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

46 pathogen *Verticillium dahliae* where they target fungal mRNAs encoding virulence
47 factors[11].
48 We hypothesized that naturally occurring small RNAs might be exchanged
49 across the *C. campestris* haustorium and affect gene expression in the recipient
50 species. To test this hypothesis, we profiled small RNA expression from *C. campestris*
51 grown on *A. thaliana* hosts using high-throughput small RNA sequencing (sRNA-seq).
52 Two biological replicates each from three tissues were analyzed: Parasite stem (PS),
53 comprising a section of *C. campestris* stem above the site of haustorium formation;
54 Interface (I), comprising *C. campestris* stem with haustoria with associated *A. thaliana*
55 stem tissue; and Host stem (HS), comprising sections of *A. thaliana* stems above the
56 interface region, as previously described[12]. Sequenced small RNAs were designated as
57 host-derived or parasite-derived based on alignment to the *A. thaliana* reference
58 genome assembly and expression analysis of HS vs. PS samples. Host-derived reads
59 were aligned to the *A. thaliana* reference genome assembly, while parasite-derived
60 sRNA-seq reads were initially clustered by sequence similarity (Supplementary Data 1).
61 151 *C. campestris* small RNAs were differentially expressed between I and PS
62 samples (Supplementary Data 2), most of which (136) were up-regulated in I relative to
63 PS (Figure 1A). RNA blots confirmed the strong accumulation in I samples for several
64 tested small RNAs (Figure 1B). Whole-genome shotgun DNA sequence data from *C.*
65 *campestris* were used to produce short, local genome assemblies containing the 136
66 interface-induced small RNAs. This allowed alignment of all parasite-derived small
67 RNAs to these genomic loci. We were then able to discern microRNA loci from other
68 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].

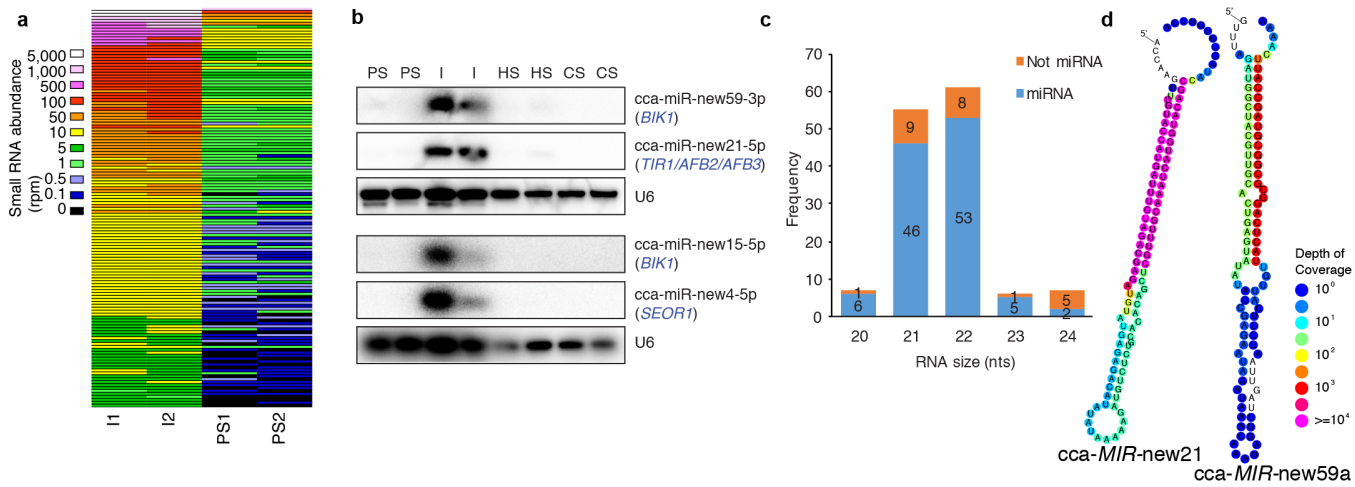


Figure 1. *C. campestris* miRNAs induced at the haustorial interface. **a)** Heatmap showing accumulation of 136 *C. campestris* small RNAs that are significantly (FDR = 0.05) more than two-fold higher in interface relative to parasite stem samples. I: Interface, PS: Parasite stem, rpm: reads per million. **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

86 structures are overlaid with colors representing depth of small RNA sequencing. See
87 Supplementary Data 3-5 for details.
88 ----------
89
90 Differential expression analysis of *A. thaliana* small RNAs found six loci induced
91 in I relative to HS samples (Figure 2A-B, Extended Data Figure 1, Supplementary Data
92 6), each deriving from *A. thaliana* mRNAs with known functions: *TIR1*, *AFB2*, and *AFB3*,
93 which encode related and partially redundant auxin receptors[16], *BIK1*, which encodes a
94 plasma membrane-localised kinase required for both pathogen-induced and
95 developmental signaling[17,18], *SEOR1*, which encodes a major phloem protein that
96 accumulates in filamentous networks in sieve tube elements and reduces photosynthate
97 loss from the phloem upon injury[19,20], and *SCHIZORHIZA/HSFB4*, which encodes a
98 predicted transcriptional repressor that is required to maintain stem-cell identity in
99 roots[21-23]. The induced siRNAs from these mRNAs resembled secondary siRNAs in
100 their size distributions, double-stranded accumulation, and phasing (Figure 2A-B;
101 Extended Data Figure 1).
102 Each of these six mRNAs has either one or two complementary sites to interface-
103 induced *C. campestris* 22 nt miRNAs (Figure 2A-B; Extended Data Figure 1). The *TIR1,*
104 *AFB2,* and *AFB3* mRNAs are also targeted by the conserved miRNA, miR393, at
105 distinct sites (Extended Data Figure 1). Analysis of uncapped mRNA fragments using 5'-
106 RNA ligase-mediated rapid amplification of cDNA ends (5'-RLM-RACE) found strong
107 evidence for miRNA-directed cleavage at all of the complementary sites to *C.*
108 *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.

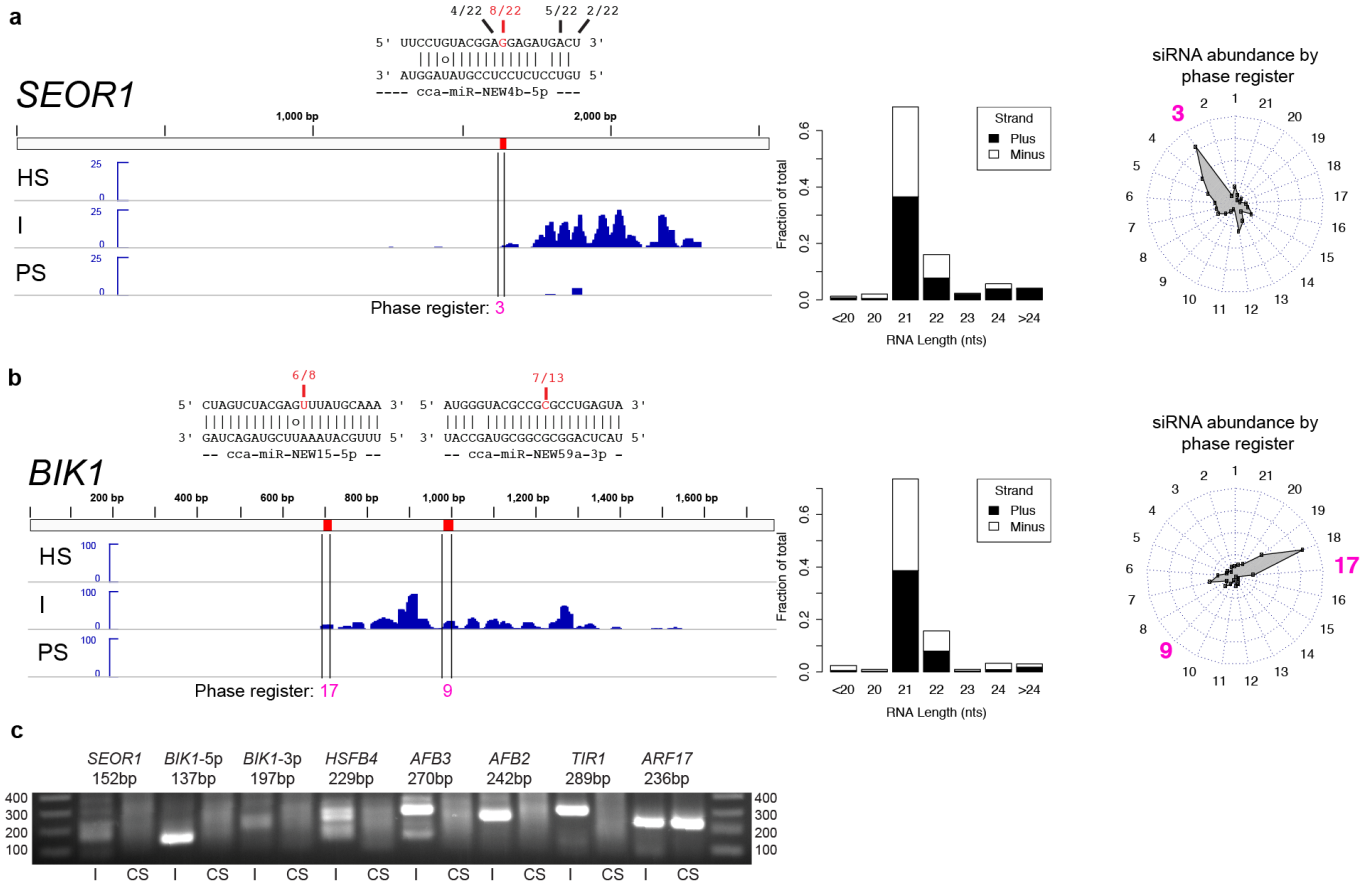
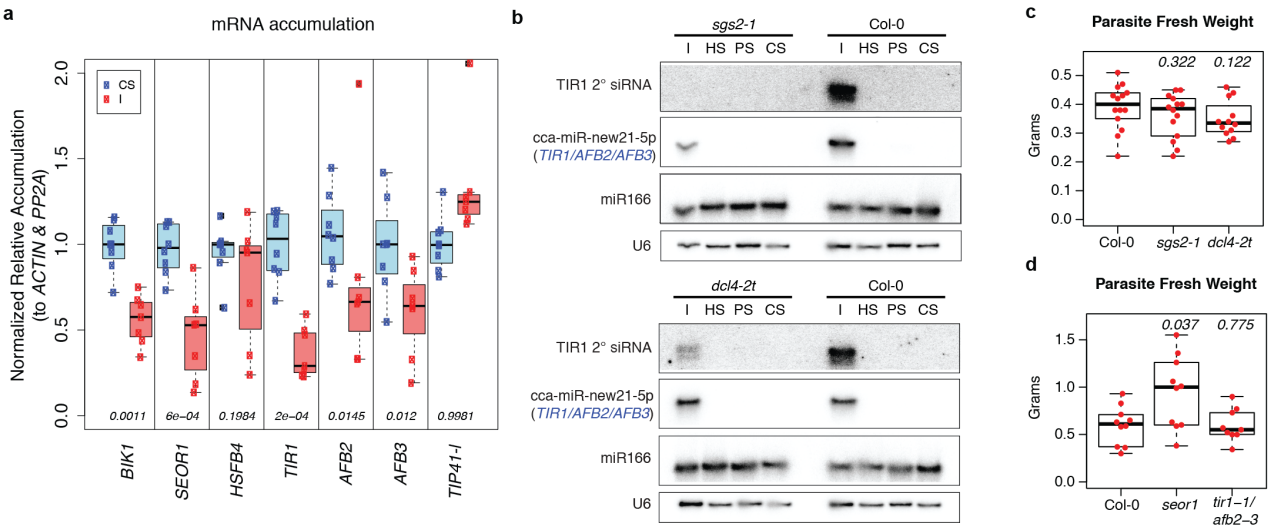


Figure 2. *C. campestris* miRNAs cause slicing and phased siRNA production from host mRNAs. **a)** Small RNA-seq coverage (reads per million) across the *A. thaliana SEOR1* transcript is shown in blue for host stem (HS), interface (I), and parasite stem (PS)

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.

144 ----------



145

146 **Figure 3**. Effects of *C. campestris* miRNAs and their targets. **a)** *A. thaliana* mRNA
147 accumulation levels in I (interface) vs. CS (control stems) during *C. campestris*
148 parasitism, assessed by qRT-PCR. Data from 7 or 8 biological replicates are plotted
149 (dots), and boxplots indicate the median (horizontal lines), 1st-3rd quartile range
150 (boxes), and up to 1.5 x the interquartile range (whiskers). Numbers indicate p-values
151 comparing CS and I samples (Wilcoxon rank-sum test, unpaired, one-tailed). *TIP41-I* is
152 a control mRNA. **b)** RNA blots from *C. campestris* infestations of the indicated *A.*
153 *thaliana* genotypes. I: interface, HS: host stem, PS: parasite stem, CS: control stem. **c)**
154 Accumulation of *C. campestris* biomass on *A. thaliana* hosts of the indicated genotypes
155 18 days post-attachment. P-values (Wilcoxon rank-sum tests, unpaired, two-tailed) from
156 comparison of mutant to wild-type (Col-0) are shown. Boxplot conventions as in panel a.
157 n=14, 14, and 12 for Col-0, *sgs2-1*, and *dcl4-2t*, respectively. **d)** As in c, except for
158 *seor1* and *tir1-1/afb2-3*. n=10, 10, and 9 for Col-0, *seor1*, and *tir1-1/afb2-3*, respectively.

159 ----------

160

161 In repeated trials with varying methodologies we did not observe consistent
162 significant differences in parasite fresh weight using *dcl4-2t* and *sgs2-1* mutants as
163 hosts (Figure 3C, Extended Data Figure 3). This implies that loss of induced secondary
164 siRNAs is not sufficient to detectably affect parasite biomass accumulation. We also
165 observed no significant difference in growth on a *tir1-1/afb2-3* double mutant (Figure
166 3D). However, there was a significant increase ($p = 0.037$) of parasite fresh weight after
167 growth on the *seor1* mutant (Figure 3D). This demonstrates that *SEOR1*, which is
168 targeted by a *C. campestris* miRNA during parasitism, restricts growth of *C. campestris*.
169 *C. campestris* has a broad host range, primarily among eudicots[24]. Therefore, we
170 searched for miRNA complementary sites for the six relevant *C. campestris* miRNAs in
171 eudicot orthologs of *TIR1, AFB2, AFB3*, *SEOR1, BIK1,* and *HSFB4*. We used *GAPDH*
172 orthologs with the same six miRNAs as a negative control query, and *PHV* orthologs
173 with a miR166 query[25] as a positive control. Probable orthologs of *BIK1, SEOR1,* and
174 *TIR/AFB* were predicted targets of interface-induced miRNAs in many eudicot species,
175 while predicted targets in probable *HSFB4* orthologs were less frequent (Figure 4A,
176 Supplementary Data 7). We conclude that the induced *C. campestris* miRNAs
177 collectively would be able to target *TIR1, AFB2, AFB3, SEOR1,* and *BIK1* orthologs in
178 many eudicot species, and *HSFB4* in some species.
179 ----------
180

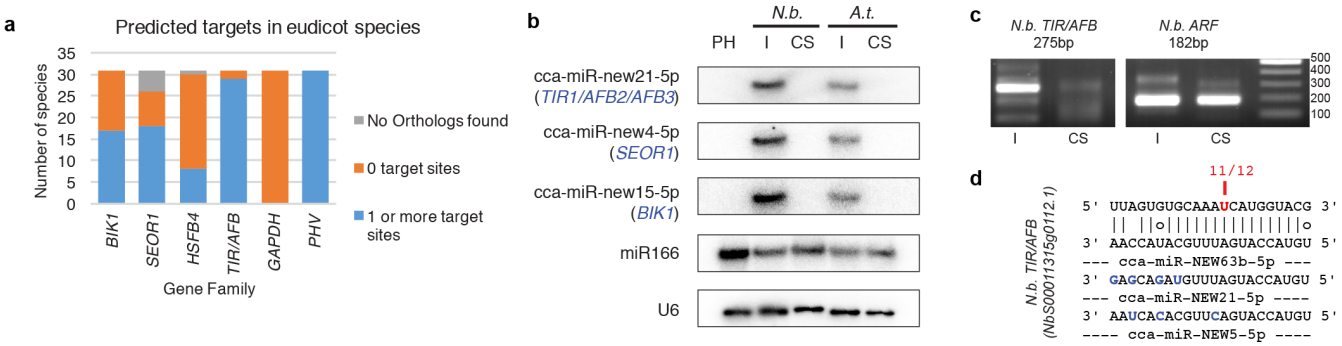


Figure 4. Conservation of host mRNA targeting by *C. campestris*. **a)** Predicted targets in the indicated eudicot gene families. See Supplementary Data 7 for details. **b)** RNA blots of the indicated small RNAs from interface (I) and control stem (CS) samples of *C. campestris* infested *N. benthamiana* (*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

200 provided strong evidence for cleavage of an *N. benthamiana TIR1/AFB* ortholog by a *C.*
201 *campestris* miRNA (Figures 4C-D; Extended Data Figure 4). This directly demonstrates
202 that some of the same *C. campestris* miRNAs are expressed upon interaction with
203 different species of host plants, and that at least one orthologous host mRNA is targeted
204 in multiple species. None of the interface-induced miRNAs we tested were detectable
205 from *C. campestris* pre-haustoria sampled from seedling tips that had coiled around
206 dead bamboo stakes instead of a live host (Figure 4B; Extended Data Figure 5). This
207 suggests that expression of these miRNAs requires prior contact with a living host.
208 These data demonstrate that *C. campestris* induces a large number of miRNAs
209 in the functional haustorium, and that some of them direct cleavage of host mRNAs and
210 reduce their accumulation. Many of the induced miRNAs are 22 nts long, and
211 associated with secondary siRNA production from their host targets. Several of the
212 targets are functionally linked to plant pathogenesis: Manipulation of *TIR1/AFB2/AFB3*
213 accumulation levels affects bacterial pathogenesis and defense signaling[26]. The BIK1
214 kinase is a central regulator of pathogen-induced signaling[27], and kinase-dependent
215 pathogen signaling is known to play a key role in tomato resistance against *C. reflexa*[28].
216 Perhaps the most intriguing target is *SEOR1*, which encodes a very abundant protein
217 present in large, filamentous agglomerations in phloem sieve-tube elements[19]. *seor1*
218 mutants have an increased loss of sugars from detached leaves[20], and our data show
219 that *seor1* mutants also support increased *C. campestris* biomass accumulation. A key
220 function of the *C. campestris* haustorium is to take nutrients from the host phloem;
221 targeting *SEOR1* could be a strategy to increase sugar uptake from host phloem for the
222 benefit of the parasite. Overall, these data suggest that *C. campestris trans*-species

223 miRNAs might function as virulence factors to remodel host gene expression to its
224 advantage during parasitism.
225
226 **Methods** [separate online only document].
227
228 **References**

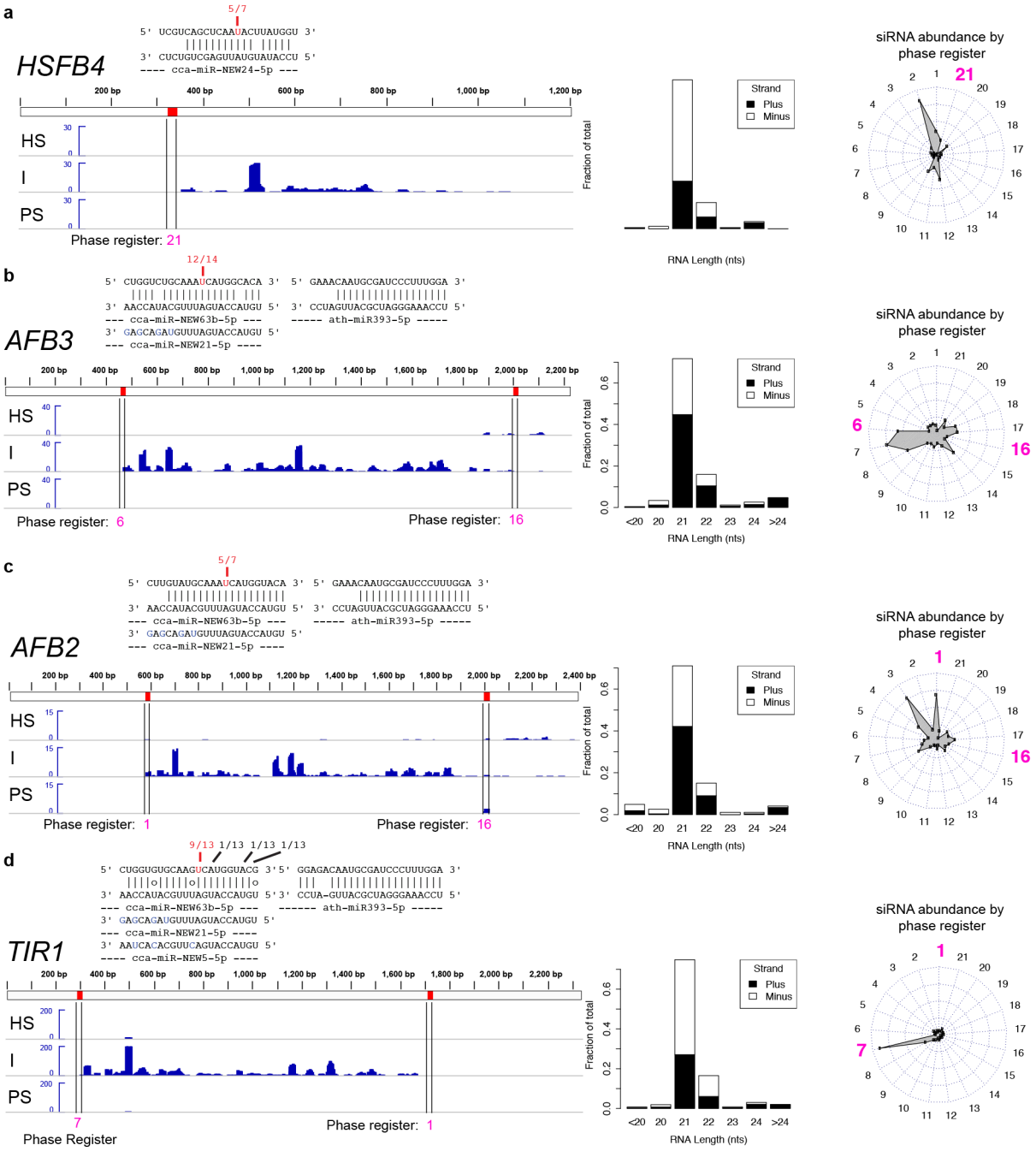
229 1. Kim, G. & Westwood, J. H. Macromolecule exchange in Cuscuta-host plant
230 interactions. *Curr Opin Plant Biol* **26,** 20–25 (2015).
231 2. Fei, Q., Xia, R. & Meyers, B. C. Phased, Secondary, Small Interfering RNAs in
232 Posttranscriptional Regulatory Networks. *Plant Cell* **25,** 2400–2415 (2013).
233 3. Baulcombe, D. C. VIGS, HIGS and FIGS: small RNA silencing in the interactions
234 of viruses or filamentous organisms with their plant hosts. *Curr Opin Plant Biol* **26,**
235 141–146 (2015).
236 4. Weiberg, A., Bellinger, M. & Jin, H. Conversations between kingdoms: small
237 RNAs. *Curr. Opin. Biotechnol.* **32,** 207–215 (2015).
238 5. Nowara, D. *et al.* HIGS: host-induced gene silencing in the obligate biotrophic
239 fungal pathogen Blumeria graminis. *Plant Cell* **22,** 3130–3141 (2010).
240 6. Huang, G., Allen, R., Davis, E. L., Baum, T. J. & Hussey, R. S. Engineering broad
241 root-knot resistance in transgenic plants by RNAi silencing of a conserved and
242 essential root-knot nematode parasitism gene. *Proc Natl Acad Sci USA* **103,**
243 14302–14306 (2006).
244 7. Baum, J. A. *et al.* Control of coleopteran insect pests through RNA interference.
245 *Nat Biotechnol* **25,** 1322–1326 (2007).

246 8. Alakonya, A. *et al.* Interspecific RNA interference of SHOOT MERISTEMLESS-
247 like disrupts Cuscuta pentagona plant parasitism. *Plant Cell* **24,** 3153–3166
248 (2012).
249 9. Weiberg, A. *et al.* Fungal small RNAs suppress plant immunity by hijacking host
250 RNA interference pathways. *Science* **342,** 118–123 (2013).
251 10. Wang, M. *et al.* Bidirectional cross-kingdom RNAi and fungal uptake of external
252 RNAs confer plant protection. *Nature Plants* **2,** 16151 (2016).
253 11. Zhang, T. *et al.* Cotton plants export microRNAs to inhibit virulence gene
254 expression in a fungal pathogen. *Nature Plants* **2,** 16153 (2016).
255 12. Kim, G., LeBlanc, M. L., Wafula, E. K., dePamphilis, C. W. & Westwood, J. H.
256 Plant science. Genomic-scale exchange of mRNA between a parasitic plant and
257 its hosts. *Science* **345,** 808–811 (2014).
258 13. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence
259 microRNAs using deep sequencing data. *Nucleic Acids Res* **42,** D68–73 (2014).
260 14. Chen, H.-M. *et al.* 22-Nucleotide RNAs trigger secondary siRNA biogenesis in
261 plants. *Proc Natl Acad Sci USA* **107,** 15269–15274 (2010).
262 15. Cuperus, J. T. *et al.* Unique functionality of 22-nt miRNAs in triggering RDR6-
263 dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat Struct*
264 *Mol Biol* (2010). doi:10.1038/nsmb.1866
265 16. Dharmasiri, N. *et al.* Plant development is regulated by a family of auxin receptor
266 F box proteins. *Dev Cell* **9,** 109–119 (2005).
267 17. Veronese, P. *et al.* The membrane-anchored BOTRYTIS-INDUCED KINASE1
268 plays distinct roles in Arabidopsis resistance to necrotrophic and biotrophic

269 pathogens. *Plant Cell* **18,** 257–273 (2006).
270 18. Lin, W. *et al.* Inverse modulation of plant immune and brassinosteroid signaling
271 pathways by the receptor-like cytoplasmic kinase BIK1. *Proc Natl Acad Sci USA*
272 **110,** 12114–12119 (2013).
273 19. Froelich, D. R. *et al.* Phloem ultrastructure and pressure flow: Sieve-Element-
274 Occlusion-Related agglomerations do not affect translocation. *Plant Cell* **23,**
275 4428–4445 (2011).
276 20. Jekat, S. B. *et al.* P-proteins in Arabidopsis are heteromeric structures involved in
277 rapid sieve tube sealing. *Front Plant Sci* **4,** 225 (2013).
278 21. Mylona, P., Linstead, P., Martienssen, R. & Dolan, L. SCHIZORIZA controls an
279 asymmetric cell division and restricts epidermal identity in the Arabidopsis root.
280 *Development* **129,** 4327–4334 (2002).
281 22. Pernas, M., Ryan, E. & Dolan, L. SCHIZORIZA controls tissue system complexity
282 in plants. *Curr Biol* **20,** 818–823 (2010).
283 23. Hove, ten, C. A. *et al.* SCHIZORIZA encodes a nuclear factor regulating
284 asymmetry of stem cell divisions in the Arabidopsis root. *Curr Biol* **20,** 452–457
285 (2010).
286 24. Dawson, J. H., Musselman, L. J., Wolswinkel, P. & Dörr, I. Biology and control of
287 Cuscuta. *Reviews of Weed Science* **6,** 265–317 (1994).
288 25. Floyd, S. K. & Bowman, J. L. Gene regulation: ancient microRNA target
289 sequences in plants. *Nature* **428,** 485–486 (2004).
290 26. Robert-Seilaniantz, A. *et al.* The microRNA miR393 re-directs secondary
291 metabolite biosynthesis away from camalexin and towards glucosinolates. *Plant J*

292 **67,** 218–231 (2011).
293 27. Lu, D. *et al.* A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin
294 receptor complex to initiate plant innate immunity. *Proc Natl Acad Sci USA* **107,**
295 496–501 (2010).
296 28. Hegenauer, V. *et al.* Detection of the plant parasite Cuscuta reflexa by a tomato
297 cell surface receptor. *Science* **353,** 478–481 (2016).
298
299
300 **Acknowledgements**
301 We thank the Penn State & Huck Institutes Genomics Core Facility for small RNA-seq
302 services. We thank Hervé Vaucheret, Michael Knoblauch, and Gabriele Monshausen for
303 gifts of *sgs2-1*, *seor1*, and *tir1-1/afb2-3* mutant seed, respectively. We thank Beth
304 Johnson for advice on growing conditions for *C. campestris*. Purchase of the Illumina
305 HiSeq2500 used for small RNA-seq was funded by a major research instrumentation
306 award from the US National Science Foundation [1229046 to MJA and CWD]. This
307 research was supported in part by an award from the US National Science Foundation
308 [1238057 to JHW and CWD] and the National Institute of Food and Agriculture [135997
309 to JHW].
310
311 **Author Contributions**: SS and MJA performed most bioinformatics analysis. SS, MJA,
312 and NRJ prepared figures and tables. GK and JHW cultivated and harvested plant
313 specimens used for initial small RNA-seq experiments. NRJ, SS, and MJA cultivated
314 and harvested plant specimens for other experiments. EW, GK, CWD, and JHW

315 performed whole-genome shotgun DNA sequencing. FW, SS, and NRJ performed RNA
316 blots. SS and MJA performed 5'-RLM-RACE and qRT-PCR. CC constructed small RNA-
317 seq libraries. NRJ and VB performed growth assays. MJA and JHW conceived of the
318 project. MJA wrote and revised the manuscript with input from all other authors.
319
320 **Author Information**
321 Small RNA-seq data from this work are available at NCBI GEO under accession
322 GSE84955.
323
324 The authors declare no competing financial interests
325
326 Correspondence and requests for materials should be addressed to Michael J. Axtell at
327 mja18@psu.edu
328
329
330

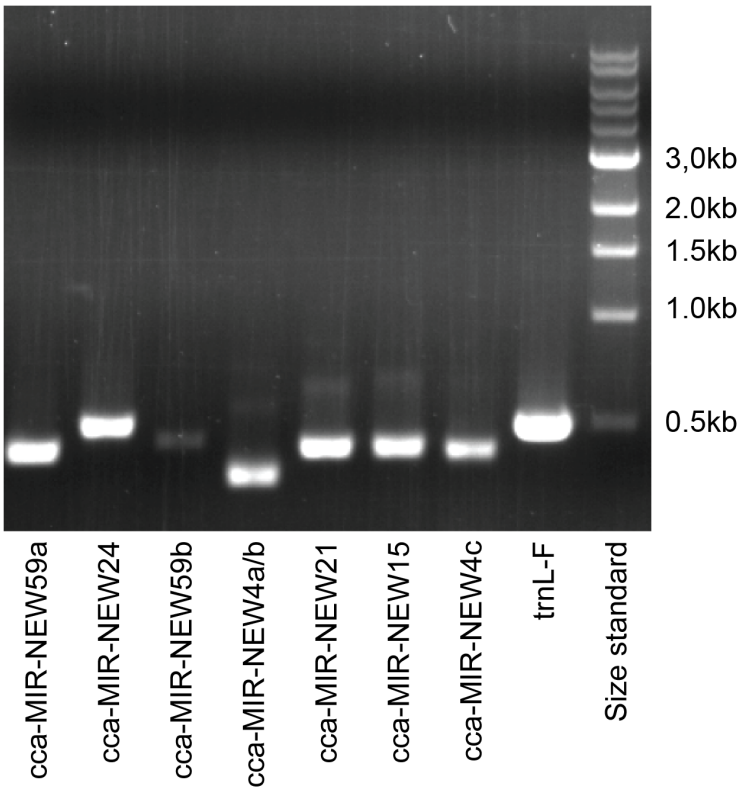


Extended Data Figure 1. *C. campestris* miRNAs cause slicing and phased siRNA production from host mRNAs. **a-d)** Small RNA-seq coverage across the indicated *A.*

335 *thaliana* transcripts are shown in blue for host stem (HS), interface (I), and parasite
336 stem (PS) samples. For display, the two biological replicates of each type were merged.
337 y-axis is in units of reads per million. Red mark and vertical lines show position of
338 complementary sites to *C. campestris* miRNAs, with the alignments shown above.
339 Fractions indicate numbers of 5'-RLM-RACE clones with 5'-ends at the indicated
340 positions; the locations in red are the predicted sites for miRNA-directed slicing
341 remnants. Barcharts show the length and polarity distribution of transcript-mapped
342 siRNAs. Radar charts show the fractions of siRNAs in each of the 21 possible phasing
343 registers; the registers highlighted in magenta are the ones predicted by the miRNA
344 target sites.
345

346

347



348

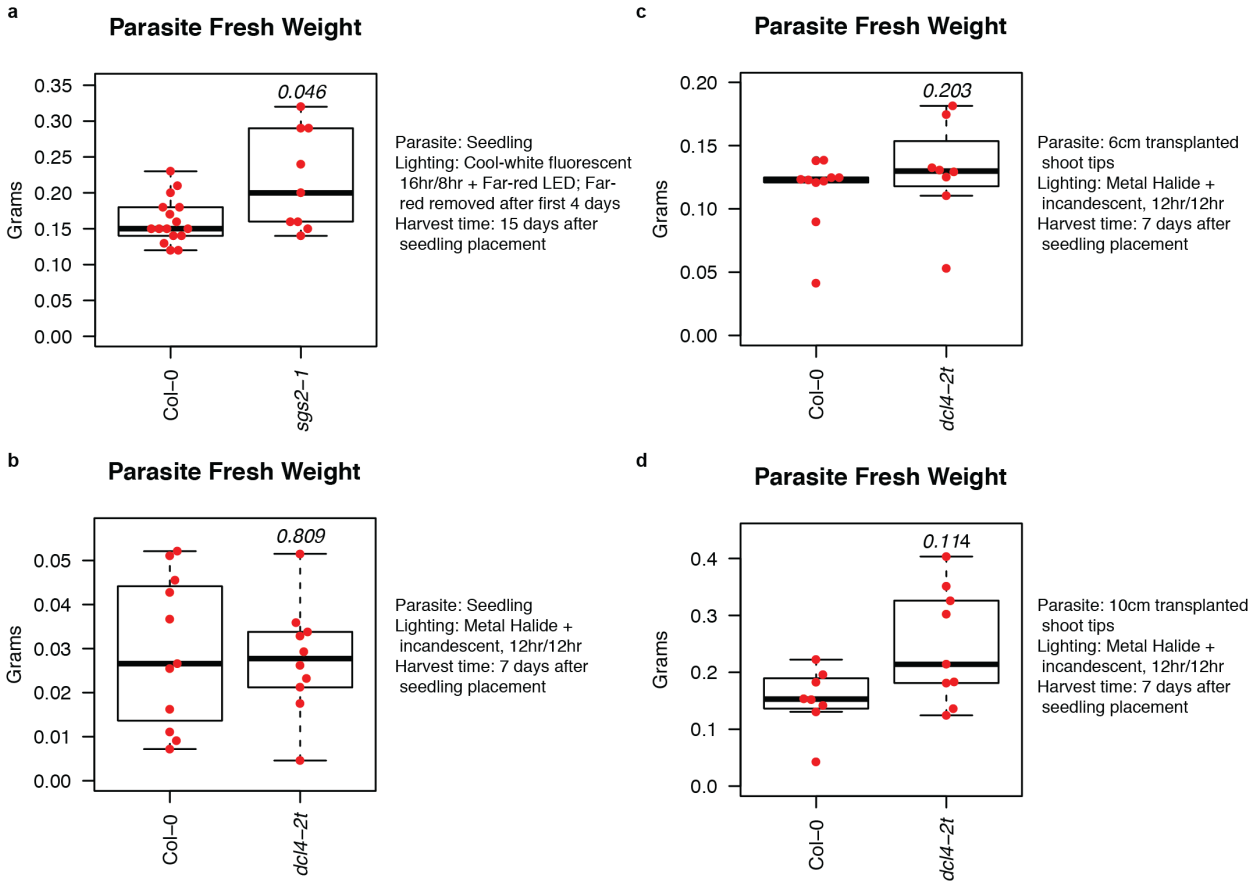
349 **Extended Data Figure 2**. PCR of *C. campestris MIRNA* loci. The template for PCR was

350 genomic DNA isolated from *C. campestris* seedlings four days after germination; the

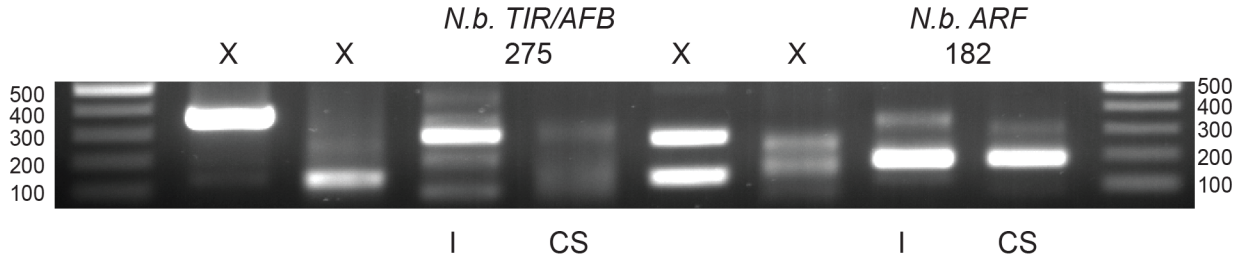
351 seedlings had never attached to nor been near a host plant, ruling out host DNA

352 contamination. trnL-F: Positive control plastid locus.

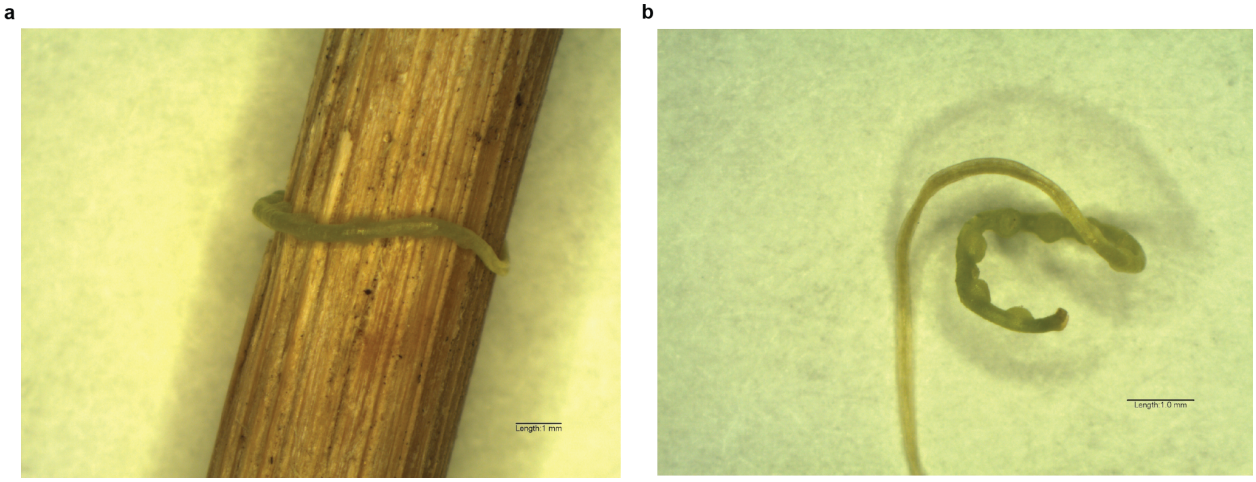
353



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.

381 **METHODS**

382 **Germplasm**

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

394

395 **Growth conditions and RNA extractions**

396 For initial experiments (small RNA-seq and RNA blots in Figure 1) *A. thaliana* (Col-0)
397 plants were grown in a growth room at 18-20°C with 12-h light per day, illuminated (200
398 μmol $m^{-2}s^{-1}$) with metal halide (400W, GE multi-vapor lamp) and spot-gro (65W,
399 Sylvania) lamps. *C. campestris* seeds were scarified in concentrated sulfuric acid for
400 45 min, followed by 5-6 rinses with distilled water and dried. *C. campestris* seeds were
401 placed in potting medium at the base of four-week-old *A. thaliana* seedlings and allowed
402 to germinate and attach to hosts. The *C. campestris* plants were allowed to grow and
403 spread on host plants for an additional three weeks to generate a supply of uniform

404 shoots for use in the experiment. Sections of *C. campestris* shoot tip (~10 cm long)
405 were placed on the floral stem of a fresh set of *A. thaliana* plants. Parasite shoots coiled
406 around the host stems and formed haustorial connections. Tissues from plants that had
407 established *C. campestris* with at least two coils around healthy host stems and clear
408 parasite growth were used in these studies. Control plants were grown under the same
409 conditions as parasitized plants, but were not exposed to *C. campestris*.
410 For the preparation of tissue-specific small RNA libraries, tissues were harvested
411 after *C. campestris* cuttings had formed active haustorial connections to the host. This
412 was evidenced by growth of the *C. campestris* shoot to a length of at least 10 cm
413 beyond the region of host attachment (7-10 d after infection). Three tissues were
414 harvested from the *A. thaliana*-*C. campestris* associations: 1) 2.5 cm of *A. thaliana* stem
415 above the region of attachment, 2) *A. thaliana* and *C. campestris* stems in the region of
416 attachment (referred to as the interface), 3) 2.5 cm of the parasite stem near the point of
417 attachment. To remove any possible cross-contamination between *A. thaliana* and *C.*
418 *campestris*, harvested regions of the parasite and host stem were taken 1 cm away
419 from the interface region and each harvested tissue was surface cleaned by immersion
420 for 5 min in 70% ethanol, the ethanol was decanted and replaced, the process was
421 repeated three times and the stems were blotted dry with a Kimwipe after the final rinse.
422 All three sections of tissue were harvested at the same time and material from 20
423 attachments were pooled for small RNA extraction. Small RNA was extracted from ~
424 100 mg of each tissue using the mirPremier microRNA Isolation Kit (Sigma-Aldrich, St.
425 Louis, MO, USA) according to the manufacturer's protocol. Small RNA was analyzed
426 using an Agilent small RNA Kit on a 2100 Bioanalyzer platform.

427 Samples used for RNA ligase-mediated 5' rapid amplification of cDNA ends (5'-
428 RLM-RACE; Figure 2C) and quantitative reverse-transcriptase polymerase chain
429 reaction (qRT-PCR; Figure 3A) analyses of *A. thaliana* targets were prepared as
430 described above with the following modifications: Col-0 *A. thaliana* hosts were cultivated
431 in a growth room with 16 hr. days, 8 hr. nights, at ~23C, under cool-white fluorescent
432 lamps, attachment of *C. campestris* cuttings was promoted by illumination with far-red
433 LED lighting for 3-5 days, and total RNA was extracted using Tri-reagent (Sigma) per
434 the manufacturer's suggestions, followed by a second sodium-acetate / ethanol
435 precipitation and wash step. Samples used for RNA blots of secondary siRNA
436 accumulation from *A. thaliana* mutants (Figure 3B) were obtained similarly, except that
437 the samples derived from the primary attachments of *C. campestris* seedlings on the
438 hosts instead of from cuttings. In these experiments, scarified *C. campestris* seedlings
439 were first germinated on moistened paper towels for three days at ~28C, then placed
440 adjacent to the host plants with their radicles submerged in a water-filled 0.125ml tube.
441 *C. campestris* pre-haustoria (Extended Data Figure 5) were obtained by
442 scarifying, germinating and placing seedlings as described above, next to bamboo
443 stakes in soil, under illumination from cool-white fluorescent lights and far-red emitting
444 LEDs. Seedlings coiled and produced pre-haustoria four days after being placed, and
445 were harvested and used for total RNA extraction (used for RNA blot in Figure 4B)
446 using Tri-reagent as described above. *Nicotiana benthamiana* was grown in a growth
447 room with 16 hr. days, 8 hr. nights, at ~23C, under cool-white fluorescent lamps. Three
448 to four week old plants served as hosts for scarified and germinated *C. campestris*
449 seedlings. Attachments were promoted by three-six days with supplementation by far-

450 red emitting LEDs. Under these conditions, *C. campestris* attached to the petioles of the
451 *N. benthamiana* hosts, not the stems. Interfaces and control petioles from un-
452 parasitized hosts were collected 7-8 days after successful attachments, and total RNA
453 (used for RNA blot in Figure 4B) recovered using Tri-reagent as described above.
454

455 **Initial sRNA-seq data processing**

456 Small RNA-seq libraries were constructed using the Illumina Tru-Seq small RNA kit per
457 the manufacturer's protocol and sequenced on an Illumina HiSeq2500 instrument. Raw
458 sRNA-seq reads were trimmed to remove 3'-adapters, and filtered for quality and
459 trimmed length >= 16 nts using cutadapt[6] version 1.9.1 with settings "-a
460 TGGAATTCTCGGGTGCCAAGG --discard-untrimmed -m 16 --max-n=0". Trimmed
461 reads that aligned with zero or one mismatch (using bowtie[7] version 1.1.2, settings "-v
462 1") to the *A. thaliana* plastid genome, the *C. gronovii* plastid genome (*C. gronovii* was
463 the closest relative to *C. campestris* that had a publically available completed plastid
464 genome assembly available), *A. thaliana* rRNAs, tRNAs, snRNAs, or snoRNAs were
465 removed. The remaining 'clean' reads were then aligned to the combined TAIR10 *A.*
466 *thaliana* reference genome and TAIR10 reference cDNAs, demanding perfect matches,
467 using bowtie[7] version 1.1.2 with settings "-v 0". Sequences that matched were initially
468 designated as host-derived, while those that didn't were initially designated as parasite-
469 derived. Species of origin assignments were then adjusted based on comparing
470 expression levels in the host stem (HS) vs. parasite stem (PS) samples. Let HS_{rpm}
471 indicate the summed reads-per-million value in both HS samples, and HS_{raw} indicate the
472 summed number of raw reads in both HS samples, and similarly for PS_{rpm} and PS_{raw}.

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

484

485 **Parasite-derived small RNA analysis**

486 Parasite-derived small RNAs between 20 and 24 nts in length were clustered according
487 to sequence similarity. Beginning with a list of parasite-derived 20-24 nt RNA
488 sequences sorted in descending order by abundance (summed across all libraries),
489 each sequence was queried against all remaining others to find less abundant
490 sequences within a Levenshtein edit distance of one or two. The most abundant
491 sequence in each resulting cluster was termed the 'head', while all other variants were
492 termed 'isos'. For computational expediency, this process was limited to 'heads' with a
493 raw read count of 20 or more. Parasite-derived sequences with total read counts of less
494 than 20 that were not found as 'isos' were discarded. This resulted in a total of 29,988

495 parasite-derived small RNA clusters. The 'head' sequences of each cluster were
496 queried against all mature miRNAs from plants from miRBase[9] version 21; hits within a
497 Levenshtein edit distance of two were recorded (Supplementary Data 1).

498 The total abundance in each cluster was calculated as the sum of read counts for
499 the 'head' and all its 'iso' sequence variants. These sums were used for differential
500 expression analysis comparing I vs. PS samples (null hypothesis: True difference no
501 more than 2-fold, Benjamini-Hochberg false discovery rate = 0.05) using the R package
502 DESeq2[10] (Supplementary Data 2). The 136 small RNA clusters that were higher in I
503 relative to PS were further analyzed to search for evidence of miRNA-like biogenesis.
504 Paired-end libraries (insert sizes of 200 bp, 340 bp, and 480 bp) and mate-pair libraries
505 (insert sizes of 3 kb and 5 kb) were constructed and sequenced on the Illumina HiSeq
506 2000 platform to obtain paired-end libraries with 100 x 100 nt reads. Data were adapter-
507 trimmed and quality-filtered using cutadapt[6] version 1.9.1 with settings "-a
508 AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A
509 AGATCGGAAGAGCGTCGTGTAGGGAAAG -m 40 -q 10 --max-n 1". The resulting
510 cleaned shotgun genomic reads were then queried to find those with exact matches to
511 the 136 parasite-derived small RNAs found to have higher accumulation in I vs. PS.
512 Genomic reads where the small RNA sequence match was within 10 nts from either the
513 5' or 3' end of the genomic read were retained and used to predict putative RNA
514 secondary structures using RNAfold[11]. Genomic reads whose predicted RNA secondary
515 structure resembled *MIRNA* hairpins were retained and aligned against one another to
516 determine local genomic assemblies. The local assemblies were consolidated based on
517 all vs. all BLAST analysis to remove redundancy (in some cases multiple parasite small

518 RNAs align to the same local assembly) and annotate families of related sequences
519 (Supplementary Data 3). The parasite-derived small RNAs were aligned against the
520 final set of local genome assemblies housing putative *MIRNA* hairpins using ShortStack
521 3.4[12] with settings "--nostitch --bowtie_cores 5 --sort_mem 4G --ranmax 20". Small
522 RNA alignment patterns relative to predicted secondary structures were visualized with
523 strucVis (https://github.com/MikeAxtell/srucVis/) version 0.2 (Supplementary Data 4,
524 Supplementary Data 5). A draft genome assembly of *C. campestris* (Westwood,
525 dePamphiis, et al., manuscript in prep.) was used to identify larger flanking regions for
526 selected *C. campestris MIRNA*s, allowing design of PCR primers (Supplementary Data
527 8) to amplify the loci directly from *C. campestris* genomic DNA (Extended Data Figure
528 2).

529

530 **Host-derived small RNA analysis**

531 The final set of host-derived reads were aligned to the TAIR10 *A. thaliana* nuclear
532 genome using ShortStack[12] version 3.4 with settings "--bowtie_cores 5 --sort_mem 4G -
533 -nostitch --mincov 1rpm", resulting in the definition of 36,918 small RNA-producing
534 genomic intervals from *A. thaliana*. Small RNA accumulation from these clusters was
535 analyzed to find differentially expressed clusters between I and HS samples (null
536 hypothesis: True difference no more than 2-fold, Benjamini-Hochberg false discovery
537 rate = 0.05) using the R package DESeq2[10] (Supplementary Data 6). Because most
538 differentially-expressed clusters from this genome-wide analysis overlapped annotated
539 transcripts (mRNAs or pri-*MIRNA*s), the host-derived small RNA reads were re-

540 analyzed by alignment to the TAIR10 representative cDNA models using ShortStack
541 version 3.4 with settings "--nohp --nostitch --bowtie_cores 5 --sort_mem 4G" and a "--
542 locifile" specifying the full-length of each transcript as a pre-defined locus. Differentially
543 expressed loci between I and HS were identified as described above (Supplementary
544 Data 6).

545

546 **RNA blots**

547 Small RNA gel blots were performed as previously described[13] with modifications. For
548 the blots in Figure 1B, small RNAs (1.8 micrograms) from each sample were separated
549 on 15% TBE-Urea Precast gels (Bio-Rad), transblotted onto the Hybond NX membrane
550 and cross-linked using 1-ethyl-3-(3-dimethylamonipropyl) carbodiimide[14]. Hybridization
551 was carried out in 5×SSC, 2×Denhardt's Solution, 20 mM sodium phosphate (pH 7.2),
552 7% SDS with 100 µg/ml salmon testes DNA (Sigma-Aldrich). Probe labeling,
553 hybridization and washing were performed as described[13]. Radioactive signals were
554 detected using Typhoon FLA 7000 (GE Healthcare). Membranes were stripped in
555 between hybridizations by washing with 1% SDS for 15 min at 80°C and exposed for at
556 least 24 h to verify complete removal of probe before re-hybridization. Sequences of
557 probes are listed below. Blots in Figures 3B and 4B were performed similarly, except
558 that 12 micrograms of total RNA were used instead. Probe sequences are listed in
559 Supplementary Data 8.

560 **5' RNA ligase-mediated rapid amplification of cDNA ends (5'-RLM-RACE)**

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

580

581 **Quantitative reverse-transcriptase PCR (qRT-PCR)**

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

594

595 ***C. campestris* growth assays**

596 *C. campestris* seedlings were scarified, pre-germinated, and placed next to hosts in
597 0.125ml water-filled tubes under cool-white fluorescent lighting supplemented with far-
598 red emitting LEDs (16hr day, 8hr night) at ~ 23C as described above. After a single
599 attachment formed (4 days), far-red light supplementation was removed to prevent
600 secondary attachments. After 18 more days of growth, entire *C. campestris* vines were
601 removed and weighed (Figures 3C-3D). Multiple additional growth trials were performed
602 specifically on the *dcl4-2t* and *sgs2-1* mutant hosts under varying conditions (Extended
603 Data Figure 4): Trial one using *sgs2*-1 used similar conditions except that harvest was

604 perform 11 days after removal of far-red light (Extended Data Figure 3A). Trial two
605 examined *dcl4-2t* and used 3cm seedlings in a 12-h per day light cycle, illuminated (200
606 µmol m-2 s -1) with metal halide (400W, GE multi-vapor lamp) and spotgro (65W,
607 Sylvania) lamps, and measured biomass seven days after attachment (Extended Data
608 Figure 3B). Trials three and four were performed on *dcl4-2t* with the same growth
609 regime and seven-day timing, except using 6cm and 10cm *C. campestris* shoot tips
610 harvested from ~ 1 month old plants grown on beets (*Beta vulgaris*) as the starting
611 material (Extended Data Figures 3C-D).

612

613 **miRNA target predictions**

614 To find probable orthologs for *Arabidopsis thaliana* genes of interest, the *A. thaliana*
615 protein sequences were used as queries for a BLASTP analysis of the 31 eudicot
616 proteomes available on Phytozome 11 (https://phytozome.jgi.doe.gov/pz/portal.html#).
617 Transcript sequences for the top 100 hits were retrieved. Two miRNA query sets were
618 prepared. The first contained mature miRNAs from interface-induced *C. campestris*
619 *MIRNA* loci. For each locus in Supplemental Table 3, the sequence with a higher total
620 read count was retained. Additionally, if the mature sequence from the other strand of
621 the hairpin had at least 100 reads and began with a 5'-U, it was also put into the
622 interface-induced query set. A second query set consisting of conserved miRNAs
623 expressed by *C. campestris* was curated by taking all small RNA 'head' sequences (see
624 above) that a) matched an annotated mature miRNA sequence from a plant species in
625 miRBase 21, and b) had a ratio of HS / (HS + PS) of <= 0.95. Targets were predicted

626 from the probable 31-species with a maximum score of 4.5 using targetfinder.pl
627 (https://github.com/MikeAxtell/TargetFinder/) version 0.1. The *BIK1, SEOR1*, *TIR/AFB*,
628 *HSFB4*, and *GAPDH* probable orthologs were searched against the interface-induced
629 queries, while the *PHB/PHV/REV* probable orthologs were searched against the
630 conserved miRNA queries.

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

635

636 **Code availability**

637 ShortStack version 3.4[12] (small RNA-seq analysis), strucVis version 0.2 (visualization of
638 predicted RNA secondary structures with overlaid small RNA-seq depths), and
639 Shuffler.pl/targetfinder.pl version 0.1 (prediction of miRNA targets controlling for false
640 discovery rate) are all freely available at https://github.com/MikeAxtell. Cutadapt version
641 1.9.1[6] is freely available at http://cutadapt.readthedocs.io/en/stable/index.html. The R
642 package DESeq2[10] is freely available at http://www.
643 bioconductor.org/packages/release/bioc/html/DESeq2.html.

644

645 **References Cited (Methods)**

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

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

649 2. Elmayan, T. *et al.* Arabidopsis mutants impaired in cosuppression. *Plant Cell* **10,**
650 1747–1758 (1998).

651 3. Xie, Z., Allen, E., Wilken, A. & Carrington, J. C. DICER-LIKE 4 functions in trans-
652 acting small interfering RNA biogenesis and vegetative phase change in
653 Arabidopsis thaliana. *Proc Natl Acad Sci USA* **102,** 12984–12989 (2005).

654 4. Froelich, D. R. *et al.* Phloem ultrastructure and pressure flow: Sieve-Element-
655 Occlusion-Related agglomerations do not affect translocation. *Plant Cell* **23,**
656 4428–4445 (2011).

657 5. Parry, G. *et al.* Complex regulation of the TIR1/AFB family of auxin receptors.
658 *Proc Natl Acad Sci USA* **106,** 22540–22545 (2009).

659 6. Martin, M. Cutadapt removes adapter sequences from high-throughput
660 sequencing reads. *EMBnet. journal* (2011).

661 7. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-
662 efficient alignment of short DNA sequences to the human genome. *Genome Biol*
663 **10,** R25 (2009).

664 8. Wang, F., Johnson, N. R., Coruh, C. & Axtell, M. J. Genome-wide analysis of
665 single non-templated nucleotides in plant endogenous siRNAs and miRNAs.
666 *Nucleic Acids Res* (2016). doi:10.1093/nar/gkw457

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

669 10. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
670 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15,** 550 (2014).

671 11. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6,** 26 (2011).
672 12. Johnson, N. R., Yeoh, J. M., Coruh, C. & Axtell, M. J. Improved Placement of
673 Multi-Mapping Small RNAs. *G3 (Bethesda)* (2016). doi:10.1534/g3.116.030452
674 13. Cho, S. H., Coruh, C. & Axtell, M. J. miR156 and miR390 regulate tasiRNA
675 accumulation and developmental timing in Physcomitrella patens. *Plant Cell* **24,**
676 4837–4849 (2012).
677 14. Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced
678 detection of small RNA. *Nat Protoc* **3,** 1077–1084 (2008).
679 15. Bombarely, A. *et al.* A draft genome sequence of Nicotiana benthamiana to
680 enhance molecular plant-microbe biology research. *Mol Plant Microbe Interact* **25,**
681 1523–1530 (2012).
682 16. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W.-R. Genome-
683 wide identification and testing of superior reference genes for transcript
684 normalization in Arabidopsis. *Plant Physiol* **139,** 5–17 (2005).
685
686

687 **SI Guide**

688

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

694

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

697

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

699

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

704

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

708

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.