus of the fibres, which is the ratio of applied stress (tension per cross-sectional area) to the resulting strain (fractional change in length) and describes the stiffness of the material. Young's modulus was higher than or comparable to other natural materials (see Supplementary Table 9.1.2), but *T. urticae* silk fibres are thinner— 54 ± 3 nm (adult silk, Fig. 5c) and 23.3 ± 0.9 nm (larval silk), that is, 435–185 times thinner—than the silk fibres of the spider *Nephila clavipes*³².

Concluding remarks

Our analysis of the *T. urticae* genome also included nuclear receptors and neuropeptide genes, immunity-related genes and RNA interference, cuticle protein genes, and DNA methylation (Supplementary Notes 7.3 and 10–12).

The first complete genome of a chelicerate species provides the opportunity for a detailed phylogenomic analysis of arthropods, the most diverse group of animals on Earth. The *T. urticae* genome illustrates the specialized life history of this polyphagous herbivorous pest.

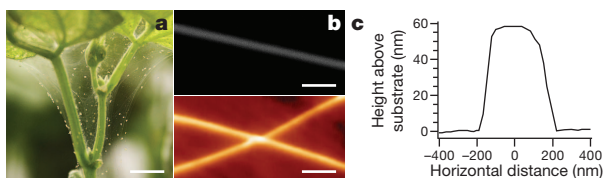


Figure 5 | *T. urticae* silk structure and dimensions. **a**, Spider mite colony on a bean plant forming characteristic silk webbing. **b**, SEM image of the spider mite larval silk filament (top), and atomic force microscopy (AFM) image of two larval spider mite silk filaments (bottom). **c**, Height profile of the adult spider mite silk filament obtained from the AFM image. Scale bars: **a**, 0.75 cm; **b**, 1 μ m.

Striking gene gains include lineage-specific expansions within detoxification gene families and lateral transfer of genes from fungi and bacteria that further expanded in *T. urticae*. The functional significance of these innovations is supported by the upregulation of many of these genes in response to feeding on less preferred host plants.

The genome of the two-spotted spider mite, together with the favourable biological features of the spider mite as a laboratory model including short generation time, easy rearing and tools for gene analysis and gene silencing³³, provide a novel resource for agriculture that should allow the dissection of pest–plant interactions and development of alternative tools for plant protection. Finally, evolutionary innovation in the process of *T. urticae* silk production expands the repertoire of potential chelicerate biomaterials (such as the well-known spider silk) with a natural biomaterial at the nanometre scale.

METHODS SUMMARY

All genomic sequencing reads were collected with standard Sanger sequencing protocols. RNA sequencing was performed with Illumina RNA-seq protocols. Annotation of the *T. urticae* genome was done using the gene prediction platform EuGene. The complete genome annotation is available at <http://bioinformatics.psb.ugent.be/webtools/bogas/>. The *T. urticae* (London) genome project was registered under the INSDC project ID 71041.

Received 30 June; accepted 17 October 2011.

- Edgecombe, G. D. Arthropod phylogeny: An overview from the perspectives of morphology, molecular data and the fossil record. *Arthropod Struct. Dev.* **39**, 74–87 (2010).
- Regier, J. C. *et al.* Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**, 1079–1083 (2010).
- Dunlop, J. A. & Selden, P. A. Calibrating the chelicerate clock: a paleontological reply to Jayaprakash and Hoy. *Exp. Appl. Acarol.* **48**, 183–197 (2009).
- Dunlop, J. A. Geological history and phylogeny of Chelicerata. *Arthropod Struct. Dev.* **39**, 124–142 (2010).
- Walter, D. E. & Proctor, H. C. *Mites: Ecology, Evolution and Behaviour* (CABI Publishing, 1999).
- Jeppson, L. R., Keifer, H. H. & Baker, E. W. *Mites Injurious to Economic Plants* (Univ. California Press, 1975).
- Gerson, U. in *Spider Mites: their Biology, Natural Enemies and Control* Vol. 1A (eds Helle, W. & Sabelis, M. W.) 223–232 (Elsevier, 1985).
- Boubou, A., Migeon, A., Roderick, G. K. & Navajas, M. Recent emergence and worldwide spread of the red tomato spider mite, *Tetranychus evansi*: genetic variation and multiple cryptic invasions. *Biol. Invasions* **13**, 81–92 (2011).
- Migeon, A. *et al.* Modelling the potential distribution of the invasive tomato red spider mite, *Tetranychus evansi* (Acari: Tetranychidae). *Exp. Appl. Acarol.* **48**, 199–212 (2009).
- Van Leeuwen, T., Vontas, J., Tsagkarakou, A., Dermauw, W. & Tirry, L. Acaricide resistance mechanisms in the two-spotted spider mite *Tetranychus urticae* and other important Acari: a review. *Insect Biochem. Mol. Biol.* **40**, 563–572 (2010).
- Khajehali, J., Van Nieuwenhuysse, P., Demaeght, P., Tirry, L. & Van Leeuwen, T. Acaricide resistance and resistance mechanisms in *Tetranychus urticae* populations from rose greenhouses in the Netherlands. *Pest Manag. Sci.* **67**, 1424–1433 (2011).
- BOGAS. Bioinformatics Gent Online Genome Annotation Service (<http://bioinformatics.psb.ugent.be/webtools/bogas/>) (2011).
- VectorBase. *Ixodes scapularis* Wikel annotation, IscaW1 (<http://iscapularis.vectorbase.org>) (2008).
- Lau, N. C. *et al.* Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line. *Genome Res.* **19**, 1776–1785 (2009).
- Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
- Richards, S. *et al.* The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**, 949–955 (2008).
- Oliver, J. H. Cytogenetics of mites and ticks. *Annu. Rev. Entomol.* **22**, 407–429 (1977).
- Kirkness, E. F. *et al.* Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl Acad. Sci. USA* **107**, 12168–12173 (2010).
- Cornman, S. R. *et al.* Genomic survey of the ectoparasitic mite *Varroa destructor*, a major pest of the honey bee *Apis mellifera*. *BMC Genom.* **11**, 602 (2010).
- Carrillo, L. *et al.* Expression of a barley cystatin gene in maize enhances resistance against phytophagous mites by altering their cysteine-proteases. *Plant Cell Rep.* **30**, 101–112 (2011).
- Feyereisen, R. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. *Biochim. Biophys. Acta* **1814**, 19–28 (2011).
- Guengerich, F. P., Wu, Z. L. & Bartleson, C. J. Function of human cytochrome P450s: characterization of the orphans. *Biochim. Biophys. Res. Commun.* **338**, 465–469 (2005).

23. Giraudo, M., Unnithan, G. C., Le Goff, G. & Feyereisen, R. Regulation of cytochrome P450 expression in *Drosophila*: Genomic insights. *Pestic. Biochem. Physiol.* **97**, 115–122 (2010).
24. Vaillancourt, F. H., Bolin, J. T. & Eltis, L. D. The ins and outs of ring-cleaving dioxygenases. *Crit. Rev. Biochem. Mol. Biol.* **41**, 241–267 (2006).
25. Moran, N. A. & Jarvik, T. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**, 624–627 (2010).
26. Veerman, A. & Helle, W. Evidence for functional involvement of carotenoids in photoperiodic reaction of spider mites. *Nature* **275**, 234 (1978).
27. Chung, J. S. Hemolymph ecdysteroids during the last three molt cycles of the blue crab, *Callinectes sapidus*: quantitative and qualitative analyses and regulation. *Arch. Insect Biochem. Physiol.* **73**, 1–13 (2010).
28. Grenier, J. K., Garber, T. L., Warren, R., Whittington, P. M. & Carroll, S. Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. *Curr. Biol.* **7**, 547–553 (1997).
29. Schwager, E. E., Schoppmeier, M., Pechmann, M. & Damen, W. G. Duplicated Hox genes in the spider *Cupiennius salei*. *Front. Zool.* **4**, 10 (2007).
30. Damen, W. G. M., Hausdorf, M., Seyfarth, E. A. & Tautz, D. A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *Proc. Natl Acad. Sci. USA* **95**, 10665–10670 (1998).
31. Mothes, U. & Seitz, K. A. Fine-structure and function of the prosomal glands of the 2-spotted spider-mite, *Tetranychus urticae* (Acari, Tetranychidae). *Cell Tissue Res.* **221**, 339–349 (1981).
32. Kluge, J. A., Rabotyagova, U., Leisk, G. G. & Kaplan, D. L. Spider silks and their applications. *Trends Biotechnol.* **26**, 244–251 (2008).
33. Grbic, M. *et al.* Mity model: *Tetranychus urticae*, a candidate for chelicerate model organism. *Bioessays* **29**, 489–496 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements M.G. and V.G. acknowledge support from NSERC Strategic Grant STPGP 322206-05, Marie Curie Incoming International Fellowship, OECD Co-operative Research Programme: Biological resource management for Sustainable Agricultural Systems JA00053351, and Ontario Research Fund–Global Leadership in Genomics and Life Sciences GL2-01-035. The genome and transcriptome sequencing projects were funded by the Government of Canada through Genome Canada and the Ontario

Genomics Institute (OGI-046), JGI Community Sequencing Program grant 777506 to M.G., a University of Utah SEED grant (to R.M.C.), and National Science Foundation (NSF) grant 0820985 (to R.M.C., Principal Investigator L. Sieburth); work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under contract No. DE-AC02-05CH11231. Y.V.d.P. acknowledges support from the Belgian Federal Science Policy Office IUAP P6/25 (BioMaGNet), the Fund for Scientific Research Flanders (FWO), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), and Ghent University (MRP N2N). T.V.L. is a post-doctoral fellow of the FWO. We acknowledge the work of J. Boore, T. Negrave, A. Migeon, P. Auger, L. Swevers and H. Van Langenhove. M.G. and V.G. thank D. Weigel, G. Schäfer, M. Gerberding, R. Sommer, J. Felix and T. Nuernberger for discussions and support. The genome annotation of *T. urticae* is available at the VIB Department of Plant Systems Biology, Ghent University (<http://bioinformatics.psb.ugent.be/webtools/bogas/>).

Author Contributions M.G. wrote the genome sequencing proposal. M.G. and V.G. generated DNA and RNA for sequencing. M.G., T.V.L., R.M.C., V.G., R.F. and Y.V.d.P. coordinated genome analysis and manuscript preparation. J.S. and E.L. coordinated genome assembly whereas P.R. and S.R. centralized and enabled the annotation process. All other authors are members of the spider mite genome sequencing consortium and contributed annotation, analyses or data to the genome project. M.G., T.V.L. and R.M.C. should be considered joint first authors. S.R., P.R. and V.G. should be considered joint second authors. M.G., R.F. and Y.V.d.P. should be considered joint corresponding authors.

Author Information Individual scaffolds of the *T. urticae* (London) genome are available through GenBank under accession numbers HE587301 to HE587940. The Illumina data for *T. urticae* strain Montpellier can be found in the Sequence Read Archive (SRA) database under accession numbers SRX030911 to SRX030913. RNA-seq data is available under Gene Expression Omnibus (GEO) super series number GSE32342. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.V.d.P. (yves.vandeppeer@psb.ugent.be) or M.G. (mgrbic@uwyo.ca).