An aphid RNA transcript migrates systemically within plants and is a virulence factor

Yazhou Chen[a,1], Archana Singh[b,1], Gemy G. Kaithakottil[b], Thomas C. Mathers[a], Matteo Gravino[a], Sam T. Mugford[a], Cock van Oosterhout[c], David Swarbreck[b], and Saskia A. Hogenhout[a,2]

[a]Department of Crop Genetics, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom; [b]Earlham Institute, Norwich Research Park, Norwich, NR4 7UH, United Kingdom; and [c]School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

Aphids are sap-feeding insects that colonize a broad range of plant species and often cause feeding damage and transmit plant pathogens, including bacteria, viruses, and viroids. These insects feed from the plant vascular tissue, predominantly the phloem. However, it remains largely unknown how aphids, and other sap-feeding insects, establish intimate long-term interactions with plants. To identify aphid virulence factors, we took advantage of the ability of the green peach aphid *Myzus persicae* to colonize divergent plant species. We found that a *M. persicae* clone of near-identical females established stable colonies on nine plant species of five representative plant eudicot and monocot families that span the angiosperm phylogeny. Members of the novel aphid gene family *Ya* are differentially expressed in aphids on the nine plant species and are coregulated and organized as tandem repeats in aphid genomes. Aphids translocate *Ya* transcripts into plants, and some transcripts migrate to distal leaves within several plant species. RNAi-mediated knockdown of *Ya* genes reduces *M. persicae* fecundity, and *M. persicae* produces more progeny on transgenic plants that heterologously produce one of the systemically migrating *Ya* transcripts as a long noncoding (lnc) RNA. Taken together, our findings show that beyond a range of pathogens, *M. persicae* aphids translocate their own transcripts into plants, including a *Ya* lncRNA that migrates to distal locations within plants, promotes aphid fecundity, and is a member of a previously undescribed host-responsive aphid gene family that operate as virulence factors.

plant–insect interactions | aphids | lncRNA | transkingdom RNA trafficking | virulence

S ap-feeding insects of the order Hemiptera include aphids, whiteflies, leafhoppers, and planthoppers that have piercing-sucking mouthparts, termed stylets, for feeding. Many species cause direct feeding damage, known, for example, as hopper burn (1, 2), although global economic yield losses caused by these insects are most often due to their abilities to transmit a diverse range of plant pathogens that include viruses, bacteria, and plasmodium-like organisms, as well as naked RNA molecules known as viroids (3–8). The majority of insect herbivores are specialized to feed on one or a few closely related plant species (9); however, some hemipteran insects are polyphagous. These include the green peach aphid *Myzus persicae*, which is known to reproduce on >400 plant species and is also able to transmit divergent plant pathogens, including >100 plant viruses (10) and the potato spindle viroid (11). The factors involved in the ability of sap-feeding insects to establish intimate interactions with their plant hosts remain largely unknown.

Hemipterans use their stylets to feed on plant sap, often from the phloem or xylem of the plant vascular tissue. How sap-feeding insects move their stylets within plant tissues to reach the vascular tissues is arguably best investigated in aphids. These insects establish a long-term feeding site in the phloem sieve cells. However, before reaching the phloem, the stylets probe epidermis, mesophyll, and other cells, with each probe consisting of a short period of cell content ingestion, often referred to as "tasting," followed by

a short period of salivation (12). As soon as the stylets reach the phloem, aphids deposit saliva into the sieve cells, followed by long periods of ingestion of phloem sap (13). The saliva introduced into cells is known as "watery" saliva and is rich in proteins (14), some of which were shown to be effectors that modulate plant processes (15–21). The cycles of tasting and salivation during the stylets path to the phloem are likely to help aphids perceive and adjust to their hosts.

We previously found that an asexually reproducing (parthenogenic) *M. persicae* colony consisting of largely genetically identical females can adjust to the divergent plant species *Brassica rapa*, *Arabidopsis thaliana*, and *Nicotiana benthamiana* via differential coregulation of tandemly repeated gene families, including that of *Cathepsin B* (*CathB*), virulence factors that optimize the ability of *M. persicae* to colonize specific plant species (22). *CathB* genes are also differentially regulated in aphids on healthy plants versus those on virus-infected plants (23), and viruses are known to modulate plant defense responses (24, 25). These studies provide evidence that *M. persicae* is a truly polyphagous/generalist insect species, as this insect has the ability to colonize plant species from

Significance

The green peach aphid *Myzus persicae* causes yield losses of diverse crops and vectors more than 100 pathogens. We found that *M. persicae* colonizes nine divergent plant species, indicating that it is a true generalist, unlike many other aphid species that have specialized to colonize a few plant species. Members of the aphid gene family *Ya* undergo coordinated expression changes in *M. persicae* depending on the plant species. The aphids translocate *Ya* transcripts into plants, and these RNAs move to distal leaves. Moreover, aphid reproduction increases on plants that stably produce the *M. persicae Ya1* RNA. Therefore, beyond transmitting a range of pathogens, *M. persicae* translocate their own transcripts into plants, including an RNA that promotes aphid performance.

AGRICULTURAL SCIENCES

different plant families and apparently can perceive the host plant status and adjust its gene expression accordingly.

Here we build on our previously reported data, taking advantage of the ability of *M. persicae* to colonize divergent plant species to better understand how aphids colonize plants. We show that a single parthenogenic *M. persicae* clone establishes stable colonies on nine plant species from five plant families. Via weighted gene coexpression network analysis (WGNCA) (26), we identified the *Ya* gene family. Members of this family are organized as tandem repeats in the *M. persicae* genome and adjust their gene expression in a coordinated manner in response to the different plant species. *Ya* transcripts translocate into plants during aphid feeding and migrate systemically, and the *M. persicae Ya1* transcript promotes aphid fecundity as a long noncoding (lnc) RNA when produced in plants. Our work shows evidence that the establishment of parasitic interactions between divergent organisms involves translocation of an lncRNA virulence factor.

Results

M. persicae Establishes Stable Colonies on Nine Plant Species That Span the Angiosperm Phylogeny.
To establish *M. persicae* colonies on divergent plant hosts, we selected plant species of representative plant families across the angiosperm (flowering plants) phylogeny (27), including *B. rapa* (Br) and *A. thaliana* (At) of the Brassicaceae, *N. benthamiana* (Nb) and *Solanum tuberosum* (St) of the Solanaceae, *Pisum sativum* (Ps) and *Phaseolus vulgaris* (Pv) of the Fabaceae, *Helianthus annuus* (Ha) and *Chrysanthemum indicum* (Ci) of the Asteraceae, and the monocot plant species, *Zea mays* (Zm) of the Poaceae (Fig. 1*A*). Individuals of asexually reproducing females of *M. persicae* clone O that had been maintained on Br for at least 3 y were transferred to Br (control) and At, Nb, St, Ps, Pv, Ha, Ci, and Zm (Fig. 1*A*). *M. persicae* clone O achieved a 100% survival rate and established stable colonies on these plant species (*SI Appendix*, Fig. S1 *A*–*D*). Aphids survived equally well on Br and At from the start and achieved a 100% survival rate at ~4 wk on Nb, St, Ps, Ci, and Zm and at ~10 wk on Pv and Ha.

Genes Differentially Expressed in *M. persicae* on Nine Plant Species Are Enriched for Salivary Gland and Candidate Long Noncoding RNAs.
To identify *M. persicae* genes that change expression levels on different plant species (i.e., host-responsive genes), we generated RNA-seq data from stable *M. persicae* colonies on nine hosts (five biological replicates each). We also used these RNA-seq data and previous RNA-seq data (LIB1777) (22) to improve the annotation. A total of 1.38 billion RNA-seq reads were assembled into transcripts using a genome-guided approach (*SI Appendix*, Table S1; GSE accession no. 129667). We identified a total of 45,972 transcripts (corresponding to 19,556 genes), including 30,127 transcripts (65.5%; 18,529 genes) annotated previously (22) and 15,845 (34.5%) additional transcripts (*SI Appendix*, Fig. S2*A*). Of the 45,972 transcripts, 6,581 (3,025 genes) had low protein-coding potential and were assigned candidate lncRNAs (*SI Appendix*, Fig. S2*B*).

To identify host-responsive genes, transcriptomes of *M. persicae* colonies on At, Nb, St, Ps, Pv, Ha, Ci, and Zm were compared with those of the colonies on Br (original host). This resulted in the identification of 2,490 (1,984 genes) that were significantly differentially expressed (DE) in the aphids on at least one of the other eight host plant species (fold change ≥ 2, $P \leq 0.05$, false discovery rate [FDR] $\leq 5\%$) (Dataset S1), and the majority of these show host-specific expression patterns (Fig. 1*B*). The DE transcripts were enriched in functions of oxidation-reduction processes, proteolysis (including *CathB*), and sensory perception of taste (*SI Appendix*, Fig. S3).

Identification of genes with expression levels that are highly correlated may help shed light on shared biological processes. Therefore, we performed WGCNA using 11,824 *M. persicae* genes (out of 19,556 total) that have mean gene expression values of >5

transcripts per million (TPM) across five biological replicates per hosts. This identified 77 coexpression modules comprising 7,864 genes (*SI Appendix*, Fig. S4 and Dataset S2). Of the 1,984 genes that we identified as DE above, 1,364 (68%) were included among the genes of the 77 coexpression modules, and 13 modules were enriched for DE genes (313 DE genes of the total DE genes [16%]; $P < 0.05$, Fisher's exact test) (Fig. 1*C* and Dataset S2). Heatmaps based on normalized TPM values of these 13 modules showed different expression patterns in aphids depending on the plant host (Fig. 1*D* and *SI Appendix*, Fig. S5), suggesting that *M. persicae* coordinates the expression of groups of genes in response to the plant species.

Four modules (darkslateblue, darkorange2, lightcoral, and thistle2) among the 13 enriched for DE genes were also enriched for candidate lncRNAs (Fig. 1*C* and Dataset S2). In addition, 8 out of 13 modules, including the darkslateblue, were enriched for *M. persicae* genes expressed in the salivary glands (28) (Fig. 1*C* and Dataset S2), and 3 of 13 modules were enriched for gut-expressed genes (*SI Appendix*, Fig. S6 and Dataset S2). We also found that 12 out of 13 modules included genes lying adjacent to each other as tandem repeats within the aphid genome (Dataset S2). The number of gene repeats varied from two to eight, with the latter in the darkslateblue module. Thus, different attributes were enriched in the darkslateblue module, including expression in the salivary glands, candidate lncRNAs, and tandemly repeated genes. The darkslateblue module is the only one among the 77 modules that contained two groups of genes with exact opposite host-responsive expression patterns (Fig. 1*E*). Therefore, genes in this module may be involved in shared biological processes and were further investigated.

### *M. persicae* Preferentially Translocates RNA Transcripts of DE and Candidate lncRNA Genes into Plants.
Because DE salivary gland genes that encode candidate lncRNAs were enriched, we investigated whether these transcripts translocate into plants when aphids feed. Leaves of 4-wk-old *A. thaliana* plants were caged with 20 adult aphids (feeding site) or no aphids (control) for 24 h (Fig. 2*A* and *SI Appendix*, Table S2). Reads that uniquely aligned to the *M. persicae* genome were identified, and at TPM ≥ 50, these corresponded to between 1,837 and 3,154 aphid transcripts, depending on the biological replicate, in leaf sections exposed to aphids only (Dataset S3). Based on the presence in at least three biological replicates, 3,186 *M. persicae* transcripts corresponding to 5% of the *M. persicae* transcriptome were found in the feeding site. These transcripts included messenger RNAs (2,985 transcripts) and candidate lncRNAs (201 transcripts) (Fig. 2*A*). The candidate lncRNAs were enriched among the transcripts found in the feeding sites (*SI Appendix*, Fig. S7). The candidate lncRNAs in the feeding sites belonged to 10 of the 13 coexpression modules that were found to be enriched for the aforementioned DE genes, one of which was the darkslateblue module (Fig. 2*B*). Moreover, the *M. persicae* transcripts in the feeding sites were enriched for DE genes and salivary gland transcripts ($P = 3.7E-13$ and $P = 0.03$, respectively, Fisher's exact test), and in both categories the transcripts were also enriched for candidate lncRNAs ($P = 0.002$ and $P = 3.8E-35$, respectively, Fisher's exact test) (Fig. 2*C*). Therefore, *M. persicae* preferentially translocates candidate lncRNAs of salivary gland-expressed genes that are DE in aphids on divergent hosts into feeding sites.

The Coregulated *M. persicae* Genes of the Darkslateblue Module Include All 30 Members of the *Ya* Family Tandem Repeat.
On further analyses of genes in the darkslateblue module, 37 out of total of 56 genes were found to encode candidate lncRNAs. Of these 37, 23 were identified to have sequence identities of ±80% and located in close proximity to one another on six scaffolds. These genes were organized in series of tandem repeats with <1-kb distance between
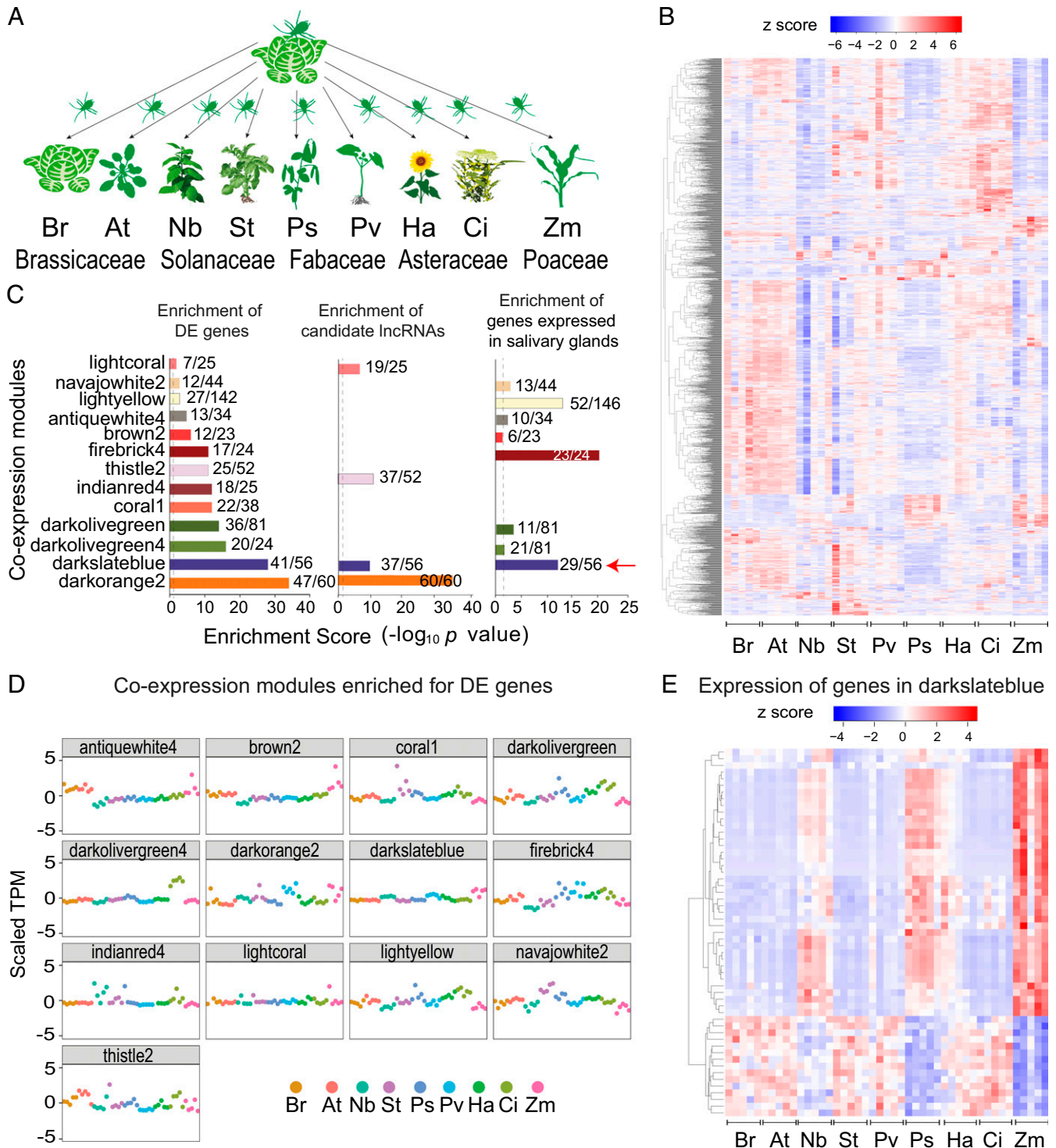
Fig. 1. Colonization of *M. persicae* on divergent plant species involves coexpression of genes enriched for lncRNAs and those expressed in the aphid salivary glands. (*A*) Schematic overview of the experimental setup with *B. rapa* (Br), *A. thaliana* (At), *N. benthamiana* (Nb), *S. tuberosum* (St), *C. indicum* (Ci), *H. annuus* (Ha), *P. sativum* (Ps), *P. vulgaris* (Pv), and *Zea mays* (Zm). (*B*) Heatmaps of log-transformed TPM values of 1,984 DE genes of *M. persicae* colonies on Br versus colonies on one of the other eight plant species as shown in *A* at five biological replicates per host. (*C*) Enrichment scores of 13 WGNCA modules. The *x*/*y* values above the bars indicate gene numbers in the enriched category (*x*) and module (*y*). Dashed vertical lines represent enrichment score 2 (*P* = 0.01). The red arrow indicates the darkslateblue module. (*D*) Scatterplots of 13 WGCNA modules enriched for DE genes. (*E*) Heatmap of log-transformed TPM values of genes in the darkslateblue module.

Chen et al.
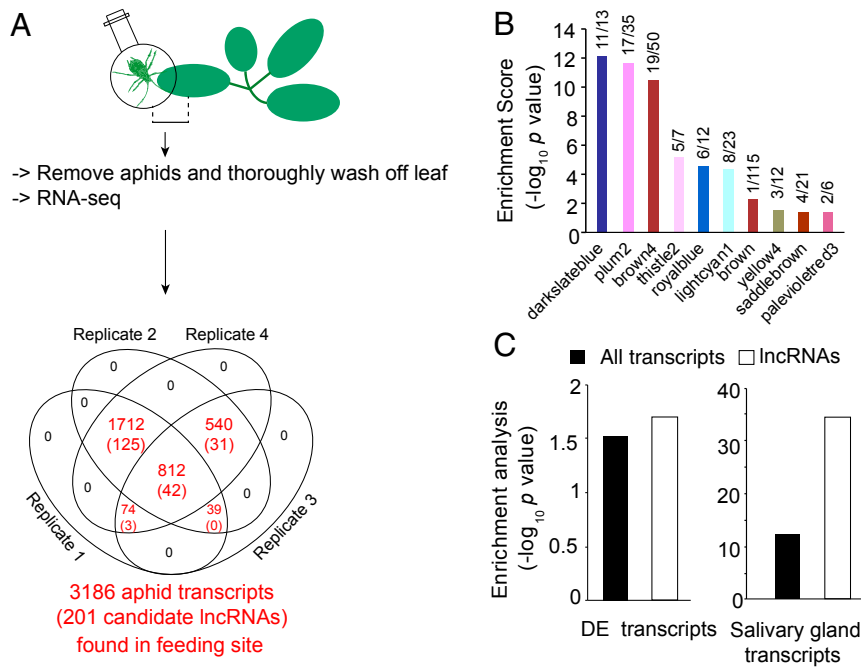
Fig. 2. Aphid transcripts in the feeding site are enriched for candidate lncRNAs and transcripts derived from genes expressed in aphid salivary glands and that are DE among aphids on nine plant species. (*A*) Schematic overview of the experimental setup. The Venn diagram shows the number of aphid transcripts found in plants at the aphid feeding site at >50 TPM in one of the replicates and/or the presence of transcripts in at least three biological replicates. The numbers in parentheses are the number of lncRNAs. (*B*) Transcripts in the feeding site were enriched for candidate lncRNAs in 10 of the 13 coexpression modules enriched for DE genes (Fig. 1*D*). The *x/y* values above the bars indicate lncRNAs (*x*) and the total number of transcripts belonging to the module found in the feeding site (*y*). (*C*) Enrichment of genes that are DE (*Left*) and expressed in salivary glands (28) (*Right*) of aphid transcripts found in the plant at aphid feeding. Black bars represent all aphid transcripts; white bars, candidate lncRNAs.

the gene copies. Given these characteristics, the genes appear to belong to a gene family.

Tandemly duplicated genes with high sequence identities are often misannotated due to incorrect alignment of RNA-seq reads. Therefore, we manually annotated this gene family by searching via BLAST the entire *M. persicae* genome for regions that align to a 148-bp nucleotide sequence that was found to be conserved among the 23 candidate lncRNA genes (*SI Appendix*, Fig. S8). This resulted in the identification of 7 additional genes to make a 30-member gene family (Dataset S4). We updated the transcriptome assembly for this gene locus with the manual annotation (GSE accession no. 129667). We named this gene family the *Ya* family (Yá means aphid in Chinese).

The 3′ ends of the 30 *Ya* transcripts were manually corrected based on the presence of a poly(A) signal and 5′ ends via identification of conserved sequences among the *Ya* transcripts. All *Ya* genes have a three-exon structure and show a modest to high sequence conservation (ranging between 84.6% and 99.1% nucleotide identities compared with *Ya1*), including a region that corresponds to a small open reading frame (ORF) that may translate into a 38-aa peptide in all 30 *Ya* transcripts (Fig. 3*A* and *SI Appendix*, Fig. S9).

Phylogenetic analyses based on nucleotide sequence alignment (*SI Appendix*, Fig. S9) group the *Ya* genes into three distinct clades (Fig. 3*B*). We also confirmed that the *Ya* genes form several tandem repeats in the *M. persicae* genome in which the *Ya* genes were often, but not always, organized in pairs facing outward on opposite genomic strands (Fig. 3*C*). The pairs belong to the same or different phylogenetic clusters of the *Ya* phylogenetic tree. Repeating the WGCNA analysis on RNA-seq of *M. persicae* on nine hosts with the new set of manually annotated 30 *M. persicae Ya* genes grouped all *Ya* members within the same

(darkslateblue) coexpression module as before (*SI Appendix*, Fig. S10), suggesting that the 30 *Ya* genes are coregulated.

We also manually annotated *Ya* genes in five other aphid species from available genomes (*Aphis glycines*, *Acyrthosiphon pisum*, *Diuraphis noxia*, *Myzus cerasi*, and *Rhopalosiphum padi*), resulting in identification of 13 candidate *Ya* genes in *A. glycines*, 16 in *A. pisum*, 17 in *D. noxia*, 29 in *M. cerasi*, and 33 in *R. padi* (Fig. 3*D* and Dataset S5; GSE accession no. 129667). These numbers of *Ya* genes may change when more complete assemblies become available. In all five aphid species, *Ya* genes are tandem repeats in genomes. We did not find *Ya* genes in genomes of hemipteran insect species beyond aphids. Therefore, *Ya* genes are likely unique to aphids and are part of larger families that form tandem repeats in aphid genomes and are DE and coregulated in *M. persicae* on nine divergent host plant species.

Translocated *M. persicae Ya* Transcripts Migrate Systemically in Plants. To evaluate whether aphid *Ya* transcripts migrate systemically within plants, we caged 20 adult aphids on leaves of 4-wk-old *A. thaliana* plants for 24 h and then examined the presence of aphid transcripts in the caged area (feeding site), in the area from the petiole to the aphid cage on the leaf (near the feeding site), and on a distal leaf (Fig. 4*A*). We designed specific primers for the nine *Ya* transcripts found in the feeding sites (*SI Appendix*, Fig. S11 and Table S3). For the six *Ya* transcripts *Ya1*, *Ya2*, *Ya3*, *Ya6*, *Ya11*, and *Ya17*, the correct sizes of amplification products were obtained, and the sequences of these amplification products matched those of the *Ya* transcripts, whereas the presence of *Ya6*, *Ya21*, and *Ya23* transcripts was not confirmed (*SI Appendix*, Fig. S11). *Ya1*, *Ya2*, and *Ya17* transcripts were detected in feeding and distal sites, indicating that these transcripts migrated away from the feeding sites to distal tissues (*SI Appendix*, Fig. S11).

Chen et al.
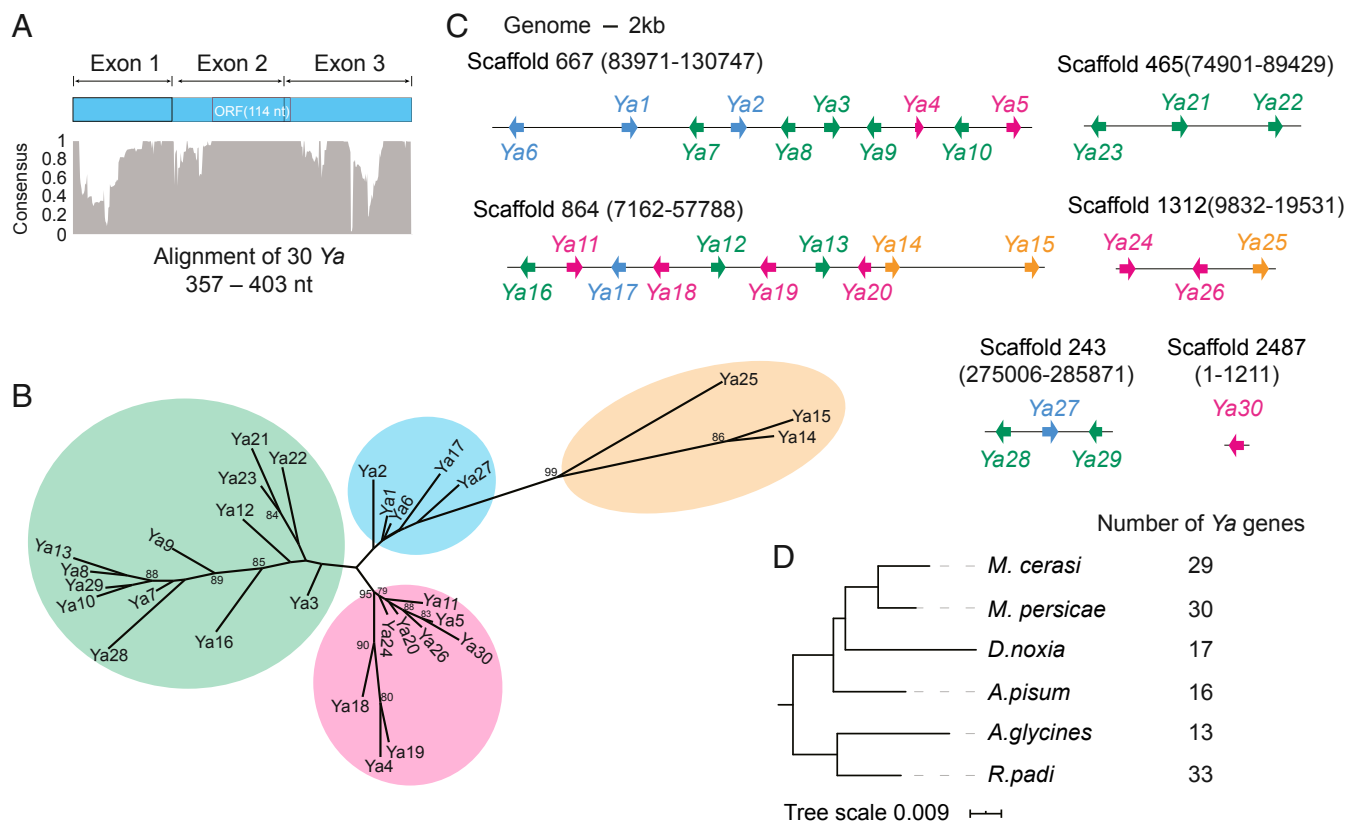
Fig. 3. Annotation of the *Ya* gene family in the *M. persicae* genome. (*A*) Level of conservation (consensus) among aligned nucleotide sequences of mature transcripts of 30 *Ya* genes (alignment shown in *SI Appendix*, Fig. S8). The small ORF may translate into a 38-aa peptide. (*B*) Phylogeny based on the alignment of 30 *Ya* transcripts shown in *A*. (*C*) Locations of *Ya* gene repeats on *M. persicae* genomic scaffolds. *Ya* genes are shown as arrows in colors matching those of the groups of the *Ya* phylogenetic tree shown in *B*. (*D*) Number of *Ya* genes in six aphid genomes. The tree is a maximum likelihood phylogeny based on a concatenated alignment of protein sequences of 5,111 genes.

We focused further analyses on the *Ya1* transcript, because the heatmap of the darkslateblue module suggested that *Ya1* is up-regulated in aphids on Br and At (*SI Appendix*, Fig. S10). The predicted size of the *Ya1* transcript is 382 nt (Fig. 4*B* and *SI Appendix*, Fig. S12*A*). The presence of *Ya1* transcript in aphids and plants was analyzed by RT-PCR with a series of specific primers and by Northern blot hybridizations with a *Ya1* probe. RT-PCR showed that a 357-nt *Ya* transcript that starts at nucleotide 25 of exon 1 was detected in aphids but not in plants, whereas a shorter *Ya1* transcript of 273 nt starting at nucleotide 110 near the start of the sequence corresponding to exon 2 was detected in the plants and migrated systemically (Fig. 4 *B* and *C* and *SI Appendix*, Fig. S12 *A* and *B*). Our 3′ rapid amplification of cDNA ends (RACE) experiments showed that the *Ya1* transcript has a poly(A) tail (*SI Appendix*, Fig. S12*C*). Northern blot analysis confirmed the sizes of the *Ya1* transcripts in aphids and plants; the fragments that hybridized to the *Ya1-SP6* in aphids were larger than the 291-nt *Ya1-SP6* transcript, whereas the *Ya1* fragments in plants were shorter than this transcript (Fig. 4 *D* and *E*). These data indicate that the first ±100 nt at the 5′ end of the 357-nt aphid *Ya1* transcript is processed to produce a 273-nt *Ya1* transcript in plants.

Northern blot analysis and qRT-PCR showed that the *Ya1* transcript gradually decreased in concentration from feeding sites to near feeding sites and distal leaves of *A. thaliana* (Fig. 4 *E* and *F* and *SI Appendix*, Fig. S13). *Ya1* also migrated systemically in *B. rapa*, *P. sativum*, and *Z. mays* exposed to *M. persicae* (Fig. 4*G* and *SI Appendix*, Fig. S14). Sequencing of the PCR products derived from of the feeding sites, near feeding sites, and distal leaves of these three hosts revealed identical sequences to *Ya1*, which differs

from the other *Ya* family members (*SI Appendix*, Fig. S15). Therefore, *M. persicae* deposits *Ya1* transcript into four divergent host plant species during feeding and the transcript migrates systemically within these plants away from aphid feeding sites.

***M. persicae Ya1* Is a Virulence Factor.** The differential expression of *Ya1* in *M. persicae* among hosts and migration of *Ya1* in various of hosts suggest that *Ya1* may have a role in aphid–host interactions. To investigate this, we generated transgenic lines expressing dsRNA corresponding to the *Ya1* sequence for plant-mediated RNA interference (RNAi) of *Ya1* in *M. persicae*. Two independent transgenic *A. thaliana* lines, 1-5 and 2-8, successfully knocked down *Ya1* expression compared with WT Col-0 and lines producing dsRNA corresponding to GFP (dsGFP) in *M. persicae* (Fig. 5*A*). *M. persicae* on *A. thaliana* lines 1-5 and 2-8 had reduced fecundity compared to aphids on WT Col-0 and dsGFP lines (Fig. 5*B*); therefore, knockdown of *Ya1* expression is correlated with reduced *M. persicae* reproduction on *A. thaliana*.

Given that aphids deposit *Ya1* into plants, and that this RNA migrates systemically, we investigated whether stable expression of *Ya1* in planta affects *M. persicae* performance. We generated stable transgenic plants that produced the 273-nt (exons 2 and 3) *Ya1* transcript (35S::Ya1 [Col-0] lines 8-8 and 9-9) and 273-nt *Ya1* mutants in which three ATG start sites within the 38-aa ORF were mutated to stop codons (35S::Ya1_3AUG [Col-0] lines 1-1, 3-3, and 4-4) (Fig. 5 *C* and *D* and *SI Appendix*, Figs. S16–S18). *M. persicae* produced more progeny on both 35S::Ya1 and 35S::Ya1_3AUG compared with 35S:GFP and Col-0 plants (Fig. 5*E* and *SI Appendix*, Fig. S19), indicating that the *Ya1* RNA
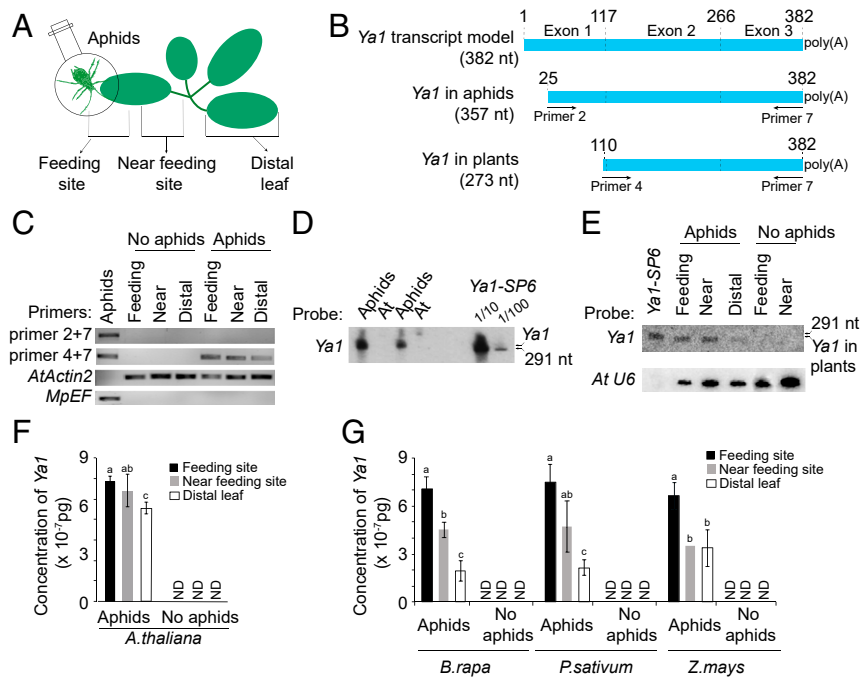
Fig. 4. The *M. persicae* *Ya1* transcript translocates and migrates to distal leaves in plants. (*A*) Schematic overview of the experimental setup. The distal leaf was chosen as described in Mousavi et al. (29). (*B*) Schematic overview of *Ya1* transcripts, showing the locations of primers used for amplification in the RT-PCR experiment of C. (*C*) RT-PCR of *Ya1* transcripts in aphids and aphid-exposed *A. thaliana* Col-0 plants. The 357-nt *Ya1* transcript (primers 2 and 7) was found in aphids but not in plants. A 273-nt transcript (primers 4 and 7) was found at aphid feeding sites and was seen to migrate systemically in plants. *A. thaliana* actin 2 (*AtActin2*) and *M. persicae* elongation factor (*MpEF*) were used to control for the presence of RNA. (*D* and *E*) Northern blot hybridizations with a *Ya1* probe to detect *Ya1* transcript in aphids (*D*) and plants (*E*). Arrows at the right of the blots indicate the locations of the in vitro synthesized 291-nt *Ya1-SP6* RNA transcript and *Ya1* transcript found in aphids and *A. thaliana* Col-0 plants. Equal RNA loading levels were assessed by stripping blots and subsequent labeling with the *A. thaliana* U6 probe. (*F* and *G*) qRT-PCR to detect systemic migration of *M. persicae* *Ya1* transcript in *A. thaliana* (*F*) and *B. rapa*, *P. sativum*, and *Z. mays* (*G*). The y-axes show *Ya1* concentrations based on a standard curve (*SI Appendix*, Fig. S13). ND, not detected. Bars represent mean ± SD concentrations of *Ya1* and two independent biological replicates. Different letters above the bars indicate significant differences between groups ($P < 0.01$, Student t test).

transcript modulates plant processes that lead to increased *M. persicae* fecundity. Taken together, our data show that *M. persicae* translocates the *Ya* transcript into plants during feeding, and that this transcript migrates systemically within plants and promotes fecundity of this aphid. Therefore, *Ya1* has the characteristics of an aphid lncRNA virulence factor.

Discussion

We found that progeny derived from a single asexually reproducing *M. persicae* clone O female stably colonized nine divergent plant species of five families that span the angiosperm phylogeny (27), demonstrating that *M. persicae* is truly polyphagous. We identified *M. persicae* genes that show coordinated up- and down-regulation depending on the plant species that the aphid is colonizing and that are organized as repeats in the aphid genome. The genes are organized in coexpression modules, 13 of which are enriched for DE genes, genes expressed in the salivary glands, and candidate lncRNAs. One of these modules includes all members of the *Ya* family with 30 genes that are tandemly repeated in the *M. persicae* genome. Moreover, transcripts from six *Ya* genes translocate into plants during aphid feeding, and three migrate systemically away from the aphid feeding site to distant leaves. RNAi-mediated knockdown of *Ya* expression in aphids reduces aphid performance on *A. thaliana*. In contrast, aphid performance is increased both on stable transgenic *A. thaliana* lines that heterologously express the aphid *Ya* gene and the *Ya* gene mutant in which stop codons were introduced into the small putative ORF. These data indicate that the *Ya1* RNA transcript acts as a lncRNA virulence factor.

Because genome annotation pipelines are generally focused on protein-coding genes, the majority of *Ya* family members were not annotated in the first version of the *M. persicae* genome (22). Moreover, because the *Ya* family is organized as tandem repeats of highly conserved sequences, genome-guided transcript assembly generated misassembled transcripts. To overcome this, we performed a thorough manual annotation of the *Ya* region and were able to identify the *Ya* family of 30 members, characterize the expression patterns of each of the family members in *M. persicae* in response to the nine plant species, and determine which family members translocate into plants during aphid feeding. We found that *M. persicae* *Ya* members are characterized by a three-exon gene model and produce transcripts ranging from 357 to 403 nt long. When analyzing other hemipterans, we identified tandemly repeated *Ya* family members in the genomes only of aphid species among hemipterans for which genome sequence data are available, suggesting that the *Ya* genes are unique to aphids.

Members of the *Ya* family are both coregulated and organized as repeated clusters in the *M. persicae* genome. The finding that *M. persicae* adjusts to divergent plant species via the coregulation of tandemly repeated gene families was reported earlier for *Cathepsin B* (*CathB*), *Cuticular Proteins* (*CutP*), and other gene families in host swap experiments involving three plant species: *B. rapa*, *A. thaliana*, and *N. benthamiana* (22). In this study, the *CathB* family members *B1* through *B7*, *B10*, and *B11* are parts of two modules that are enriched for DE genes and that show up-regulation of the *CathB* genes in aphids on *B. rapa* and *A. thaliana* versus *N. benthamiana*, confirming previous data (22). The majority of the *CutP* genes grouped together in the skyblue2

Chen et al.
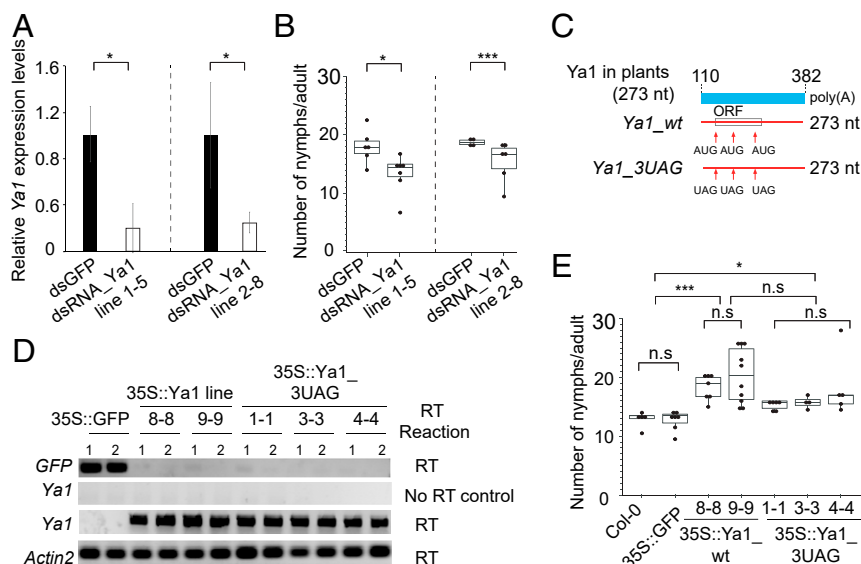
Fig. 5. *M. persicae Ya* lncRNA promotes *M. persicae* colonization on *A. thaliana*. (*A*) qRT-PCR showing knockdown of *Ya1* in *M. persicae* reared on plants that stably produce dsRNA_Ya1 relative to those that produce dsRNA_GFP (controls). The *y*-axis indicates the *Ya1* expression levels relative to *M. persicae* EF reference gene. Bars represent relative mean ± SD expression values at *n* = 4 to 8, in which *n* is one aphid per plant. *P < 0.01, Student *t* test. (*B*) Reduction in fecundity of Ya1-RNAi aphids tested for *Ya1* knockdown in *A*. Each data point (black dot) represents the number of nymphs produced by one adult female per plant. The boxplots show the distribution of data points collected from *n* = 4 to 8 female aphids per *A. thaliana* line as shown on the *x*-axis. (*C*) Mutation of the three AUGs within the ORF of 38 amino acids of the *Ya1* transcript into UAG stop codons. (*D*) Detection of transgene transcripts by RT-PCR at two plants per line. Detection of the *A. thaliana Actin2* transcript served as a control for the presence of RNA. (*E*) Stable expression of *Ya1_wt* and *Ya1_3UAG* promotes *M. persicae* reproduction on plants. The boxplots show the distribution of nymphs produced by an adult female aphid per plant (black dots) collected from *n* = 4 to 10 female aphids per *A. thaliana* line. *P < 0.05; ***P < 0.001, ANOVA followed by the Tukey–Kramer post hoc test.

module. Although this module was not enriched for DE genes, several *CutP* genes were significantly DE and, as before (22), were up-regulated on *N. benthamiana* and down-regulated on *B. rapa* and *A. thaliana*. Therefore, the previous study (22), as well as the present work, demonstrate that the coregulation of gene families that cluster within the *M. persicae* genome play fundamental roles in the ability of this aphid to colonize plants.

Although *Ya* genes are annotated as genes with low coding potential, *Ya* transcripts have a conserved small ORF that potentially translates into a 38-aa peptide. Mutations within the *Ya1* transcript that prevent translation of this peptide in plants do not affect the ability of *Ya1* to promote aphid fecundity, indicating that *Ya1* has virulence activity as an lncRNA. ORFs have been detected in number of transcripts known to function as lncRNAs, including RNAs associated with ribosome function (30, 31) and an X-inactive specific transcript (*Xist*) that regulates the X chromosome inactivation process (32), and it is being debated whether lncRNA ORFs may be translated in some situations (33). Whether the small ORF found to be conserved across *Ya* genes has a function remains to be determined, but based on our annotations and functional analyses of *Ya1*, at least one member of the *Ya* gene family appears to function as an lncRNA.

Several parasites translocate small RNAs into their hosts (34–36). The functions of these transkingdom small RNAs are known for only a few parasites. For example, small RNAs of the fungal plant pathogen *Botrytis cinerea* interact with the AGO1 protein of the *A. thaliana* RNAi machinery to suppress plant defense genes (37), and microRNAs of the parasitic plant dodder (*Cuscuta pentagona*) target host messenger RNAs involved in plant defense (38). In the opposite direction, plants export specific microRNAs to control virulence of a pathogenic fungus (39). In some parasite–host interactions, a large number of long transcripts (>200 nt) were found in hosts, including RNA transcripts of *Cryptosporidium parvum* in the nuclei of human intestinal epithelial cells (40) and transcripts of *C. pentagona* that systemically

migrate in *Solanum lycopersicum* and *A. thaliana* (41). However, whether these larger parasite RNAs modulate parasite–host interactions is unclear. Here we show that *M. persicae* RNA transcripts translocate into plants; these include transcripts of the *Ya* family that migrate systemically. Knockdown of *Ya* gene expression via RNAi reduces aphid fecundity, whereas in planta expression of *Ya1* as an lncRNA promotes aphid fecundity. This suggests that the aphid *Ya* genes are virulence factors, and that the *Ya1* transcript controls aphid performance via the plant.

The *M. persicae Ya1* lncRNA may be an effector that modulates specific plant processes. Interestingly, the entire *M. persicae Ya* family of 30 members is part of the darkslateblue module, which is the only module among the 77 coexpression modules that consists of two groups of genes with exact opposite expression patterns. The module also includes several protein-coding genes with similarities to hemolymph juvenile hormone binding and WD40 and EGF-like domain-containing proteins that have roles in signal transduction in insects. Given this finding, a non-mutually exclusive possibility is that *Ya1* and other *Ya* transcripts may have a sensing role within aphids; for example, lncRNAs not degraded in plants may migrate back into the aphid to regulate aphid gene expression, in agreement with lncRNAs often having functions in gene expression regulation (42–44). Whereas the mechanism by which *Ya1* controls aphid fecundity remains to be determined, our present results indicate that members of the *Ya* gene family are aphid-specific virulence factors.

Materials and Methods

Aphids from a *M. persicae* clone O colony on *B. rapa* were transferred to the nine plant species as shown in Fig. 1*A*. They were reared on these plants for at least 10 generations, and stable colonies at five biological replicates per plant species were processed for RNA extraction, RNA-seq, and downstream transcriptome analyses as described previously (22).

A genome-guided transcriptome assembly was generated with RNA-seq data of the 45 libraries of the nine host experiments and RNA-seq data generated from library LIB1777 (22). Transcripts with an FPKM ≤0.2 were

Chen et al.

removed from downstream analyses. The computational workflow for *M. persicae* transcript annotation and lncRNA identification is shown in *SI Appendix*, Fig. S2. DE transcripts of *M. persicae* colonies on the nine plant species were identified as described previously (22), with lowly expressed transcripts (mean count <10) removed on the basis of normalized counts and transcripts considered DE if they had a *P* value <0.05 after accounting for a 5% FDR according to the Benjamini–Hochberg procedure and if log2 fold change was >1 (Dataset S1). Genes with >5 TPM in at least one sample per plant host were included in the coexpression analysis using WGCNA and hierarchical clustering on the basis of dissimilarity of gene connectivity (1-TOM) (26).

M. persicae Ya genes were manually annotated by selecting gene models and corresponding transcripts that align to a conserved 148-bp nucleotide sequence among *Ya* transcripts (*SI Appendix*, Fig. S8). Selected transcripts were further curated by manually annotating the 3′ ends of each of the transcripts based on the presence of a poly-A tail. The 5′ ends were identified based on the most conserved sequence among all transcripts combined with existing RT-PCR data for *Ya1* (*SI Appendix*, Fig. S12). *Ya* genes were manually annotated in five other aphid species using publicly available RNA-seq data and genome assemblies (Dataset S5).

Translocation and systemic migration of aphid transcripts in plants were determined by caging a leaf section with aphids and detecting aphid transcripts in the caged area (feeding site), next to the caged area of the same leaf (near-feeding site), and on a distal leaf (distal site). Plants exposed to cages without aphids served as controls. For RNA-seq analyses, leaf areas covered by the cages were carefully washed three times with deionized water and three time with nuclease-free water. RNAs were isolated from four independent biological replicates of aphid-exposed leaves and nonexposed control leaves and processed for RNA-seq. Reads were trimmed to remove sequencing adapters and aligned to *A. thaliana* genome (TAIR10 database; https://www.arabidopsis.org/) and the *M. persicae* G006 genome (22) with HISAT2 v2.0.5. Reads mapped to the *M. persicae* genome were retrieved and subjected to further filtering by mapping them back to the *A. thaliana* genome. Reads that did not align to the *A. thaliana* genome in the last step

were considered unique *M. persicae* mapping reads. Transcripts with ≥50 TPM in at least one sample and that were present in at least three samples were selected for further analysis.

DNA amplification, cDNA synthesis, RT-PCR, Sanger sequencing, quantification of transcript levels by qPCR, and Northern blot hybridization were conducted using standard procedures and specific primers (*SI Appendix*, Table S3).

Stable transgenic *A. thaliana* lines were generated by cloning specific genes into pJawohl8-RNAi and pBI121 to generate pJawohl8-RNAi_Ya1, pBI121_35S::Ya1, and pBI121_35S::Ya1_3UAG plasmids, which were introduced into *A. tumefaciens* strain GV3101 that carried the helper plasmid pMP90RK for subsequent transformation of *A. thaliana* Col-0 using the floral dip method (45). Transgenic seeds were selected on phosphinothricin (BASTA) or kanamycin. F2 seedlings with 3:1 alive/dead segregation were taken forward to the F3 stage, and F3 lines with 100% survival ratio (homozygous) were selected for aphid experiments. Assessments of gene knockdown in aphids and aphid fecundity assays were conducted as described previously (22).

All data analyses were performed in R Stats Package. All statistical tests are described in the figure legends. More details on the materials and methods used in this study are provided in *SI Appendix*.

1. V. G. Kaplin, Y. A. Sharapova, Influence of the Russian wheat aphid Diuraphis noxia (Kurdjumov) (Homoptera, Aphididae) on productive qualities of spring bread wheat and barley grown from the seeds from aphid-infested spikes. *Entomol. Rev. (Engl. Transl.)* **97**, 415–424 (2017).

2. E. A. Backus, M. S. Serrano, C. M. Ranger, Mechanisms of hopperburn: An overview of insect taxonomy, behavior, and physiology. *Annu. Rev. Entomol.* **50**, 125–151 (2005).

3. M. Heck, Insect transmission of plant pathogens: A systems biology perspective. *mSystems* **3**, e00168-e17 (2018).

4. N. Van Bogaert, K. De Jonghe, E. J. Van Damme, M. Maes, G. Smagghe, Quantitation and localization of pospiviroids in aphids. *J. Virol. Methods* **211**, 51–54 (2015).

5. E. Jaskowska, C. Butler, G. Preston, S. Kelly, Phytomonas: Trypanosomatids adapted to plant environments. *PLoS Pathog.* **11**, e1004484 (2015).

6. Y. Walia, S. Dhir, A. A. Zaidi, V. Hallan, Apple scar skin viroid naked RNA is actively transmitted by the whitefly Trialeurodes vaporariorum. *RNA Biol.* **12**, 1131–1138 (2015).

7. S. A. Hogenhout, D. Ammar, A. E. Whitfield, M. G. Redinbaugh, Insect vector interactions with persistently transmitted viruses. *Annu. Rev. Phytopathol.* **46**, 327–359 (2008).

8. J. C. Ng, B. W. Falk, Virus-vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. *Annu. Rev. Phytopathol.* **44**, 183–212 (2006).

9. L. M. Schoonhoven, J. J. A. van Loon, M. Dicke, *Insect-Plant Biology*, (Oxford University Press, ed. 2, 2005).

10. Center for Agriculture and Bioscience International, *Myzus persicae* (green peach aphid). Invasive Species Compendium. https://www.cabi.org/isc/datasheet/35642. Accessed April 9, 2020.

11. J. Syller, W. Marczewski, J. Pawłowicz, Transmission by aphids of potato spindle tuber viroid encapsidated by potato leafroll luteovirus particles. *Eur. J. Plant Pathol.* **103**, 285–289 (1997).

12. W. F. Tjallingii, Salivary secretions by aphids interacting with proteins of phloem wound responses. *J. Exp. Bot.* **57**, 739–745 (2006).

13. W. F. Tjallingii, H. Esch, Fine-structure of aphid stylet routes in plant-tissues in correlation with EPG signals. *Physiol. Entomol.* **18**, 317–328 (1993).

14. T. Will, A. C. Furch, M. R. Zimmermann, How phloem-feeding insects face the challenge of phloem-located defenses. *Front Plant Sci* **4**, 336 (2013).

15. N. S. Mutti *et al.*, A protein from the salivary glands of the pea aphid, Acyrthosiphon pisum, is essential in feeding on a host plant. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9965–9969 (2008).

16. J. I. Bos *et al.*, A functional genomics approach identifies candidate effectors from the aphid species Myzus persicae (green peach aphid). *PLoS Genet.* **6**, e1001216 (2010).

17. M. Pitino, S. A. Hogenhout, Aphid protein effectors promote aphid colonization in a plant species-specific manner. *Mol. Plant Microbe Interact.* **26**, 130–139 (2013).

18. G. J. Kettles, I. Kaloshian, The potato aphid salivary effector Me47 is a glutathione-S-transferase involved in modifying plant responses to aphid Infestation. *Front Plant Sci* **7**, 1142 (2016).

19. S. T. Mugford, E. Barclay, C. Drurey, K. C. Findlay, S. A. Hogenhout, An immunosuppressive aphid saliva protein is delivered into the cytosol of plant mesophyll cells during feeding. *Mol. Plant Microbe Interact.* **29**, 854–861 (2016).

20. P. A. Rodriguez, C. Escudero-Martinez, J. I. Bos, An aphid effector targets trafficking protein VPS52 in a host-specific manner to promote virulence. *Plant Physiol.* **173**, 1892–1903 (2017).

21. R. Chaudhary *et al.*, Aphid effector Me10 interacts with tomato TFT7, a 14-3-3 isoform involved in aphid resistance. *New Phytol.* **221**, 1518–1528 (2019).

22. T. C. Mathers *et al.*, Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol.* **18**, 27 (2017).

23. P. V. Pinheiro *et al.*, Host plants indirectly influence plant virus transmission by altering gut cysteine protease activity of aphid vectors. *Mol. Cell. Proteomics* **16** (suppl. 1), S230–S243 (2017).

24. K. E. Mauck, Q. Chesnais, L. R. Shapiro, Evolutionary determinants of host and vector manipulation by plant viruses. *Adv. Virus Res.* **101**, 189–250 (2018).

25. J. P. Carr, A. M. Murphy, T. Tungadi, J. Y. Yoon, Plant defense signals: Players and pawns in plant-virus-vector interactions. *Plant Sci.* **279**, 87–95 (2019).

26. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

27. S. D. Gamboa-Tuz *et al.*, New insights into the phylogeny of the TMBIM superfamily across the tree of life: Comparative genomics and synteny networks reveal independent evolution of the BI and LFG families in plants. *Mol. Phylogenet. Evol.* **126**, 266–278 (2018).

28. J. S. Ramsey *et al.*, Genomic resources for Myzus persicae: EST sequencing, SNP identification, and microarray design. *BMC Genomics* **8**, 423 (2007).

29. S. A. Mousavi, A. Chauvin, F. Pascaud, S. Kellenberger, E. E. Farmer, GLUTAMATE RECEPTOR-LIKE genes mediate leaf-to-leaf wound signalling. *Nature* **500**, 422–426 (2013).

30. C. Zeng, M. Hamada, Identifying sequence features that drive ribosomal association for lncRNA. *BMC Genomics* **19** (suppl. 10), 906 (2018).

31. C. Zeng, T. Fukunaga, M. Hamada, Identification and analysis of ribosome-associated lncRNAs using ribosome profiling data. *BMC Genomics* **19**, 414 (2018).

32. A. Sahakyan, Y. Yang, K. Plath, The role of xist in X-chromosome dosage compensation. *Trends Cell Biol.* **28**, 999–1013 (2018).

33. Z. Ji, R. Song, A. Regev, K. Struhl, Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **4**, e08890 (2015).

34. J. H. Zhao, H. S. Guo, Trans-kingdom RNA interactions drive the evolutionary arms race between hosts and pathogens. *Curr. Opin. Genet. Dev.* **58–59**, 62–69 (2019).

35. P. J. van Kleeff, M. Galland, R. C. Schuurink, P. M. Bleeker, Small RNAs from Bemisia tabaci are transferred to Solanum lycopersicum phloem during feeding. *Front Plant Sci* **7**, 1759 (2016).

36. T. Meningher et al., Schistosomal microRNAs isolated from extracellular vesicles in sera of infected patients: A new tool for diagnosis and follow-up of human schisto-somiasis. J. Infect. Dis. 215, 378–386 (2017).

37. A. Weiberg et al., Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. Science 342, 118–123 (2013).

38. S. Shahid et al., MicroRNAs from the parasitic plant Cuscuta campestris target host messenger RNAs. Nature 553, 82–85 (2018).

39. T. Zhang et al., Cotton plants export microRNAs to inhibit virulence gene expression in a fungal pathogen. Nat. Plants 2, 16153 (2016).

40. Y. Wang et al., Delivery of parasite RNA transcripts into infected epithelial cells during Cryptosporidium infection and its potential impact on host gene transcription. J. Infect. Dis. 215, 636–643 (2017).

41. G. Kim, M. L. LeBlanc, E. K. Wafula, C. W. dePamphilis, J. H. Westwood, Genomic-scale exchange of mRNA between a parasitic plant and its hosts. Science 345, 808–811 (2014).

42. A. Di Ruscio et al., DNMT1-interacting RNAs block gene-specific DNA methylation. Nature 503, 371–376 (2013).

43. B. K. Sun, A. M. Deaton, J. T. Lee, A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol. Cell 21, 617–628 (2006).

44. K. M. Schmitz, C. Mayer, A. Postepska, I. Grummt, Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. 24, 2264–2269 (2010).

45. N. Bechtold, J. Ellis, G. Pelletier, In planta Agrobacterium-mediated gene transfer by infiltration of adult Arabidopsis thaliana plants. C. R. Acad. Sci. Ser. III Sci. Vie. Life Sci 316, 1194–1199 (1993).

AGRICULTURAL SCIENCES