MicroRNAs from the parasitic plant *Cuscuta campestris* target host messenger RNAs

Saima Shahid[1,2], Gunjune Kim[3], Nathan R. Johnson[1,2], Eric Wafula[2], Feng Wang[1,2,4], Ceyda Coruh[1,2,5], Vivian Bernal-Galeano[3], Claude W. dePamphilis[1,2], James H. Westwood[3], and Michael J. Axtell[1,2]

[1] Intercollege Ph.D. Program in Plant Biology, Huck Institutes of the Life Sciences, Penn State University, University Park, PA 16802 USA

[2] Department of Biology, Penn State University, University Park, PA 16802 USA

[3] Department of Plant Pathology, Physiology and Weed Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA.

[4] Current Address: Department of Biology, Indiana University, Bloomington IN 47405, USA

[5] Current Address: Salk Institute for Biological Studies, La Jolla, CA 92037, USA

First paragraph:

Dodders (*Cuscuta* spp.) are obligate parasitic plants that obtain water and nutrients from the stems of host plants via specialized feeding structures called haustoria. Dodder haustoria facilitate bi-directional movement of viruses, proteins, and mRNAs between host and parasite[1], but the functional effects of these movements are not clear. Here we show that *C. campestris* haustoria accumulate high levels of many novel microRNAs (miRNAs) while parasitizing *Arabidopsis thaliana* hosts. Many of

these miRNAs are 22 nts long, a usually rare size of plant miRNA associated with amplification of target silencing through secondary small interfering RNA (siRNA) production[2]. Several *A. thaliana* mRNAs are targeted by *C. campestris* 22 nt miRNAs during parasitism, resulting in mRNA cleavage, high levels of secondary siRNA production, and decreased mRNA accumulation levels. Hosts with a mutation in the target *SIEVE ELEMENT OCCLUSION RELATED 1* (*SEOR1*) supported significantly higher growth of *C. campestris.* Homologs of target mRNAs from diverse plants also have predicted target sites to induced *C. campestris* miRNAs, and several of the same miRNAs are expressed when *C. campestris* parasitizes a second host, *Nicotiana benthamiana*. These data show that *C. campestris* miRNAs act as *trans*-species regulators of host gene expression, and suggest that they may act as virulence factors during parasitism.

Host-induced gene silencing (HIGS) involves plant transgenes that produce siRNAs which can silence targeted pathogen/parasite mRNAs in *trans*[3,4]. Plant-based HIGS is effective against fungi[5], nematodes[6], insects[7], and the parasitic plant *C. pentagona*[8]. The apparent ease with which plant-based HIGS functions suggests that plants might also exchange naturally occurring small RNAs during pathogen and parasite interactions. Consistent with this hypothesis, small RNAs from the plant pathogenic fungus *Botrytis cinerea* target host mRNAs during infection[9], and HIGS targeting *Botrytis Dicer-Like* mRNAs reduces pathogen virulence[10]. Conversely, the conserved miRNAs miR159 and miR166 can be exported from cotton into the fungal pathogen *Verticillium dahliae* where they target fungal mRNAs encoding virulence factors[11].

We hypothesized that naturally occurring small RNAs might be exchanged across the *C. campestris* haustorium and affect gene expression in the recipient species. To test this hypothesis, we profiled small RNA expression from *C. campestris* grown on *A. thaliana* hosts using high-throughput small RNA sequencing (sRNA-seq). Two biological replicates each from three tissues were analyzed: Parasite stem (PS), comprising a section of *C. campestris* stem above the site of haustorium formation; Interface (I), comprising *C. campestris* stem with haustoria with associated *A. thaliana* stem tissue; and Host stem (HS), comprising sections of *A. thaliana* stems above the interface region, as previously described[12]. Sequenced small RNAs were designated as host-derived or parasite-derived based on alignment to the *A. thaliana* reference genome assembly and expression analysis of HS vs. PS samples. Host-derived reads were aligned to the *A. thaliana* reference genome assembly, while parasite-derived sRNA-seq reads were initially clustered by sequence similarity (Supplementary Data 1).

151 *C. campestris* small RNAs were differentially expressed between I and PS samples (Supplementary Data 2), most of which (136) were up-regulated in I relative to PS (Figure 1A). RNA blots confirmed the strong accumulation in I samples for several tested small RNAs (Figure 1B). Whole-genome shotgun DNA sequence data from *C. campestris* were used to produce short, local genome assemblies containing the 136 interface-induced small RNAs. This allowed alignment of all parasite-derived small RNAs to these genomic loci. We were then able to discern microRNA loci from other small RNA loci by virtue of stereotypical predicted hairpins and accumulation of

miRNA/miRNA* duplexes (Supplementary Data 3-5). Most of the interface up-regulated *C. campestris* small RNAs were found to be 22 or 21 nt miRNAs (Figures 1C-D; Supplementary Data 3-5). The prevalence of 22 nt miRNAs in *C. campestris* is unprecedented: in all plant species previously annotated 21 nt miRNAs are much more prevalent than 22 nt miRNAs (miRBase version 21[13]). 22 nt plant miRNAs are strongly associated with secondary siRNA accumulation from their targets[14,15]. Secondary siRNAs are thought to amplify the strength of miRNA-directed gene silencing[2].

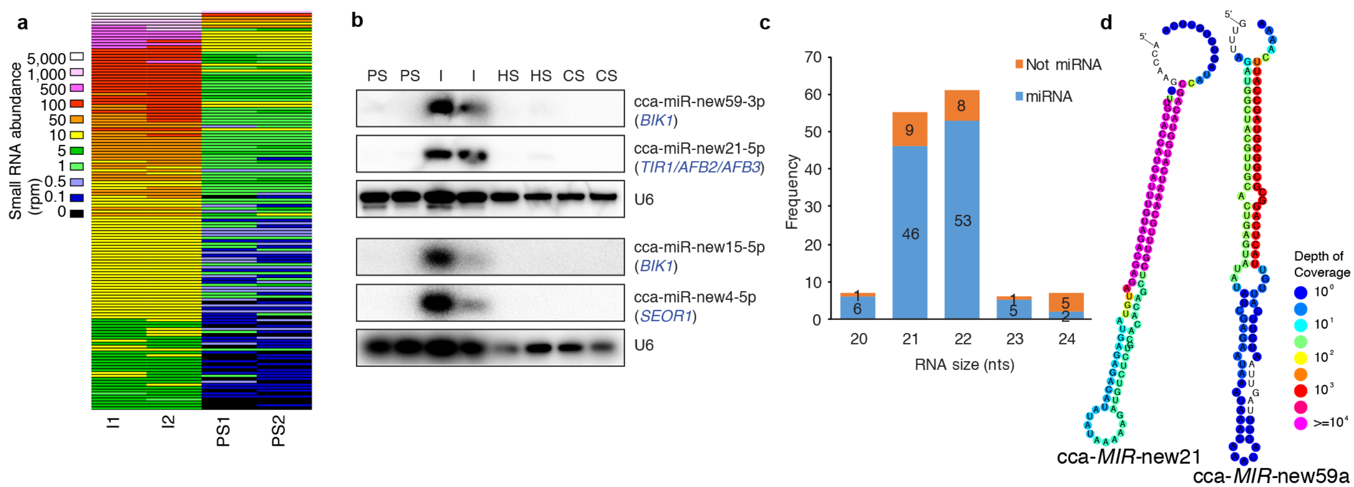


Figure 1. *C. campestris* miRNAs induced at the haustorial interface. **a)** Heatmap showing accumulation of 136 *C. campestris* small RNAs that are significantly (FDR = 0.05) more than two-fold higher in interface relative to parasite stem samples. I: Interface, PS: Parasite stem, rpm: reads per million. **b)** Small RNA blots. *A. thaliana* gene names in parentheses are targets of the probed small RNA. HS: *A.* thaliana host stems above the point of attachment, CS: control (un-parasitized) *A. thaliana* stems. **c)** Tally of sizes and types for the 136 induced *C. campestris* small RNAs. **d)** Examples of *C. campestris MIRNA* hairpins producing 22 nt miRNAs. Predicted RNA secondary

structures are overlaid with colors representing depth of small RNA sequencing. See Supplementary Data 3-5 for details.

Differential expression analysis of *A. thaliana* small RNAs found six loci induced in I relative to HS samples (Figure 2A-B, Extended Data Figure 1, Supplementary Data 6), each deriving from *A. thaliana* mRNAs with known functions: *TIR1*, *AFB2*, and *AFB3*, which encode related and partially redundant auxin receptors[16], *BIK1*, which encodes a plasma membrane-localised kinase required for both pathogen-induced and developmental signaling[17,18], *SEOR1*, which encodes a major phloem protein that accumulates in filamentous networks in sieve tube elements and reduces photosynthate loss from the phloem upon injury[19,20], and *SCHIZORHIZA/HSFB4*, which encodes a predicted transcriptional repressor that is required to maintain stem-cell identity in roots[21-23]. The induced siRNAs from these mRNAs resembled secondary siRNAs in their size distributions, double-stranded accumulation, and phasing (Figure 2A-B; Extended Data Figure 1).

Each of these six mRNAs has either one or two complementary sites to interface-induced *C. campestris* 22 nt miRNAs (Figure 2A-B; Extended Data Figure 1). The *TIR1, AFB2,* and *AFB3* mRNAs are also targeted by the conserved miRNA, miR393, at distinct sites (Extended Data Figure 1). Analysis of uncapped mRNA fragments using 5'-RNA ligase-mediated rapid amplification of cDNA ends (5'-RLM-RACE) found strong evidence for miRNA-directed cleavage at all of the complementary sites to *C. campestris* miRNAs, specifically from interface samples but not from control stem samples (Figure 2; Extended Data Figure 1). Several of the key *MIRNA* loci were detected by PCR of *C. campestris* genomic DNA prepared from four-day old seedlings that had never interacted with a host plant (Extended Data Figure 2). This rules out the possibility that these miRNAs arose from a cryptic region of the *A. thaliana* genome. We conclude that 22 nt miRNAs from *C. campestris* target these *A. thaliana* mRNAs during parasitism, triggering production of secondary siRNAs.

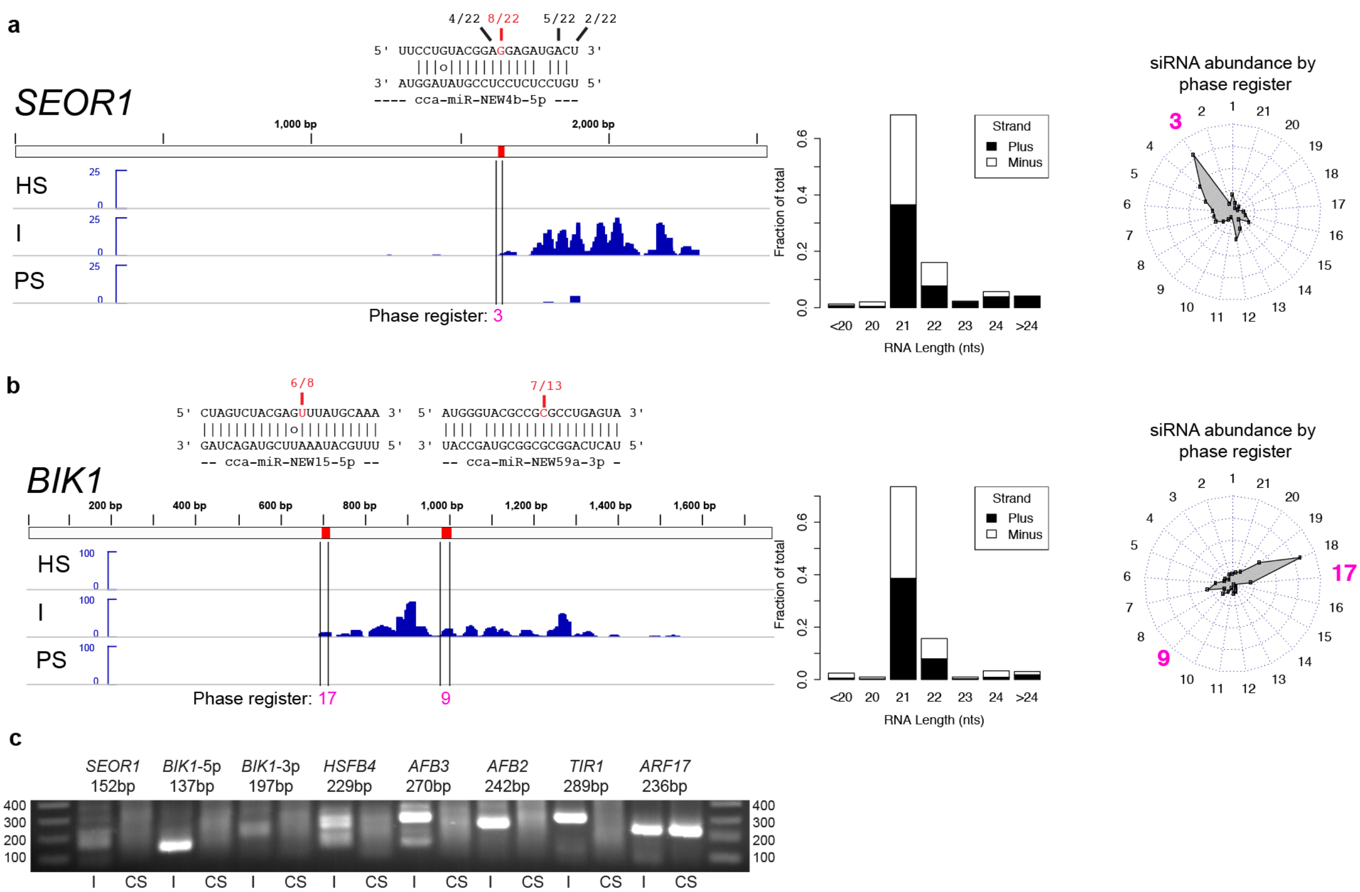


Figure 2. *C. campestris* miRNAs cause slicing and phased siRNA production from host mRNAs. **a)** Small RNA-seq coverage (reads per million) across the *A. thaliana SEOR1* transcript is shown in blue for host stem (HS), interface (I), and parasite stem (PS)

samples. miRNA complementary sites are shown with alignments above. Fractions indicate numbers of 5'-RLM-RACE clones with 5'-ends at the indicated positions; red color highlights the predicted slicing site. Barchart shows the length and polarity distribution of *SEOR1*-mapped siRNAs. Radar chart shows the fraction of siRNAs in each of the 21 possible phasing registers; the register highlighted in magenta is the one predicted by the miRNA target site. **b)** As in a, except for *A. thaliana BIK1*. Extended Data Figure 1 shows similar plots for four other mRNAs. **c)** 5'-RLM-RACE products from nested amplifications for the indicated cDNAs. Numbers below cDNA names give the expected size of the products in base-pairs (bp). First and last lanes are size standards. I: interface, CS: control stem. *ARF17*, a known target of miR160, is a positive control.

Accumulation of five of the six secondary siRNA-producing targets was significantly reduced in stems parasitized by *C. campestris* compared to un-parasitized stems ($p < 0.05$; Figure 3A), consistent with miRNA-mediated repression. The true magnitude of repression for these targets could be even greater, since many miRNAs also direct translational repression. Accumulation of many known *A. thaliana* secondary siRNAs is often partially dependent on the endonuclease Dicer-Like 4 (DCL4) and wholly dependent on RNA-Dependent RNA polymerase 6 (RDR6/SGS2/SDE1)[2]. Accumulation of an abundant secondary siRNA from *TIR1* was eliminated entirely in the *sgs2-1* mutant, and reduced in the *dcl4-2t* mutant (Figure 3B). Thus host *DCL4* and *RDR6/SGS2/SDE1* are required for secondary siRNA production, strongly implying that the *C. campestris* derived miRNAs are active inside of host cells.

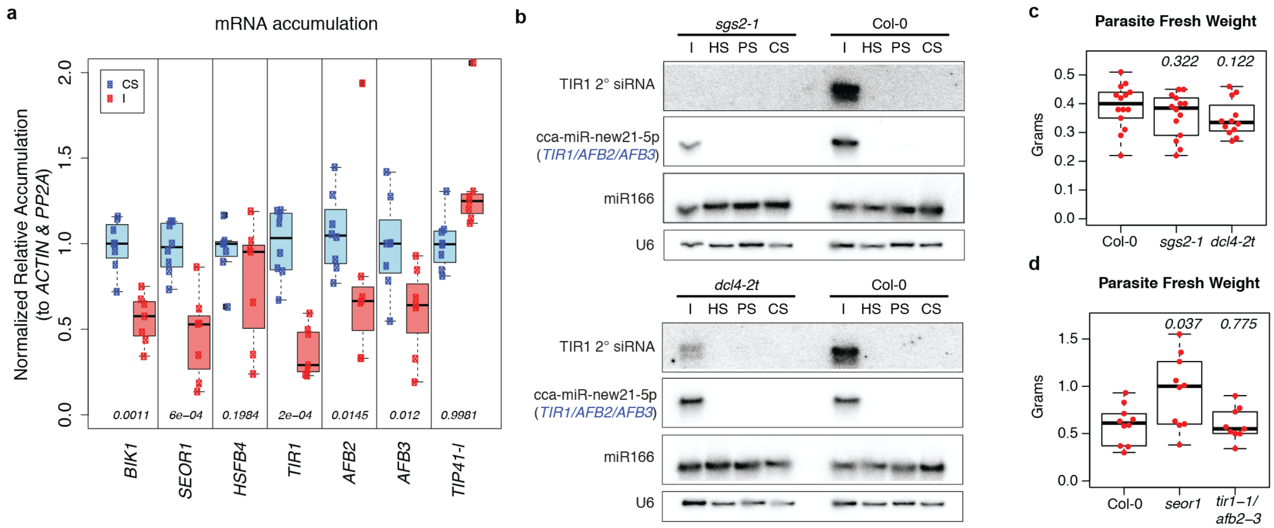


Figure 3. Effects of *C. campestris* miRNAs and their targets. **a)** *A. thaliana* mRNA accumulation levels in **I** (interface) vs. CS (control stems) during *C. campestris* parasitism, assessed by qRT-PCR. Data from 7 or 8 biological replicates are plotted (dots), and boxplots indicate the median (horizontal lines), 1st-3rd quartile range (boxes), and up to 1.5 x the interquartile range (whiskers). Numbers indicate p-values comparing CS and **I** samples (Wilcoxon rank-sum test, unpaired, one-tailed). *TIP41-l* is a control mRNA. **b)** RNA blots from *C. campestris* infestations of the indicated *A. thaliana* genotypes. **I**: interface, HS: host stem, PS: parasite stem, CS: control stem. **c)** Accumulation of *C. campestris* biomass on *A. thaliana* hosts of the indicated genotypes 18 days post-attachment. P-values (Wilcoxon rank-sum tests, unpaired, two-tailed) from comparison of mutant to wild-type (Col-0) are shown. Boxplot conventions as in panel a. n=14, 14, and 12 for Col-0, *sgs2-1*, and *dcl4-2t*, respectively. **d)** As in c, except for *seor1* and *tir1-1/afb2-3*. n=10, 10, and 9 for Col-0, *seor1*, and *tir1-1/afb2-3*, respectively.

In repeated trials with varying methodologies we did not observe consistent significant differences in parasite fresh weight using *dcl4-2t* and *sgs2-1* mutants as hosts (Figure 3C, Extended Data Figure 3). This implies that loss of induced secondary siRNAs is not sufficient to detectably affect parasite biomass accumulation. We also observed no significant difference in growth on a *tir1-1/afb2-3* double mutant (Figure 3D). However, there was a significant increase ($p = 0.037$) of parasite fresh weight after growth on the *seor1* mutant (Figure 3D). This demonstrates that *SEOR1*, which is targeted by a *C. campestris* miRNA during parasitism, restricts growth of *C. campestris*.

C. campestris has a broad host range, primarily among eudicots[24]. Therefore, we searched for miRNA complementary sites for the six relevant *C. campestris* miRNAs in eudicot orthologs of *TIR1, AFB2, AFB3*, *SEOR1, BIK1,* and *HSFB4*. We used *GAPDH* orthologs with the same six miRNAs as a negative control query, and *PHV* orthologs with a miR166 query[25] as a positive control. Probable orthologs of *BIK1, SEOR1,* and *TIR/AFB* were predicted targets of interface-induced miRNAs in many eudicot species, while predicted targets in probable *HSFB4* orthologs were less frequent (Figure 4A, Supplementary Data 7). We conclude that the induced *C. campestris* miRNAs collectively would be able to target *TIR1, AFB2, AFB3, SEOR1,* and *BIK1* orthologs in many eudicot species, and *HSFB4* in some species.

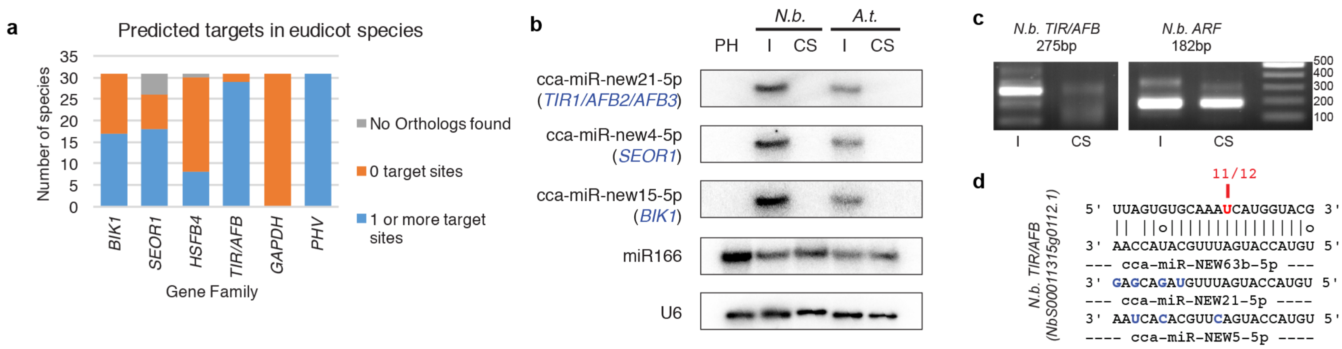


Figure 4. Conservation of host mRNA targeting by *C. campestris*. **a)** Predicted targets in the indicated eudicot gene families. See Supplementary Data 7 for details. **b)** RNA blots of the indicated small RNAs from interface (I) and control stem (CS) samples of *C. campestris* infested *N. benthamiana* (*N.b.*) and *A. thaliana* (*A.t.*) hosts, as well as from *C. campestris* pre-haustoria (PH) induced by seedlings twining on a dead bamboo stake (Extended Data Figure 5). **c)** 5'-RLM-RACE products from nested amplifications for the indicated *N. benthamiana* cDNAs. Numbers below labels are predicted sizes of the products in base pairs (bp); last lane is a DNA size standard. I: interface, CS: control stem. *N.b. ARF*, is an *A.t. ARF17* ortholog with a target site for miR160, and serves as a positive control. The image was cropped to remove irrelevant lanes; the uncropped image is in Extended Data Figure 4. **d)** Alignment of *C. campestris* interface-induced miRNAs with a *N. benthamiana TIR/AFB* ortholog. Fraction indicates termini of 5'-RLM-RACE products; the predicted position of miRNA-directed slicing is shown in red.

Interface samples from *C. campestris*-infested *Nicotiana benthamiana* petioles have strong accumulation of several of the same *C. campestris* miRNAs we identified from our study of parasitized *A. thaliana* (Figure 4B). Analysis of uncapped RNA ends provided strong evidence for cleavage of an *N. benthamiana TIR1/AFB* ortholog by a *C. campestris* miRNA (Figures 4C-D; Extended Data Figure 4). This directly demonstrates that some of the same *C. campestris* miRNAs are expressed upon interaction with different species of host plants, and that at least one orthologous host mRNA is targeted in multiple species. None of the interface-induced miRNAs we tested were detectable from *C. campestris* pre-haustoria sampled from seedling tips that had coiled around dead bamboo stakes instead of a live host (Figure 4B; Extended Data Figure 5). This suggests that expression of these miRNAs requires prior contact with a living host.

These data demonstrate that *C. campestris* induces a large number of miRNAs in the functional haustorium, and that some of them direct cleavage of host mRNAs and reduce their accumulation. Many of the induced miRNAs are 22 nts long, and associated with secondary siRNA production from their host targets. Several of the targets are functionally linked to plant pathogenesis: Manipulation of *TIR1/AFB2/AFB3* accumulation levels affects bacterial pathogenesis and defense signaling[26]. The BIK1 kinase is a central regulator of pathogen-induced signaling[27], and kinase-dependent pathogen signaling is known to play a key role in tomato resistance against *C. reflexa*[28]. Perhaps the most intriguing target is *SEOR1*, which encodes a very abundant protein present in large, filamentous agglomerations in phloem sieve-tube elements[19]. *seor1* mutants have an increased loss of sugars from detached leaves[20], and our data show that *seor1* mutants also support increased *C. campestris* biomass accumulation. A key function of the *C. campestris* haustorium is to take nutrients from the host phloem; targeting *SEOR1* could be a strategy to increase sugar uptake from host phloem for the benefit of the parasite. Overall, these data suggest that *C. campestris trans*-species miRNAs might function as virulence factors to remodel host gene expression to its advantage during parasitism.

Methods [separate online only document].

References

1. Kim, G. & Westwood, J. H. Macromolecule exchange in Cuscuta-host plant interactions. *Curr Opin Plant Biol* **26,** 20–25 (2015).
2. Fei, Q., Xia, R. & Meyers, B. C. Phased, Secondary, Small Interfering RNAs in Posttranscriptional Regulatory Networks. *Plant Cell* **25,** 2400–2415 (2013).
3. Baulcombe, D. C. VIGS, HIGS and FIGS: small RNA silencing in the interactions of viruses or filamentous organisms with their plant hosts. *Curr Opin Plant Biol* **26,** 141–146 (2015).
4. Weiberg, A., Bellinger, M. & Jin, H. Conversations between kingdoms: small RNAs. *Curr. Opin. Biotechnol.* **32,** 207–215 (2015).
5. Nowara, D. *et al.* HIGS: host-induced gene silencing in the obligate biotrophic fungal pathogen Blumeria graminis. *Plant Cell* **22,** 3130–3141 (2010).
6. Huang, G., Allen, R., Davis, E. L., Baum, T. J. & Hussey, R. S. Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. *Proc Natl Acad Sci USA* **103,** 14302–14306 (2006).
7. Baum, J. A. *et al.* Control of coleopteran insect pests through RNA interference. *Nat Biotechnol* **25,** 1322–1326 (2007).

8. Alakonya, A. *et al.* Interspecific RNA interference of SHOOT MERISTEMLESS-like disrupts Cuscuta pentagona plant parasitism. *Plant Cell* **24,** 3153–3166 (2012).

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

10. Wang, M. *et al.* Bidirectional cross-kingdom RNAi and fungal uptake of external RNAs confer plant protection. *Nature Plants* **2,** 16151 (2016).

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

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

13. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42,** D68–73 (2014).

14. Chen, H.-M. *et al.* 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc Natl Acad Sci USA* **107,** 15269–15274 (2010).

15. Cuperus, J. T. *et al.* Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat Struct Mol Biol* (2010). doi:10.1038/nsmb.1866

16. Dharmasiri, N. *et al.* Plant development is regulated by a family of auxin receptor F box proteins. *Dev Cell* **9,** 109–119 (2005).

17. Veronese, P. *et al.* The membrane-anchored BOTRYTIS-INDUCED KINASE1 plays distinct roles in Arabidopsis resistance to necrotrophic and biotrophic

pathogens. *Plant Cell* **18,** 257–273 (2006).

18. Lin, W. *et al.* Inverse modulation of plant immune and brassinosteroid signaling pathways by the receptor-like cytoplasmic kinase BIK1. *Proc Natl Acad Sci USA* **110,** 12114–12119 (2013).
19. Froelich, D. R. *et al.* Phloem ultrastructure and pressure flow: Sieve-Element-Occlusion-Related agglomerations do not affect translocation. *Plant Cell* **23,** 4428–4445 (2011).
20. Jekat, S. B. *et al.* P-proteins in Arabidopsis are heteromeric structures involved in rapid sieve tube sealing. *Front Plant Sci* **4,** 225 (2013).
21. Mylona, P., Linstead, P., Martienssen, R. & Dolan, L. SCHIZORIZA controls an asymmetric cell division and restricts epidermal identity in the Arabidopsis root. *Development* **129,** 4327–4334 (2002).
22. Pernas, M., Ryan, E. & Dolan, L. SCHIZORIZA controls tissue system complexity in plants. *Curr Biol* **20,** 818–823 (2010).
23. Hove, ten, C. A. *et al.* SCHIZORIZA encodes a nuclear factor regulating asymmetry of stem cell divisions in the Arabidopsis root. *Curr Biol* **20,** 452–457 (2010).
24. Dawson, J. H., Musselman, L. J., Wolswinkel, P. & Dörr, I. Biology and control of Cuscuta. *Reviews of Weed Science* **6,** 265–317 (1994).
25. Floyd, S. K. & Bowman, J. L. Gene regulation: ancient microRNA target sequences in plants. *Nature* **428,** 485–486 (2004).
26. Robert-Seilaniantz, A. *et al.* The microRNA miR393 re-directs secondary metabolite biosynthesis away from camalexin and towards glucosinolates. *Plant J*

67, 218–231 (2011).

27. Lu, D. *et al.* A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. *Proc Natl Acad Sci USA* **107,** 496–501 (2010).

28. Hegenauer, V. *et al.* Detection of the plant parasite Cuscuta reflexa by a tomato cell surface receptor. *Science* **353,** 478–481 (2016).

Acknowledgements

We thank the Penn State & Huck Institutes Genomics Core Facility for small RNA-seq services. We thank Hervé Vaucheret, Michael Knoblauch, and Gabriele Monshausen for gifts of *sgs2-1*, *seor1*, and *tir1-1/afb2-3* mutant seed, respectively. We thank Beth Johnson for advice on growing conditions for *C. campestris*. Purchase of the Illumina HiSeq2500 used for small RNA-seq was funded by a major research instrumentation award from the US National Science Foundation [1229046 to MJA and CWD]. This research was supported in part by an award from the US National Science Foundation [1238057 to JHW and CWD] and the National Institute of Food and Agriculture [135997 to JHW].

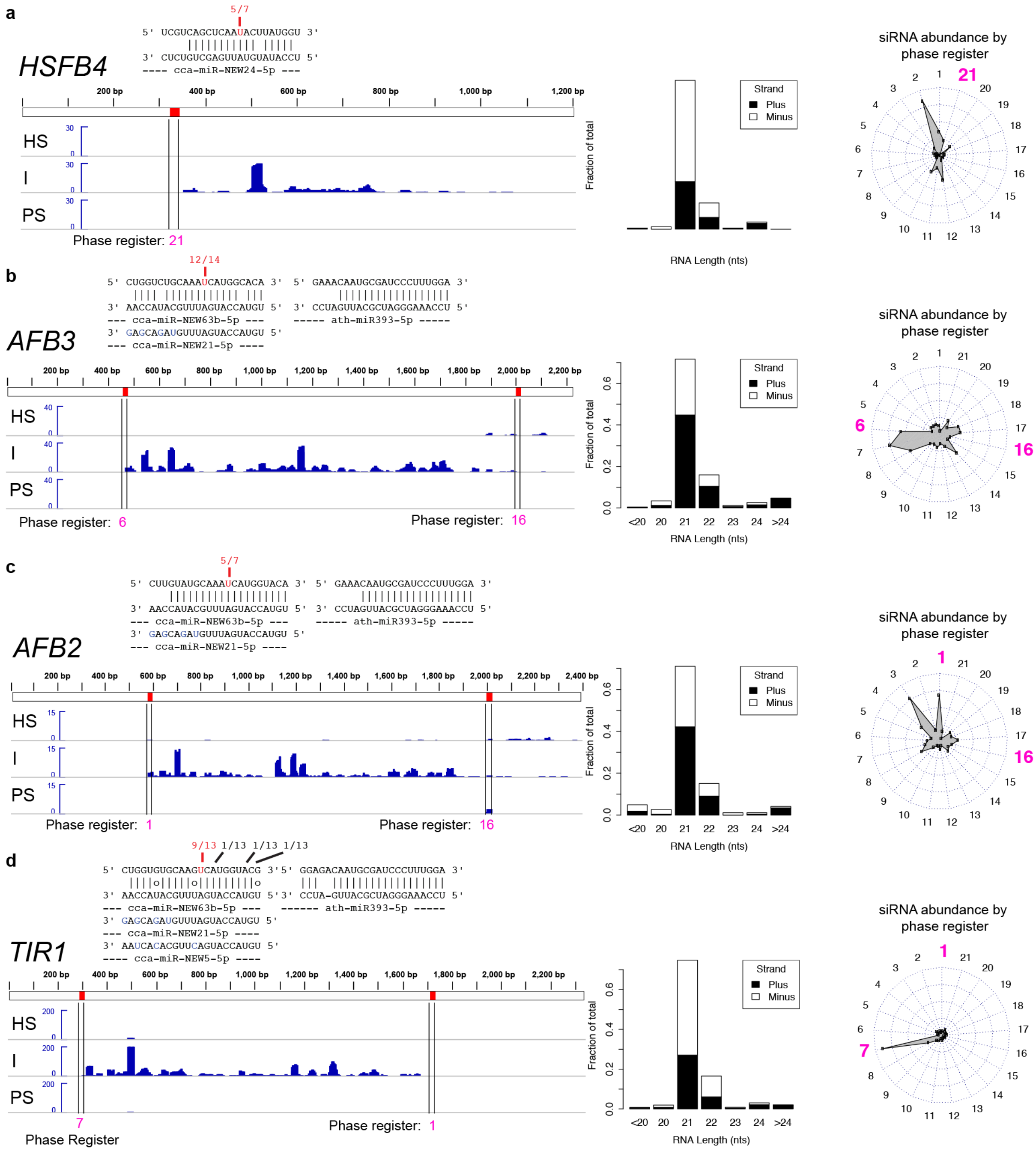
Author Contributions: SS and MJA performed most bioinformatics analysis. SS, MJA, and NRJ prepared figures and tables. GK and JHW cultivated and harvested plant specimens used for initial small RNA-seq experiments. NRJ, SS, and MJA cultivated and harvested plant specimens for other experiments. EW, GK, CWD, and JHW performed whole-genome shotgun DNA sequencing. FW, SS, and NRJ performed RNA blots. SS and MJA performed 5'-RLM-RACE and qRT-PCR. CC constructed small RNA-seq libraries. NRJ and VB performed growth assays. MJA and JHW conceived of the project. MJA wrote and revised the manuscript with input from all other authors.

Author Information

Small RNA-seq data from this work are available at NCBI GEO under accession GSE84955.

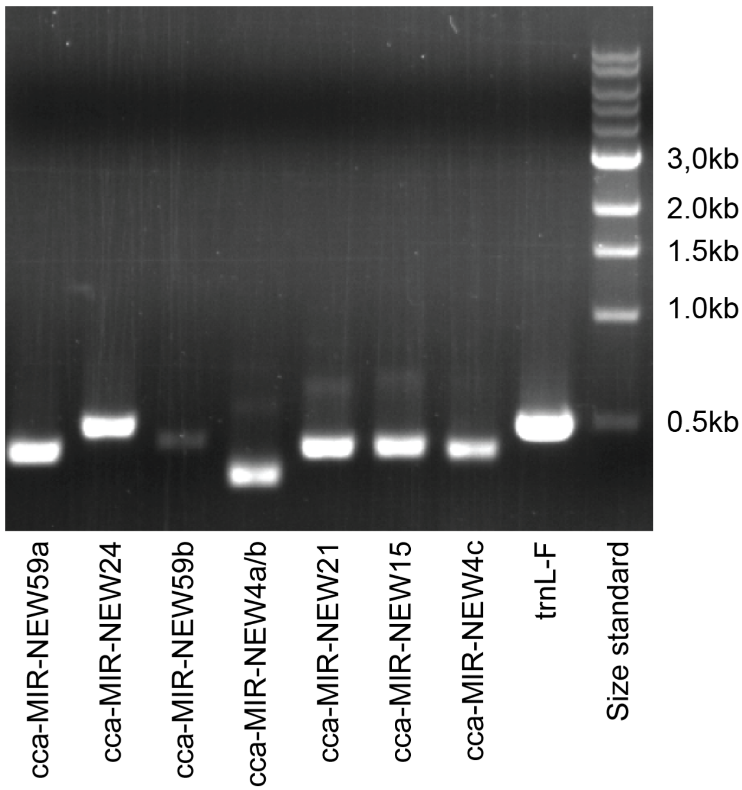
The authors declare no competing financial interests

Correspondence and requests for materials should be addressed to Michael J. Axtell at mja18@psu.edu

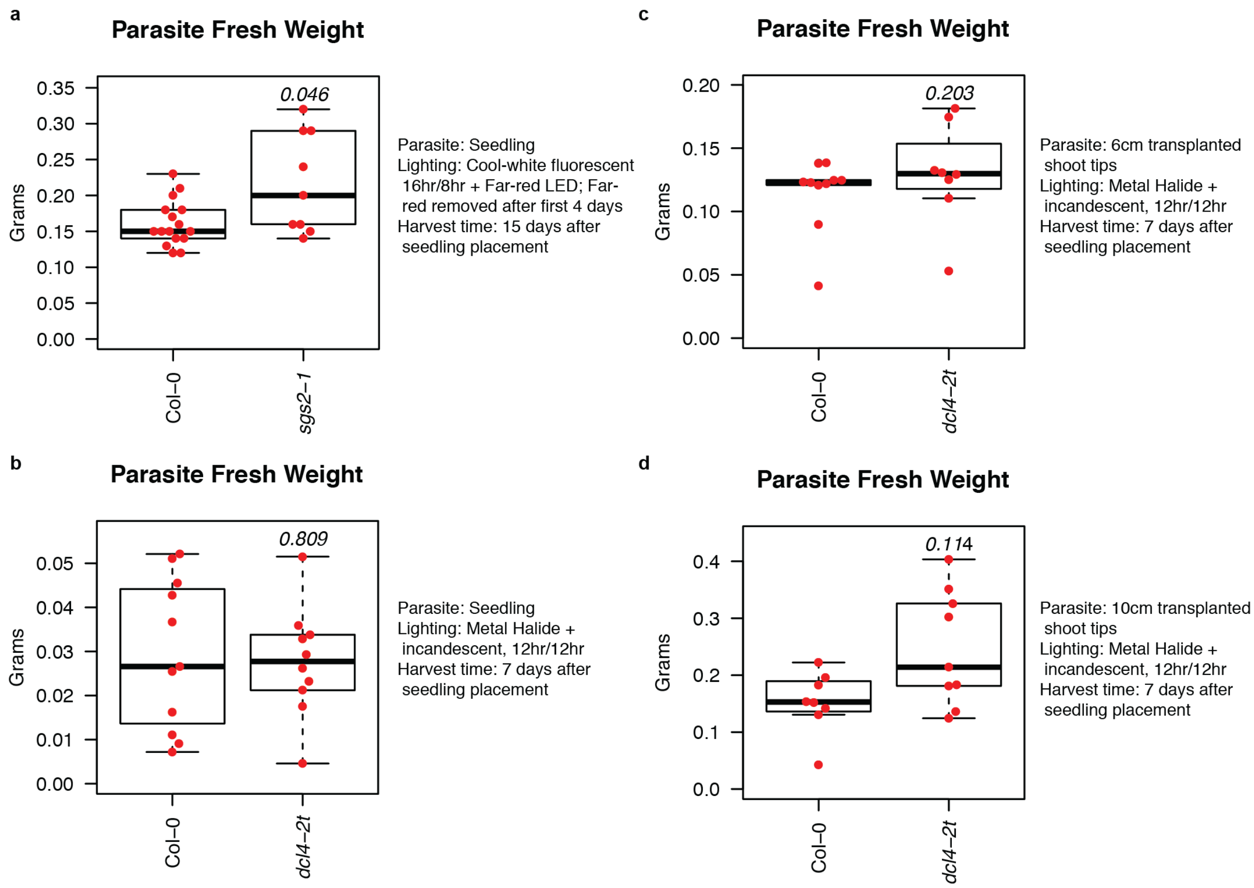


Extended Data Figure 1. *C. campestris* miRNAs cause slicing and phased siRNA production from host mRNAs. **a-d)** Small RNA-seq coverage across the indicated *A.*

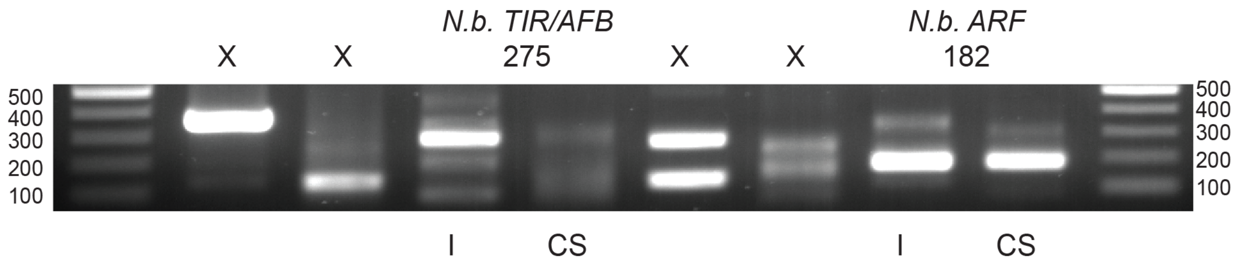
thaliana transcripts are shown in blue for host stem (HS), interface (I), and parasite stem (PS) samples. For display, the two biological replicates of each type were merged. y-axis is in units of reads per million. Red mark and vertical lines show position of complementary sites to *C. campestris* miRNAs, with the alignments shown above. Fractions indicate numbers of 5'-RLM-RACE clones with 5'-ends at the indicated positions; the locations in red are the predicted sites for miRNA-directed slicing remnants. Barcharts show the length and polarity distribution of transcript-mapped siRNAs. Radar charts show the fractions of siRNAs in each of the 21 possible phasing registers; the registers highlighted in magenta are the ones predicted by the miRNA target sites.



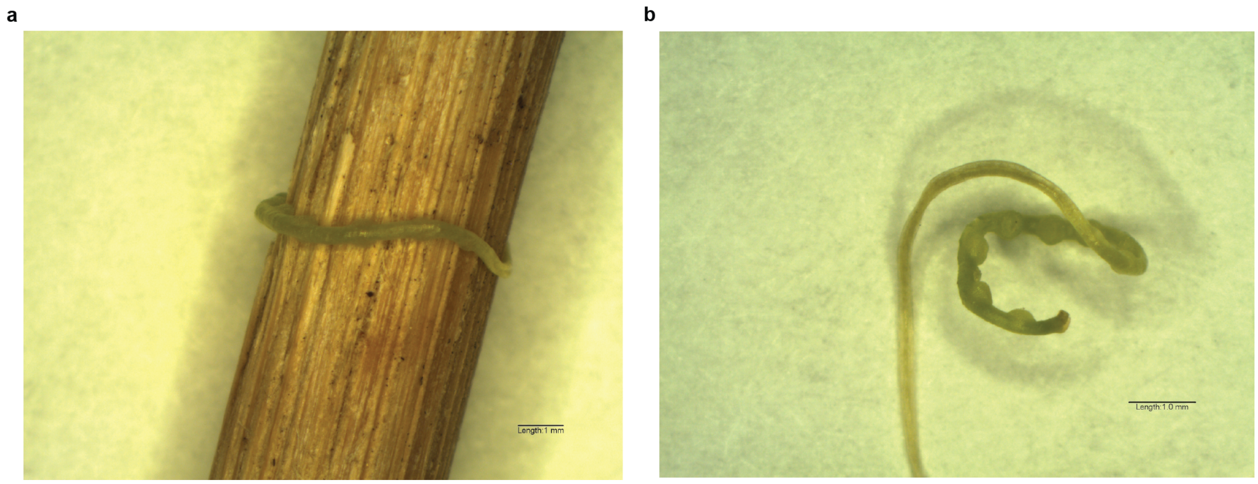
Extended Data Figure 2. PCR of *C. campestris MIRNA* loci. The template for PCR was genomic DNA isolated from *C. campestris* seedlings four days after germination; the seedlings had never attached to nor been near a host plant, ruling out host DNA contamination. trnL-F: Positive control plastid locus.



Extended Data Figure 3. Growth of *C. campestris* on *A. thaliana sgs2-1* and *dcl4-2t* mutants with varying methodologies, as indicated. P-values (Wilcoxon rank-sum tests, unpaired, two-tailed) from comparison of mutant to wild-type (Col-0) are shown. Dots show all data points. Boxplots represent medians (horizontal lines), the central half of the data (boxes), and other data out to 1.5 times the interquartile range (whiskers). **a)** n=16 and 9 for Col-0 and *sgs2-1*, respectively. **b)** n=11 and 10 for Col-0 and *dcl4-2t*, respectively. **c)** n=10 and 8 for Col-0 and *dcl4-2t*, respectively. **d)** n=8 and 9 for Col-0 and *dcl4-2t*, respectively.



Extended Data Figure 4. Uncropped image of *N. benthamiana* 5'-RLM-RACE products. Lanes with 'X' are irrelevant to this study. This is the uncropped version of the image in Figure 4C.



Extended Data Figure 5. *C. campestris* prehaustoria. **a)** *C. campestris* seedling wound around a bamboo stake. **b)** The same seedling, removed from the stake to show the prominent pre-haustorial bumps. Seedling was scarified, germinated on moist paper towels for three days at ~28C, and then placed next to bamboo stake for four days with far-red LED lighting. Approximately 30 such seedlings were used for the 'PH' RNA in Figure 4B. Scales bars: 1mm.

METHODS

Germplasm

Cuscuta was initially obtained from a California tomato field, and seed stocks derived from self-pollination through several generations in the Westwood laboratory. The isolate was initially identified as *Cuscuta pentagona* (Engelm.) *C. pentagona* is very closely related to *C. campestris* (Yunck.), and the two are distinguished by microscopic differences in floral anatomy; because of this they have often been confused[1]. We subsequently determined that our isolate is indeed *C. campestris*. *Arabidopsis thaliana sgs2-1* mutants[2] were a gift from Hervé Vaucheret. *A. thaliana dcl4-2t* mutants (GABI_160G05[3]) were obtained from the Arabidopsis Biological Resource Center. *A. thaliana seor* mutants (GABI-KAT 609F04[4]) were a gift from Michael Knoblauch. *A. thaliana tir1-1/afb2-3* double-mutants[5] were a gift from Gabriele Monshausen. All *A. thaliana* mutants were in the Col-0 background.

Growth conditions and RNA extractions

For initial experiments (small RNA-seq and RNA blots in Figure 1) *A. thaliana* (Col-0) plants were grown in a growth room at 18-20°C with 12-h light per day, illuminated (200 μmol $m^{-2}s^{-1}$) with metal halide (400W, GE multi-vapor lamp) and spot-gro (65W, Sylvania) lamps. *C. campestris* seeds were scarified in concentrated sulfuric acid for 45 min, followed by 5-6 rinses with distilled water and dried. *C. campestris* seeds were placed in potting medium at the base of four-week-old *A. thaliana* seedlings and allowed to germinate and attach to hosts. The *C. campestris* plants were allowed to grow and spread on host plants for an additional three weeks to generate a supply of uniform shoots for use in the experiment. Sections of *C. campestris* shoot tip (~10 cm long) were placed on the floral stem of a fresh set of *A. thaliana* plants. Parasite shoots coiled around the host stems and formed haustorial connections. Tissues from plants that had established *C. campestris* with at least two coils around healthy host stems and clear parasite growth were used in these studies. Control plants were grown under the same conditions as parasitized plants, but were not exposed to *C. campestris*.

For the preparation of tissue-specific small RNA libraries, tissues were harvested after *C. campestris* cuttings had formed active haustorial connections to the host. This was evidenced by growth of the *C. campestris* shoot to a length of at least 10 cm beyond the region of host attachment (7-10 d after infection). Three tissues were harvested from the *A. thaliana*-*C. campestris* associations: 1) 2.5 cm of *A. thaliana* stem above the region of attachment, 2) *A. thaliana* and *C. campestris* stems in the region of attachment (referred to as the interface), 3) 2.5 cm of the parasite stem near the point of attachment. To remove any possible cross-contamination between *A. thaliana* and *C. campestris*, harvested regions of the parasite and host stem were taken 1 cm away from the interface region and each harvested tissue was surface cleaned by immersion for 5 min in 70% ethanol, the ethanol was decanted and replaced, the process was repeated three times and the stems were blotted dry with a Kimwipe after the final rinse. All three sections of tissue were harvested at the same time and material from 20 attachments were pooled for small RNA extraction. Small RNA was extracted from ~ 100 mg of each tissue using the mirPremier microRNA Isolation Kit (Sigma-Aldrich, St. Louis, MO, USA) according to the manufacturer's protocol. Small RNA was analyzed using an Agilent small RNA Kit on a 2100 Bioanalyzer platform.

Samples used for RNA ligase-mediated 5' rapid amplification of cDNA ends (5'-RLM-RACE; Figure 2C) and quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR; Figure 3A) analyses of *A. thaliana* targets were prepared as described above with the following modifications: Col-0 *A. thaliana* hosts were cultivated in a growth room with 16 hr. days, 8 hr. nights, at ~23C, under cool-white fluorescent lamps, attachment of *C. campestris* cuttings was promoted by illumination with far-red LED lighting for 3-5 days, and total RNA was extracted using Tri-reagent (Sigma) per the manufacturer's suggestions, followed by a second sodium-acetate / ethanol precipitation and wash step. Samples used for RNA blots of secondary siRNA accumulation from *A. thaliana* mutants (Figure 3B) were obtained similarly, except that the samples derived from the primary attachments of *C. campestris* seedlings on the hosts instead of from cuttings. In these experiments, scarified *C. campestris* seedlings were first germinated on moistened paper towels for three days at ~28C, then placed adjacent to the host plants with their radicles submerged in a water-filled 0.125ml tube.

C. campestris pre-haustoria (Extended Data Figure 5) were obtained by scarifying, germinating and placing seedlings as described above, next to bamboo stakes in soil, under illumination from cool-white fluorescent lights and far-red emitting LEDs. Seedlings coiled and produced pre-haustoria four days after being placed, and were harvested and used for total RNA extraction (used for RNA blot in Figure 4B) using Tri-reagent as described above. *Nicotiana benthamiana* was grown in a growth room with 16 hr. days, 8 hr. nights, at ~23C, under cool-white fluorescent lamps. Three to four week old plants served as hosts for scarified and germinated *C. campestris* seedlings. Attachments were promoted by three-six days with supplementation by far-red emitting LEDs. Under these conditions, *C. campestris* attached to the petioles of the *N. benthamiana* hosts, not the stems. Interfaces and control petioles from un-parasitized hosts were collected 7-8 days after successful attachments, and total RNA (used for RNA blot in Figure 4B) recovered using Tri-reagent as described above.

Initial sRNA-seq data processing

Small RNA-seq libraries were constructed using the Illumina Tru-Seq small RNA kit per the manufacturer's protocol and sequenced on an Illumina HiSeq2500 instrument. Raw sRNA-seq reads were trimmed to remove 3'-adapters, and filtered for quality and trimmed length >= 16 nts using cutadapt[6] version 1.9.1 with settings "-a TGGAATTCTCGGGTGCCAAGG --discard-untrimmed -m 16 --max-n=0". Trimmed reads that aligned with zero or one mismatch (using bowtie[7] version 1.1.2, settings "-v 1") to the *A. thaliana* plastid genome, the *C. gronovii* plastid genome (*C. gronovii* was the closest relative to *C. campestris* that had a publically available completed plastid genome assembly available), *A. thaliana* rRNAs, tRNAs, snRNAs, or snoRNAs were removed. The remaining 'clean' reads were then aligned to the combined TAIR10 *A. thaliana* reference genome and TAIR10 reference cDNAs, demanding perfect matches, using bowtie[7] version 1.1.2 with settings "-v 0". Sequences that matched were initially designated as host-derived, while those that didn't were initially designated as parasite-derived. Species of origin assignments were then adjusted based on comparing expression levels in the host stem (HS) vs. parasite stem (PS) samples. Let HS_{rpm} indicate the summed reads-per-million value in both HS samples, and HS_{raw} indicate the summed number of raw reads in both HS samples, and similarly for PS_{rpm} and PS_{raw}.

For *A. thaliana*-matched sequences, if $HS_{rpm} / (HS_{rpm} + PS_{rpm}) = 0$ or if $HS_{rpm} / (HS_{rpm} + PS_{rpm}) \leq 0.05$ and $HS_{raw} + PS_{raw} \geq 5$, the sequence was re-assigned to be in the parasite-derived set. Similarly, for sequences not exactly matched to *A. thaliana*, if $HS_{rpm} / (HS_{rpm} + PS_{rpm}) = 1$ or if $HS_{rpm} / (HS_{rpm} + PS_{rpm}) \geq 0.95$ and $HS_{raw} + PS_{raw} \geq 5$, the sequence was re-assigned to the host-derived set. *A. thaliana*-matched RNAs switched to the parasite-derived set likely include small RNAs conserved in both species but primarily expressed by *C. campestris*. RNAs that don't match the *A. thaliana* genome or transcriptome that were switched to the host-derived set likely include RNAs with sequencing errors, or non-templated nucleotides, which are frequent[8]. Note that conserved small RNAs expressed by both species will be generally assigned as host-derived by this method.

Parasite-derived small RNA analysis

Parasite-derived small RNAs between 20 and 24 nts in length were clustered according to sequence similarity. Beginning with a list of parasite-derived 20-24 nt RNA sequences sorted in descending order by abundance (summed across all libraries), each sequence was queried against all remaining others to find less abundant sequences within a Levenshtein edit distance of one or two. The most abundant sequence in each resulting cluster was termed the 'head', while all other variants were termed 'isos'. For computational expediency, this process was limited to 'heads' with a raw read count of 20 or more. Parasite-derived sequences with total read counts of less than 20 that were not found as 'isos' were discarded. This resulted in a total of 29,988 parasite-derived small RNA clusters. The 'head' sequences of each cluster were queried against all mature miRNAs from plants from miRBase[9] version 21; hits within a Levenshtein edit distance of two were recorded (Supplementary Data 1).

The total abundance in each cluster was calculated as the sum of read counts for the 'head' and all its 'iso' sequence variants. These sums were used for differential expression analysis comparing I vs. PS samples (null hypothesis: True difference no more than 2-fold, Benjamini-Hochberg false discovery rate = 0.05) using the R package DESeq2[10] (Supplementary Data 2). The 136 small RNA clusters that were higher in I relative to PS were further analyzed to search for evidence of miRNA-like biogenesis. Paired-end libraries (insert sizes of 200 bp, 340 bp, and 480 bp) and mate-pair libraries (insert sizes of 3 kb and 5 kb) were constructed and sequenced on the Illumina HiSeq 2000 platform to obtain paired-end libraries with 100 x 100 nt reads. Data were adapter-trimmed and quality-filtered using cutadapt[6] version 1.9.1 with settings "-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A AGATCGGAAGAGCGTCGTGTAGGGAAAG -m 40 -q 10 --max-n 1". The resulting cleaned shotgun genomic reads were then queried to find those with exact matches to the 136 parasite-derived small RNAs found to have higher accumulation in I vs. PS. Genomic reads where the small RNA sequence match was within 10 nts from either the 5' or 3' end of the genomic read were retained and used to predict putative RNA secondary structures using RNAfold[11]. Genomic reads whose predicted RNA secondary structure resembled *MIRNA* hairpins were retained and aligned against one another to determine local genomic assemblies. The local assemblies were consolidated based on all vs. all BLAST analysis to remove redundancy (in some cases multiple parasite small

RNAs align to the same local assembly) and annotate families of related sequences (Supplementary Data 3). The parasite-derived small RNAs were aligned against the final set of local genome assemblies housing putative *MIRNA* hairpins using ShortStack 3.4[12] with settings "--nostitch --bowtie_cores 5 --sort_mem 4G --ranmax 20". Small RNA alignment patterns relative to predicted secondary structures were visualized with strucVis (https://github.com/MikeAxtell/srucVis/) version 0.2 (Supplementary Data 4, Supplementary Data 5). A draft genome assembly of *C. campestris* (Westwood, dePamphiis, et al., manuscript in prep.) was used to identify larger flanking regions for selected *C. campestris MIRNA*s, allowing design of PCR primers (Supplementary Data 8) to amplify the loci directly from *C. campestris* genomic DNA (Extended Data Figure 2).

Host-derived small RNA analysis

The final set of host-derived reads were aligned to the TAIR10 *A. thaliana* nuclear genome using ShortStack[12] version 3.4 with settings "--bowtie_cores 5 --sort_mem 4G --nostitch --mincov 1rpm", resulting in the definition of 36,918 small RNA-producing genomic intervals from *A. thaliana*. Small RNA accumulation from these clusters was analyzed to find differentially expressed clusters between I and HS samples (null hypothesis: True difference no more than 2-fold, Benjamini-Hochberg false discovery rate = 0.05) using the R package DESeq2[10] (Supplementary Data 6). Because most differentially-expressed clusters from this genome-wide analysis overlapped annotated transcripts (mRNAs or pri-*MIRNA*s), the host-derived small RNA reads were reanalyzed by alignment to the TAIR10 representative cDNA models using ShortStack version 3.4 with settings "--nohp --nostitch --bowtie_cores 5 --sort_mem 4G" and a "--locifile" specifying the full-length of each transcript as a pre-defined locus. Differentially expressed loci between I and HS were identified as described above (Supplementary Data 6).

RNA blots

Small RNA gel blots were performed as previously described[13] with modifications. For the blots in Figure 1B, small RNAs (1.8 micrograms) from each sample were separated on 15% TBE-Urea Precast gels (Bio-Rad), transblotted onto the Hybond NX membrane and cross-linked using 1-ethyl-3-(3-dimethylamonipropyl) carbodiimide[14]. Hybridization was carried out in 5×SSC, 2×Denhardt's Solution, 20 mM sodium phosphate (pH 7.2), 7% SDS with 100 μg/ml salmon testes DNA (Sigma-Aldrich). Probe labeling, hybridization and washing were performed as described[13]. Radioactive signals were detected using Typhoon FLA 7000 (GE Healthcare). Membranes were stripped in between hybridizations by washing with 1% SDS for 15 min at 80°C and exposed for at least 24 h to verify complete removal of probe before re-hybridization. Sequences of probes are listed below. Blots in Figures 3B and 4B were performed similarly, except that 12 micrograms of total RNA were used instead. Probe sequences are listed in Supplementary Data 8.

5' RNA ligase-mediated rapid amplification of cDNA ends (5'-RLM-RACE)

Five micrograms of total RNA were ligated to one microgram of a 44 nucleotide RNA adapter (Supplementary Data 8) using a 20ul T4 RNA ligase 1 reaction (NEB) per the manufacturer's instructions for a one-hour incubation at 37C. The reaction was then diluted with 68ul of water and 2ul of 0.5M EDTA pH 8.0, and incubated at 65C for 15 minutes to inactivate the ligase. Sodium acetate pH5.2 was added to a final concentration of 0.3M, and the RNA precipitated with ethanol. The precipitated and washed RNA was resuspended in 10ul of water. 3.33ul of this sample was used as the template in a reverse transcription reaction using random primers and the Protoscript II reverse transcriptase (NEB) per the manufacturer's instructions. The resulting cDNA was used as template in first round PCR using a 5' primer matching the RNA adapter and a 3' gene-specific primer (Supplementary Data 8). 1ul of the first round PCR product was used as the template for nested PCR with nested primers (Supplementary Data 8). Gene-specific primers for *A. thaliana* cDNAs were based on the representative TAIR10 transcript models, while those for *N. benthamiana* cDNAs were based on the version 0.4.4 transcripts (Sol Genomics Network[15]). *N. benthamiana TIR/AFB* is transcript ID *NbS00011315g0112.1*; *N. benthamiana ARF* is transcript ID *NbS00059497g0003.1*. Bands were gel-purified from agarose gels and cloned into pCR4-TOPO (Life Tech). Inserts from individual clones were recovered by colony PCR and subject to Sanger sequencing.

Quantitative reverse-transcriptase PCR (qRT-PCR)

Total RNA used for qRT-PCR was first treated with DNaseI (RNase-free; NEB) per the manufacturer's instructions, ethanol precipitated, and resuspended. 2 micrograms of treated total RNA was used for cDNA synthesis using the High Capacity cDNA Synthesis Kit (Applied Biosystems) per the manufacturer's instructions. PCR reactions used PerfeCTa SYBR Green FastMix (Quanta bio) on an Applied Biosystems StepONE-Plus quantitative PCR system per the manufacturer's instructions. Primers (Supplementary Data 8) were designed to span the miRNA target sites, to ensure that only uncleaved mRNAs were measured. Three reference mRNAs were used: *ACTIN*, *PP2A (PP2A sub-unit PDF2*; *At1g13320*), and *TIP41-l* (*TIP41-like; At4g34270*)[16]. Raw Ct values were used to calculate relative normalized expression values to each reference mRNA separately, and the final analysis took the median relative expression values between the *ACTIN*- and *TIP41-l* normalized data.

C. campestris growth assays

C. campestris seedlings were scarified, pre-germinated, and placed next to hosts in 0.125ml water-filled tubes under cool-white fluorescent lighting supplemented with far-red emitting LEDs (16hr day, 8hr night) at ~ 23C as described above. After a single attachment formed (4 days), far-red light supplementation was removed to prevent secondary attachments. After 18 more days of growth, entire *C. campestris* vines were removed and weighed (Figures 3C-3D). Multiple additional growth trials were performed specifically on the *dcl4-2t* and *sgs2-1* mutant hosts under varying conditions (Extended Data Figure 4): Trial one using *sgs2*-1 used similar conditions except that harvest was perform 11 days after removal of far-red light (Extended Data Figure 3A). Trial two examined *dcl4-2t* and used 3cm seedlings in a 12-h per day light cycle, illuminated (200 μmol m-2 s -1) with metal halide (400W, GE multi-vapor lamp) and spotgro (65W, Sylvania) lamps, and measured biomass seven days after attachment (Extended Data Figure 3B). Trials three and four were performed on *dcl4-2t* with the same growth regime and seven-day timing, except using 6cm and 10cm *C. campestris* shoot tips harvested from ~ 1 month old plants grown on beets (*Beta vulgaris*) as the starting material (Extended Data Figures 3C-D).

miRNA target predictions

To find probable orthologs for *Arabidopsis thaliana* genes of interest, the *A. thaliana* protein sequences were used as queries for a BLASTP analysis of the 31 eudicot proteomes available on Phytozome 11 (https://phytozome.jgi.doe.gov/pz/portal.html#). Transcript sequences for the top 100 hits were retrieved. Two miRNA query sets were prepared. The first contained mature miRNAs from interface-induced *C. campestris* *MIRNA* loci. For each locus in Supplemental Table 3, the sequence with a higher total read count was retained. Additionally, if the mature sequence from the other strand of the hairpin had at least 100 reads and began with a 5'-U, it was also put into the interface-induced query set. A second query set consisting of conserved miRNAs expressed by *C. campestris* was curated by taking all small RNA 'head' sequences (see above) that a) matched an annotated mature miRNA sequence from a plant species in miRBase 21, and b) had a ratio of HS / (HS + PS) of <= 0.95. Targets were predicted from the probable 31-species with a maximum score of 4.5 using targetfinder.pl (https://github.com/MikeAxtell/TargetFinder/) version 0.1. The *BIK1, SEOR1*, *TIR/AFB*, *HSFB4*, and *GAPDH* probable orthologs were searched against the interface-induced queries, while the *PHB/PHV/REV* probable orthologs were searched against the conserved miRNA queries.

N. benthamiana orthologs of *A. thaliana TIR1/AFB2/AFB3* and of *ARF17* were found based on BLAST-P searches against the version 0.4.4 *N. benthamiana* protein models at Sol Genomics Network[15], and miRNA target sites for cca-miR-NEW21 and miR160, respectively, predicted using targetfinder.pl as above.

Code availability

ShortStack version 3.4[12] (small RNA-seq analysis), strucVis version 0.2 (visualization of predicted RNA secondary structures with overlaid small RNA-seq depths), and Shuffler.pl/targetfinder.pl version 0.1 (prediction of miRNA targets controlling for false discovery rate) are all freely available at https://github.com/MikeAxtell. Cutadapt version 1.9.1[6] is freely available at http://cutadapt.readthedocs.io/en/stable/index.html. The R package DESeq2[10] is freely available at http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html.

References Cited (Methods)

1. Costea, M., García, M. A., Baute, K. & Stefanović, S. Entangled evolutionary history of Cuscuta pentagona clade: A story involving hybridization and Darwin in

the Galapagos. *Taxon* **64,** 1225–1242 (2015).

2. Elmayan, T. *et al.* Arabidopsis mutants impaired in cosuppression. *Plant Cell* **10,** 1747–1758 (1998).
3. Xie, Z., Allen, E., Wilken, A. & Carrington, J. C. DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. *Proc Natl Acad Sci USA* **102,** 12984–12989 (2005).
4. Froelich, D. R. *et al.* Phloem ultrastructure and pressure flow: Sieve-Element-Occlusion-Related agglomerations do not affect translocation. *Plant Cell* **23,** 4428–4445 (2011).
5. Parry, G. *et al.* Complex regulation of the TIR1/AFB family of auxin receptors. *Proc Natl Acad Sci USA* **106,** 22540–22545 (2009).
6. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* (2011).
7. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10,** R25 (2009).
8. Wang, F., Johnson, N. R., Coruh, C. & Axtell, M. J. Genome-wide analysis of single non-templated nucleotides in plant endogenous siRNAs and miRNAs. *Nucleic Acids Res* (2016). doi:10.1093/nar/gkw457
9. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42,** D68–73 (2014).
10. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15,** 550 (2014).

11. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6,** 26 (2011).

12. Johnson, N. R., Yeoh, J. M., Coruh, C. & Axtell, M. J. Improved Placement of Multi-Mapping Small RNAs. *G3 (Bethesda)* (2016). doi:10.1534/g3.116.030452

13. Cho, S. H., Coruh, C. & Axtell, M. J. miR156 and miR390 regulate tasiRNA accumulation and developmental timing in Physcomitrella patens. *Plant Cell* **24,** 4837–4849 (2012).

14. Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced detection of small RNA. *Nat Protoc* **3,** 1077–1084 (2008).

15. Bombarely, A. *et al.* A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research. *Mol Plant Microbe Interact* **25,** 1523–1530 (2012).

16. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W.-R. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol* **139,** 5–17 (2005).

SI Guide

Supplementary Data 1: Clustered parasite-derived small RNAs. Tab-delimited text file, .zip compressed. Columns 1: Sequence name, Column 2: Sequence ID with embedded raw read counts from each of the six libraries, Column 3: Sequence, Column 4: Total counts, Column 5: Similarity to annotated mature miRNAs in miRBase 21. (x: no similarity within edit distance of 2, NA: not analyzed).

Supplementary Data 2: Differential expression analysis of *C. campestris* small RNA clusters. Excel (.xlsx) format.

Supplementary Data 3: *C. campestris MIRNA* loci. Excel (.xlsx) format.

Supplementary Data 4: Details of *C. campestris MIRNA* loci: Text-based sequences, predicted secondary structures, and aligned small RNA reads (all six libraries). Lower-case letters indicate small RNA bases that are mismatched to the genomic sequence. Plain-text (ASCII) format.

Supplementary Data 5: Post-script files showing *C. campestris* hairpins overlaid with color-codes representing total read-depth (all six libraries). Tar-achived/gzip-compressed set of post-script files.

Supplementary Data 6: Differential expression analysis of *A. thaliana* small RNA clusters. Excel (.xlsx) format.

Supplementary Data 7: Predicted miRNA targets in multiple species for selected gene families. Excel (.xlsx) format.

Supplementary Data 8: Oligonucleotide sequences. Excel (.xlsx) format.