Multiple horizontal acquisitions of plant genes in the whitefly *Bemisia tabaci*

Clément Gilbert[1*] and Florian Maumus[2*]

* equal contribution

1- Université Paris-Saclay, CNRS, IRD, UMR Evolution, Génomes, Comportement et Ecologie, Gif-sur-Yvette, 91198, France
2- Université Paris-Saclay, INRAE, URGI, 78026, Versailles, France

Emails: clement.gilbert@egce.cnrs-gif.fr and florian.maumus@inrae.fr

Competing interest statement: All authors declare to have no conflict of interest

Abstract

The extent to which horizontal gene transfer (HGT) has shaped eukaryote evolution remains an open question. Two recent studies reported four plant-like genes acquired through two HGT events by the whitefly *Bemisia tabaci*, a major agricultural pest (Lapadula et al. 2020; Xia et al. 2021). Here, we performed a systematic search for plant-to-insect HGT in *B. tabaci* and uncovered a total of 50 plant-like genes deriving from at least 24 independent HGT events. Most of these genes are present in three cryptic *B. tabaci* species, show high level of amino-acid identity to plant genes (mean = 64%), are phylogenetically nested within plant sequences, and are expressed and evolve under purifying selection. The predicted functions of these genes suggest that most of them are involved in plant-insect interactions. Thus, substantial plant-to-insect HGT may have facilitated the evolution of *B. tabaci* towards adaptation to a large host spectrum. Our study shows that eukaryote-to-eukaryote HGT may be relatively common in some lineages and it provides new candidate genes that may be targeted to improve current control strategies against whiteflies.

Introduction

Horizontal gene transfer (HGT) is the passage of genetic material between organisms by means other than reproduction. The patterns, mechanisms and vectors of HGT are well-characterized in prokaryotes, in which these transfers are ubiquitous and a major source of innovation (1). In eukaryotes, HGT has long been considered anecdotal because of multiple barriers that should impede such transfers, or controversial, as resulting from phylogenetic artifacts or contaminant sequences (2). Yet, dozens of recent studies have reported robust HGT events in various eukaryotic organisms, sometimes with demonstrated impact on host phenotype (3, 4). For example, two studies reported plant genes (two BAHD acyltransferases called BtPMaT1 and BtPMaT2 and ribosome inactivating proteins [RIPs]) acquired by horizontal transfer by the sweet potato whitefly *Bemisia tabaci* (5, 6). Whiteflies (family Aleyrodidae) are herbivorous hemipteran insects that are important agricultural pests because of their feeding habits and the many viruses they transmit to plants. Functional assays revealed that the BtPMaT1 protein detoxifies plant phenolic glucosides that are normally used by plants to protect themselves against insect herbivores. Thus, the acquisition of a plant gene by whiteflies through HT enabled them to thwart plant defenses and may in part explain why these insects have become generalist phloem-feeders (6). Interestingly, whiteflies feeding on transformed tomato plants expressing BtPMaT1 gene-silencing fragments showed increased mortality and reduced fecundity (6). Therefore, identifying genes of plant origin in herbivore insects can provide targets to engineer new pest control strategies.

Results and discussion

To assess the extent to which plant-to-insect HGT may have shaped interactions between whiteflies and their host plants, we conducted a systematic search for genes of plant origin in the *B. tabaci* MEAM1 genome (7). To this end, we performed last common ancestor (LCA) prediction for each *B. tabaci* predicted protein using MMseqs2 taxonomy (8) against the Uniref90 database from which Aleyrodidae proteins were excluded (Supplementary Materials and Methods). In this approach, the placement of a protein LCA in plants (Viridiplantae) indicates potential HGT from plants to *B. tabaci* after the emergence of Aleyrodidae. The deeper the LCA, the more conserved the protein across the breadth of plant taxonomy. In total, we identified 65 *B. tabaci* proteins for which LCA was inferred deeper than genus level in plants and which were hence considered potentially transferred from plants to Aleyrodidae. This included the two BAHD acyltransferase genes (BtPMaT1 and BtPMaT2) reported by Xia et al. (6) and the two RIPs identified by Lapadula et al. (5).

To further characterize the taxonomic origin of candidate *B. tabaci* plant-derived genes, we grouped them in 41 clusters of orthologous proteins and used one representative sequence per cluster to perform similarity searches against all Aleyrodidae proteomes as well as against the UniRef90 protein database of streptophytes, that of eukaryotes without streptophytes and Aleyrodidae, and those of prokaryotes and viruses (Supplementary Materials and Methods). The ten best hits from each search supplemented with the twenty most diverse hits against streptophytes were used to build a multiple alignment which was submitted to phylogenetic analysis. Remarkably, while the phylogenetic signal observed in 17 out of 41 trees was in disagreement with plant-to-Aleyrodidae gene transfer (data not shown), the topology of 24 trees was consistent with this scenario. In 22 of the 24 trees, the *B. tabaci* proteins were nested within clades of proteins of streptophyte origin (Figure 1A and Supplementary Dataset), as in (5, 6). In the remaining two trees (Bta13103 and Bta14885), the *B. tabaci* proteins are sister to all proteins of streptophyte origin (Figure 1B). The topology of these two trees could be explained either by extensive sequence adaptation after transfer or by donor plant lineage being extinct or under-represented in genome databases. In support of a plant origin, the level of amino-acid (aa) similarity between the corresponding 24 *B. tabaci* representative proteins and their plant

homologues varied from 28% to 84% (mean=64%; Table 1) and homologous genes are virtually absent from the proteomes of non-Aleyrodidae insects. In addition, 23 clusters are shared between MEAM1 and at least one other *B. tabaci* cryptic species (MED1 and SSA-ECA) independently sequenced in different laboratories, indicating that they are *bona fide B. tabaci* genes, and not contaminating plant genes. Altogether, these evidences indicate that as proposed for BAHD acyltransferases and RIPs (5, 6), these 24 representative genes were acquired by an ancestor of *B. tabaci* through plant-to-insect HGT.

In total, the 24 corresponding Aleyrodidae orthogroups encompass 138 proteins, of which 133 have best hits against plant homologs including 46 from SSA-ECA, 35 from MED, and 50 from MEAM1 *B. tabaci* cryptic species as well as two from the outgroup species *Trialeurodes vaporariorum* (Tv) (Supplementary Dataset). Protein clustering combined to phylogenetic analysis shows that at least 24 independent HGT events are necessary to explain the presence of these genes in *B. tabaci* genomes. This number of plant-to-insect HGT is remarkable given that most HGT events reported so far in animals (excluding HT of transposable elements) involve genes of microbial origins, with only few cases of gene transfers from eukaryotes to animals (9). It is noteworthy that fifteen out of twenty clusters are shared between the MEAM1 and the SSA-ECA cryptic species, indicating that most HGTs likely occurred before the split of these species, i.e., between 19 and 40 million years ago (10, 11). Several transfers were apparently followed by gene duplications. The largest cluster, with predicted delta(12) fatty acid desaturase (FAD) function, comprises 38 Aleyrodidae members including 15 MEAM1 proteins. Eight of the FAD genes are organized in a genomic region spanning about 120kb (scaffold 995: 912024-1039680) which is also observed in syntenic position at least in the SSA-ECA genome (scaffold 436:909978- 1039696), indicating that this hotspot evolved before the split of these species. In combination with phylogenetic analysis, the amplification of the FAD genes can be explained by a mixture of local and distal duplication events post HGT (Figure 1A). We also note that, in the FAD tree, the plant-related *B. tabaci* proteins are distributed on two distinct branches and we cannot rule out that it could reflect two independent transfer events of FAD genes (Figure 1A).

To search for evidence of functionality of plant-like genes in *B. tabaci*, we measured the evolutionary pressures acting on them by calculating pairwise ratios of non-synonymous over synonymous substitutions (Ka/Ks) between sequences within each of the 24 clusters. We found that 593 out of the 612 ratios we obtained were lower than 0.5, indicating that most if not all plant-like genes are evolving under purifying selection. Furthermore, we found transcripts supporting expression of at least one gene for 18 out of the 24 clusters, often in multiple independent transcriptome assemblies (Table 1). Together with conservation across cryptic species and evidence of gene duplication, these data indicate that most if not all *B. tabaci* plant-like genes are functional.

Interestingly, most of these genes (21 out of 24 clusters) have a predicted function based on similarity with their nearest plant relative. As for the recently described malonyltransferase (6, 12), there is direct or indirect evidence that many of them are involved in plant-pathogen interactions (Table 1). For example, the delta(12) FAD are known to produce polyunsaturated linoleic acid in plants, which are involved in response to pathogens (13). Likewise, members of subtilisin-like protease and pathogen-related protein families can both be induced following pathogen infection (14, 15). Ornithine decarboxylase is also worth noting as this enzyme synthesizes putrescine in plants, a polyamine involved in pathogen response (16) and suspected to be usurped by Hessian fly larvae to facilitate nutrient acquisition while feeding on wheat (17). In the same vein, a gene resembling the nicotianamine synthase, involved in the transport of various metal ions in plants (18), may facilitate acquisition of micronutrient by whiteflies. Finally, pectinesterase is a plant cell wall degrading enzyme (PCWDE) that is also found in plant and fungal pathogens causing maceration and

soft-rotting of plant tissues. In fact, horizontal acquisition of PCWDE has already been documented in insects, but the source of the gene was bacterial (19).

To conclude, our study reveals that in addition to bacterial genes, which repeatedly entered arthropod genomes and fueled the evolution of herbivory (20), numerous plant genes have been acquired through HGT by *B. tabaci*, likely contributing to the highly polyphagous nature of this species. The significant representation of predicted functions potentially involved in parasitism suggests that these genes were selected from an important set of transferred genes. Using the same approach on *Drosophila melanogaster*, we found no gene of potential plant origin, showing that plant-to-insect HGT is not ubiquitous and suggesting that it could be facilitated by specific vectors in association with *B. tabaci*. It is noteworthy that viruses have been proposed to act as vectors of HT in eukaryotes (21, 22) and that *B. tabaci* is known to act as a vector of dozens of plant viruses, some of which are able to replicate in insect cells (23). Our results call for a large-scale evaluation of plant-to-insect HGT and for a detailed functional characterization of *B. tabaci* plant-like genes, which may further contribute to control this pest.

Data availability

Supplementary Dataset (on figshare repository): this folder contains a table listing all transcripts covering at least 75% of the *B. tabaci* MEAM1 plant-derived genes with at least 95% identity, a table providing features of all *B. tabaci* plant-derived genes included in this study and references supporting involvement in plant-pathogen interactions, Supplementary Material and Methods (direct link), a fasta file combining the plant-derived MEAM1 representative sequences, an archive comprising the initial (.aln files) and trimmed (.trim files) protein alignments in fasta format as well as the phylogenetic trees in Newick format (.annot files) for each cluster of plant-derived genes, and the tree images combined in a single file (.pdf). The phylogenetic trees can be viewed interactively at https://itol.embl.de/shared/fmaumus.

Author contributions

CG and FM contributed equally to perform analyses, interpret the results and write the manuscript. FM conceived the study.

Table 1. Characteristics of representative sequences of each cluster of plant-derived *Bemisia tabaci* genes.

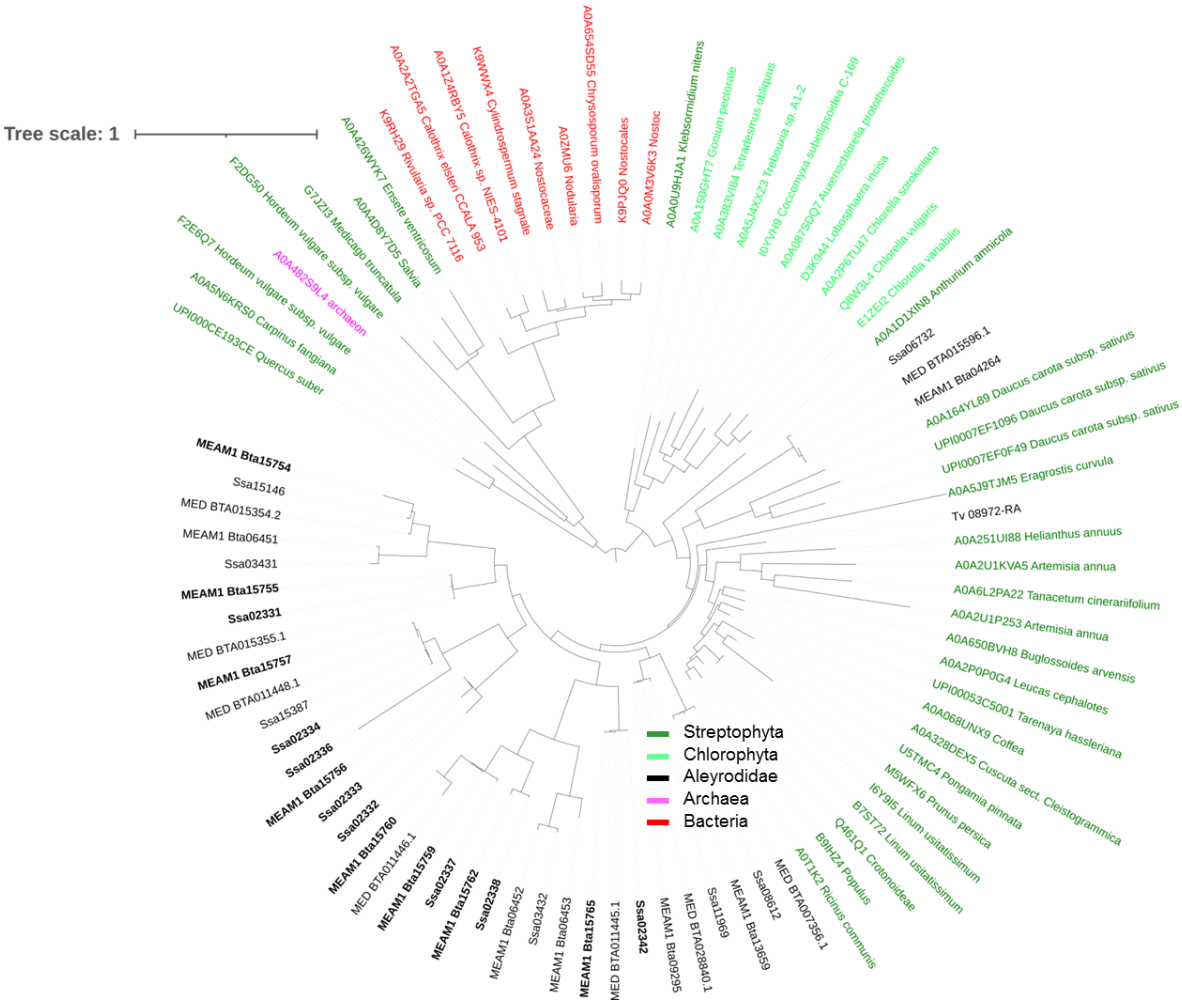
MEAM1 gene representative of cluster at 35% ID	% ID with best plant hit	Nb of seq. in cluster	Nb of Ka/Ks<0.5	Nb of TSA datasets	Nb of TSA seq. (>75% cov; >95 ID4)	Predicted function*
Bta01560	57.8	9	35/35 (9)	1	1	BAHD acyltransferase
Bta05675	63.3	5	6/6 (4)	-	-	ornithine decarboxylase
Bta14372	61.4	6	14/15 (6)	-	-	HXXXD-type acyl-transferase
Bta14387	84.3	3	3/3 (3)	8	13	pathogen-related protein
Bta14404	76.4	2	1/1 (2)	-	-	Glutelin
Bta14885	61.8	5	5/6 (4)	6	7	beta-(1,2)-xylosyltransferase
Bta01681	70.2	9	36/36 (9)	7	8	methyltransferase DDB
Bta07485	66.5	5	6/6 (4)	4	5	purple acid phosphatase
Bta07668	70.4	3	3/3 (3)	-	-	AAA-ATPase ASD
Bta07786	39.7	6	15/15 (6)	4	6	phenolic glucoside malonyltransferase
Bta08104	57.6	3	2/3 (3)	9	9	subtilisin-like protease/Pectinesterase
Bta04871	63.2	3	3/3(3)	5	8	Nicotianamine synthase
Bta08851	76.1	4	3/3 (3)	4	5	alpha-(1,4)-fucosyltransferase
Bta11221	60.5	4	5/6 (4)	-	-	Pectinesterase
Bta06453	58.7	38	423/435 (30)	6	61	delta(12) fatty acid desaturase
Bta11576	70.3	1	NA	8	15	delta(7)-sterol-C5(6)-desaturase
Bta07117	65.9	2	0/1 (2)	9	10	unknown function
Bta11842	80.4	2	1/1 (2)	8	8	unknown function
Bta13103	28.7	5	9/10 (5)	7	13	Ricin-like
Bta02746	49.5	4	6/6 (4)	1	2	unknown function
Bta06115	72.3	4	6/6 (4)	6	9	glucan endo-1,3-beta-glucosidase
Bta15659	74.5	3	3/3 (3)	3	6	Beta-1,4-N-acetylglucosaminyltransferase
Bta13961	64.5	4	5/6 (4)	5	6	Thaumatin
Bta08293	60	3	3/3 (3)	-	-	chalcone_synthase

*References supporting a potential role in host-pathogen interaction are provided in Supplementary Dataset

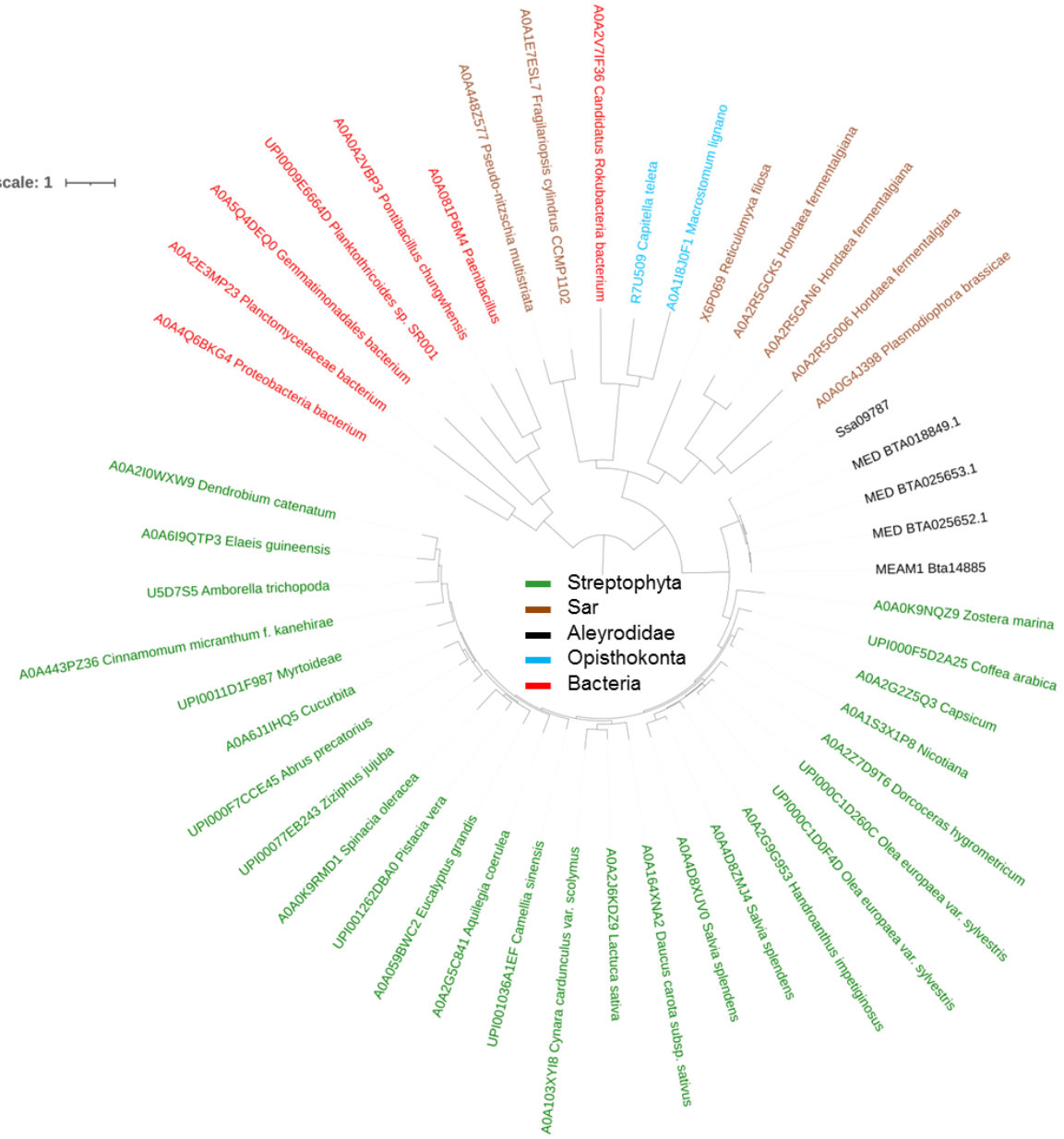
Figures

Figure 1 – Phylogenetic relationships of potential plant-derived *Bemisia tabaci* genes

A)



B)



Phylogenetic trees showing the Bta06453 cluster, annotated as delta(12) fatty acid desaturase family (A) and the Bta14885 cluster, annotated as beta-(1,2)-xylosyltransferase family (B). Each tree is inferred from the multiple sequence alignment of the cluster representative sequence with its homologs found across Aleyrodidae proteins and Uniref90 proteins from eukaryotes, prokaryotes and viruses. The labels for Uniref90 proteins are either species names or the names of the first taxonomic level comprising all the proteins of a specific Uniref90 cluster. The SSA-ECA proteins are labelled as « Ssa ». In (A), the labels of the *B. tabaci* proteins present in the FAD genomic hotspots found in the MEAM1 and SSA-ECA genomes are in bold and the *Trialeurodes vaporariorum* protein is labelled as "Tv". The color legend indicating the taxonomy of proteins is indicated inset. For the ease of representation, the trees are presented in a circular mode rooted at midpoint. Unrooted trees are available in the Supplementary dataset.

References

1. S. M. Soucy, J. Huang, J. P. Gogarten, Horizontal gene transfer: building the web of life. *Nature Reviews: Genetics* **16**, 472–82 (2015).

2. W. F. Martin, Too Much Eukaryote LGT. *BioEssays* **39**, 1700115 (2017).

3. J. Van Etten, D. Bhattacharya, Horizontal Gene Transfer in Eukaryotes: Not if, but How Much? *Trends in Genetics* **36**, 915–925 (2020).

4. M. M. Leger, L. Eme, C. W. Stairs, A. J. Roger, Demystifying Eukaryote Lateral Gene Transfer (Response to Martin 2017 DOI: 10.1002/bies.201700115). *BioEssays* **40**, 1700242 (2018).

5. W. J. Lapadula, M. L. Mascotti, M. Juri Ayub, Whitefly genomes contain ribotoxin coding genes acquired from plants. *Sci Rep* **10**, 15503 (2020).

6. J. Xia, *et al.*, Whitefly hijacks a plant detoxification gene that neutralizes plant toxins. *Cell* **184**, 1693-1705.e17 (2021).

7. W. Chen, *et al.*, The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biol* **14**, 110 (2016).

8. M. Mirdita, M. Steinegger, J. Söding, MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858 (2019).

9. L. A. Graham, P. L. Davies, Horizontal Gene Transfer in Vertebrates: A Fishy Tale. *Trends in Genetics* **37**, 501–503 (2021).

10. D. Santos-Garcia, C. Vargas-Chavez, A. Moya, A. Latorre, F. J. Silva, Genome Evolution in the Primary Endosymbiont of Whiteflies Sheds Light on Their Divergence. *Genome Biology and Evolution* **7**, 873–888 (2015).

11. H. Mugerwa, *et al.*, African ancestry of New World, Bemisia tabaci-whitefly species. *Sci Rep* **8**, 2734 (2018).

12. G. Taguchi, *et al.*, Malonylation is a key reaction in the metabolism of xenobiotic phenolic glucosides in Arabidopsis and tobacco: Phenolic-xenobiotics metabolism in Arabidopsis. *The Plant Journal* **63**, 1031–1041 (2010).

13. A. A. Dar, A. R. Choudhury, P. K. Kancharla, N. Arumugam, The FAD2 Gene in Plants: Occurrence, Regulation, and Role. *Front. Plant Sci.* **8**, 1789 (2017).

14. L. C. van Loon, M. Rep, C. M. J. Pieterse, Significance of Inducible Defense-related Proteins in Infected Plants. *Annu. Rev. Phytopathol.* **44**, 135–162 (2006).

15. A. Figueiredo, F. Monteiro, M. Sebastiana, Subtilisin-like proteases in plantâ€"pathogen recognition and immune priming: a perspective. *Front. Plant Sci.* **5** (2014).

16. H. J. Rhee, E.-J. Kim, J. K. Lee, Physiological polyamines: simple primordial stress molecules. *J Cellular Mol Med* **11**, 685–703 (2007).

17. S. Subramanyam, *et al.*, Hessian fly larval feeding triggers enhanced polyamine levels in susceptible but not resistant wheat. *BMC Plant Biol* **15**, 3 (2015).

18. C. Laffont, P. Arnoux, The ancient roots of nicotianamine: diversity, role, regulation and evolution of nicotianamine-like metallophores. *Metallomics* **12**, 1480–1493 (2020).

19. M. Shelomi, *et al.*, Horizontal Gene Transfer of Pectinases from Bacteria Preceded the Diversification of Stick and Leaf Insects. *Sci Rep* **6**, 26388 (2016).

20. N. Wybouw, Y. Pauchet, D. G. Heckel, T. Van Leeuwen, Horizontal Gene Transfer Contributes to the Evolution of Arthropod Herbivory. *Genome Biol Evol* **8**, 1785–801 (2016).

21. M. Catoni, *et al.*, Virus-mediated export of chromosomal DNA in plants. *Nature Communications* **9** (2018).

22. C. Gilbert, R. Cordaux, Viruses as vectors of horizontal transfer of genetic material in eukaryotes. *Current Opinion in Virology* **25**, 16–22 (2017).

23. Y.-Z. He, *et al.*, A plant DNA virus replicates in the salivary glands of its insect vector via recruitment of host DNA synthesis machinery. *Proc Natl Acad Sci USA* **117**, 16928–16937 (2020).