Endogenous viral element-derived piRNAs are not required for production of ping-pong-dependent piRNAs from Diaphorina citri densovirus

Jared C. Nigg[1], Yen-Wen Kuo, and Bryce W. Falk

Department of Plant Pathology, University of California, 1 Shields Avenue, Davis, CA 95616

[1] Present address: Viruses and RNA Interference Unit, Institut Pasteur, UMR3569, CNRS, Paris, France

Abstract

Partial integrations of DNA and non-retroviral RNA virus genomes, termed endogenous viral elements (EVEs), are abundant in arthropod genomes and often produce PIWI-interacting RNAs (piRNAs) speculated to target cognate viruses through the ping-pong cycle, a post-transcriptional RNA silencing mechanism. Here we describe a Diaphorina citri densovirus (DcDV)-derived EVE in the genome of *Diaphorina citri*. We found that this EVE gives rise to DcDV-specific primary piRNAs and is unevenly distributed among *D. citri* populations. Unexpectedly, we found that DcDV is targeted by ping-pong-dependent viral piRNAs (vpiRNAs) in *D. citri* lacking the DcDV-derived EVE, while four naturally infecting RNA viruses of *D. citri* are not targeted by vpiRNAs. Furthermore, a recombinant Cricket paralysis virus containing a portion of the DcDV genome corresponding to the DcDV-derived EVE was not targeted by vpiRNAs during infection in *D. citri* harboring the EVE. These results represent the first report of ping-pong-dependent vpiRNAs outside of mosquitoes.

Introduction

24 The Piwi-interacting RNA (piRNA) pathway is a small RNA (sRNA) guided gene silencing
25 mechanism responsible for repressing transposable elements (TEs) in animals and emerging
26 evidence supports a role for this pathway in antiviral responses in some mosquito species and
27 mosquito-derived cell lines (1–5). In *Drosophila melanogaster*, biogenesis of primary piRNAs
28 begins with transcription of single-stranded piRNA precursor transcripts from discrete genomic
29 regions called piRNA clusters (6). TE sequences integrated into piRNA clusters can become
30 transcribed as part of a piRNA precursor transcript and these precursor transcripts are processed
31 into primary piRNAs that direct transcriptional and post-transcriptional silencing of TEs by
32 association with the Piwi-family Argonaute proteins Piwi and Aubergine, respectively (1).
33 During the ping-pong cycle, cleavage of sense TE RNA directed by Aubergine-bound antisense
34 primary piRNAs triggers production of sense secondary piRNAs from cleaved TE RNA and, in
35 turn, sense secondary piRNAs direct cleavage of antisense piRNA precursor transcripts via
36 association with another Piwi-family Argonaute protein, Argonaute 3 (Ago3), to specifically
37 amplify the response against active TEs (6). piRNAs are distinguished from other sRNAs by
38 their size (24-32 nt), nucleotide biases (uridine as the first nucleotide for primary piRNAs and
39 adenine as the tenth nucleotide for secondary piRNAs, known as the 1U and 10A bias), and
40 association with a Piwi-family Argonaute protein (1). Additionally, complementary piRNAs
41 produced by the ping-pong cycle have 5' ends separated by exactly 10 nt, known as the ping-
42 pong signature (1).
43
44 The presence of ping-pong-dependent virus-derived piRNAs (vpiRNAs) during infection with
45 several RNA viruses in *Aedes* and *Culex* mosquitoes and cell lines has been reported (2–5,7–12)
46 and reduced expression of piRNA pathway components leads to increased replication of Semliki

47 Forest virus, Bunyamwera virus, Cache Valley virus, and Rift Valley Fever virus in *Aedes*
48 *aegypti*-derived Aag2 cells (2–5). Despite the prevalence of virus-derived piRNAs and evidence
49 supporting their antiviral role, little is known regarding their biogenesis or antiviral function *in*
50 *vivo*. *Ae. aegypti* expresses an expanded group of Piwi-family Argonaute proteins containing
51 eight proteins (Piwi1-7 and Ago3) rather than the three proteins seen in *D. melanogaster* (Piwi,
52 Aub, and Ago3) (13). In Aag2 cells, Piwi5 and Ago3 are the primary Piwi-family Argonaute
53 proteins required for biogenesis of vpiRNAs upon infection with Sindbis virus (9). Importantly,
54 induction of the ping-pong cycle in mosquitoes is thought to involve recognition of exogenous
55 viral RNA, triggering biogenesis of primary vpiRNAs directly from viral RNA without the need
56 for primary piRNAs derived from endogenous loci (9,11). This is in contrast to the canonical
57 ping-pong cycle, which produces primary piRNAs only from endogenous transcripts (6). Thus,
58 mosquitoes possess a unique piRNA pathway that produces vpiRNAs by virtue of novel
59 functionality that has so far not been found in other groups of organisms. Indeed, the piRNA
60 pathway was not found to play an antiviral role in *D. melanogaster* and no evidence for an
61 antiviral piRNA pathway, or even the presence of ping-pong-dependent vpiRNAs, has been
62 reported outside of the mosquito lineage (14).
63
64 Arthropod genomes contain an abundance of sequences derived from DNA viruses and non-
65 retroviral RNA viruses (15–20) . These sequences, termed endogenous viral elements (EVEs),
66 are often enriched within piRNA clusters and are known to serve as sources of primary piRNAs
67 in a wide variety of arthropod species (15–17). These observations have led to speculation that
68 EVE-derived primary piRNAs may play an antiviral role by targeting cognate viruses in a
69 manner analogous to piRNA-mediated repression of TEs via the ping-pong cycle (15–17).

70 Importantly, such a role for EVE-derived piRNAs could theoretically be mediated by the
71 canonical ping-pong cycle and could thus expand the antiviral potential of the piRNA pathway
72 beyond mosquitoes. In support of this idea, Whitfield et al. found that a single EVE-derived
73 piRNA maps to the genome of Phasi Charoen-like virus (PCLV) in Aag2 cells, which are
74 persistently infected with PCLV, and that a vpiRNA is produced from the complementary strand
75 (16). Moreover, the authors found that knockdown of Piwi4 led to a ~2 fold increase in the
76 abundance of PCLV RNA in Aag2 cells (16). While these results are intriguing, it is difficult to
77 assess how the EVE-derived piRNA affects PCLV RNA abundance in the context of the Piwi4
78 knockdown, as only global sRNA mapping data to the entire PCLV nucleocapsid-encoding
79 segment are presented, the abundance of the EVE-derived piRNA and its potential PCLV-
80 derived ping-pong partner were not discussed, and Piwi4 knockdown has been shown to reduce
81 both siRNA and piRNA production (16,21). More recently, Tassetto et al. found that Piwi4 is
82 required for vpiRNA maturation in Aag2 cells and binds specifically to piRNAs derived from
83 EVEs (21). Furthermore, they found that insertion of EVE sequences into the 3' UTR of Sindbis
84 virus reduced viral replication in a Piwi4-dependent manner in Aag2 cells (21). These results
85 raise the possibility that EVE-derived piRNAs may target cognate viruses in these cells through
86 an interaction with Piwi4, however, detailed sRNA profiles and more thorough examinations of
87 the interplay between virus-specific piRNAs derived from endogenous and exogenous sources
88 are needed to establish the link between EVE-derived piRNAs and an antiviral response.
89
90 We previously identified the EVEs present within the genomes of 48 arthropod species (15).
91 Among the EVEs that we estimated to be similar enough to their corresponding exogenous
92 viruses at the nucleotide level such that piRNAs derived from them could potentially target viral

93 RNA during infection, Diaphorina citri densovirus (DcDV) and a DcDV-derived EVE located
94 within a piRNA cluster on *Diaphorina citri* genomic scaffold 2850 stood out as an EVE-virus
95 pair sharing exceptionally high nucleotide identity over a relatively long region (15). *D. citri*,
96 also known as the Asian citrus psyllid, is a hemipteran insect pest and serves as a vector of
97 *Candidatus* Liberibacter species that cause Huanglongbing, or citrus greening disease, in all
98 types of commercially cultivated citrus throughout the world (22). This disease represents the
99 greatest threat to the citrus industry worldwide and has devastated many citrus growing regions,
100 such as in the US state of Florida, where citrus production has decreased by 74% since the first
101 report of citrus greening disease in the state in 2005 (23). Control strategies for the disease are
102 limited and primarily involve removal of infected trees and management of *D. citri* populations
103 with chemical insecticides (24). The continued spread of citrus greening disease and the
104 development of insecticide resistance in field populations of *D. citri* highlight the need for new
105 management approaches (24). Biological control strategies relying on infectious agents have
106 proven to be effective methods for controlling insect pests and the use of parasites and viruses
107 have been proposed as possible vector control approaches for *D. citri* (25–28). Current
108 understanding of immune processes in *D. citri* is limited and studies in this field are needed to
109 facilitate the development of effective biological control strategies. In particular, antiviral
110 mechanisms in *D. citri* have not been experimentally evaluated. The presence of a highly similar
111 virus-EVE pair, a lack of novel Piwi-family of Argonaute proteins like those observed in
112 mosquitoes, and the need to study antiviral mechanisms in *D. citri* make this species an excellent
113 and relevant model to study the interactions between EVE-derived piRNAs and corresponding
114 viruses.
115

116 Here we further characterized the DcDV-derived EVE present within the *D. citri* genome. We
117 found that piRNAs are produced from this EVE in multiple tissue types and that this EVE is
118 conserved among some geographically distinct populations of *D. citri*, but is absent from other
119 populations. By analyzing sRNA profiles in *D. citri* insects infected with DcDV, we found that
120 this virus is targeted by ping-pong-dependent vpiRNAs independent of DcDV-derived EVEs and
121 endogenous DcDV-specific piRNAs. Additionally, analysis of sRNA profiles during infection
122 with other viruses suggests that RNA viruses are not a target of vpiRNAs in *D. citri*.
123
124
125 **Results**
126 *A DcDV-derived EVE with high nucleotide identity to DcDV is conserved in some populations*
127 *of D. citri, but absent in others*
128 We previously identified a 621 bp DcDV-derived EVE located within a piRNA cluster on *D.*
129 *citri* genomic scaffold 2850 by comparing deduced virus protein sequences to deduced *D. citri*
130 genome-encoded protein sequences using BLASTx (15). This EVE was 85.5% identical to the
131 corresponding region of the DcDV genome at the deduced amino acid level and resided just
132 downstream from another 624 bp EVE derived from a different portion of the DcDV genome. To
133 characterize these EVEs at the nucleotide level, we aligned the nucleotide sequence of the DcDV
134 genome to the region of the *D. citri* genome harboring these DcDV-derived EVEs. We found
135 that this region of the *D. citri* genome contains sequences corresponding to the DcDV inverted
136 terminal repeats (ITRs) and to two regions of the non-structural protein (NS) gene cassette
137 together spanning 1,396 bp. For simplicity we refer to these regions as endogenous ITR (EITR)
138 and endogenous NS (ENS) (Fig. 1a). Notably, the portion of ENS spanning nucleotides 7,523 to

139 8,175 within genomic scaffold 2850 shares 86% nucleotide identity with the corresponding
140 region of the DcDV genome.
141
142 Because our EVE prediction shown in Fig. 1a was based only on BLAST results and a previous
143 *D. citri* genome assembly, we wanted to confirm the presence, sequence, and organization of the
144 DcDV-derived EVEs. To confirm the presence of these EVEs, we designed PCR primers to
145 amplify this region from *D. citri* genomic DNA (Fig. 1a). We obtained PCR products of
146 unexpected size (based on the available sequence of scaffold 2850 and the organization depicted
147 in Fig. 1a) for both primer sets (Fig. 1b). Sequencing of these PCR products revealed that the
148 portion of ENS spanning nucleotides 5,248 to 5,871 within genomic scaffold 2850 was not
149 present and the distance between the ITR-like fragments was 3,885 bp rather than 11,038 bp.
150 Additionally, EITR was 210 bp instead of 114 bp. The portion of ENS corresponding to the
151 region between DcDV nucleotides 263-916 was present and closely matched the sequence
152 observed by BLAST. Figure 1c shows the correct organization of this region of the DcDV
153 genome, as confirmed by sequencing of the PCR products shown in Fig. 1b. This organization
154 and sequence is supported by a more recent draft of the *D. citri* genome assembled using PacBio
155 long reads (Diaci2.0,
156 ftp://ftp.citrusgreening.org/genomes/Diaphorina_citri/assembly/DIACI_v2.0/).
157
158 Our PCR reactions were performed using DNA extracted from *D. citri* insects collected from a
159 colony located at the University of California Davis Contained Research Facility (CRF) that was
160 originally started using insects collected in the US state of California (this colony is designated
161 CRF-CA). The *D. citri* reference genome used for BLAST searches and nucleotide alignments

162 was produced by sequencing DNA from insects collected in the US state of Florida (29).
163 Haplotype network analysis of global *D. citri* populations based on the mitochondrial
164 cytochrome oxidase subunit I gene indicates the existence of 44 *D. citri* haplotypes belonging to
165 two distinct lineages, denoted lineages A and B (30–32). The invasion history of *D. citri* out of
166 Southern Asia has resulted in segregation of the two lineages, such that only lineage B is found
167 in North America, while lineage A predominates in Southeast Asia, Africa, and South America
168 (30). Besides CRF-CA, three other *D. citri* colonies are maintained at the CRF and these were
169 started using *D. citri* insects collected in Taiwan, Uruguay, and the US state of Hawaii
170 (designated CRF-TW, CRF-Uru, and CRF-HI, respectively). To determine whether ENS is
171 conserved among geographically distinct *D. citri* populations, we performed PCR using primers
172 flanking ENS and DNA extracted from insects collected from CRF-CA, CRF-TW, CRF-Uru,
173 and CRF-HI. We also included DNA extracted from field collected *D. citri* insects from
174 Pakistan, Brazil, and the US states of Arizona and Florida. We obtained nearly identical PCR
175 products from insects from CRF-CA, CRF-HI, Pakistan, Arizona, and Florida, but no PCR
176 products from insects from CRF-TW, CRF-Uru, or Brazil (Fig. 2a & S1). Based on the known
177 distribution of the two *D. citri* lineages, these results suggest that ENS is absent in *D. citri* from
178 lineage A.
179
180 Sequences derived from some insect infecting viruses are maintained as circular episomal
181 molecules that produce virus-specific sRNAs (12,33). Given the heterogenous distribution of
182 ENS among *D. citri* populations, we examined whether ENS is maintained episomally rather
183 than representing a true EVE. When we performed a Southern blot with DNA extracted from
184 CRF-CA *D. citri* using an RNA probe based on the sequence of ENS we observed a single DNA

185 species at high molecular weight (>12,000 bp) from undigested DNA and two species of lower
186 molecular weight in DNA digested with PstI and HindIII, indicating that ENS resides in high
187 molecular weight genomic DNA rather than low molecular weight episomal DNA (Fig. 2b). The
188 observation of two DNA species in the digested sample does not represent a second DcDV-like
189 EVE because ENS contains a PstI clevage site. To confirm that ENS is a true EVE, we treated *D.*
190 *citri* genomic DNA with an exonuclease to remove genomic DNA, but not circular episomal
191 DNA, and performed PCR with ENS-specific primers using exonuclease-treated DNA as
192 template. We obtained a PCR product from non-digested DNA, but not from digested DNA (Fig.
193 2c). Together, these results indicate the ENS is indeed integrated into the *D. citri* genome.
194 Finally, strand specific RT-PCR indicates that ENS is bidirectionally transcribed, although the
195 majority of transcripts are antisense to the corresponding DcDV transcript (Fig. 2d).
196
197
198 *ENS and EITR give rise to DcDV-specific primary piRNAs*
199 To evaluate whether ENS and EITR give rise to virus-specific primary piRNAs as described for
200 other EVEs, we mapped sRNAs from CRF-CA *D. citri* not infected with DcDV to the DcDV
201 genome. We found that sRNAs purified from CRF-CA *D. citri* mapped to EITR and to the
202 negative strand within the portion of the DcDV genome corresponding to ENS (i.e. antisense to
203 DcDV transcripts), but not to other regions (Fig. 3a). The size of these sRNAs was characteristic
204 of piRNAs and they possessed a strong 1U bias (Fig. 3b & c). Similar results were obtained for
205 other *D. citri* populations that are not infected with DcDV and for which sRNA datasets are
206 available (Fig. S2a & b). As expected based on the lack of ENS in *D. citri* insects collected in
207 Brazil or from CRF-Uru, DcDV-specific piRNAs were not observed in these insects (Fig. S2c &

208 d). We note that the DcDV ITRs are comprised of a 210 nt hairpin present on both ends of the
209 DcDV genome. Thus, it is not possible to determine whether the sRNAs mapping to EITR are
210 specific to the positive or the negative strand. For this reason, because of the relatively small size
211 of EITR compared to ENS, and because EITR does not correspond to a transcribed region of the
212 DcDV genome (34), we chose to focus our analysis on piRNAs derived from ENS.
213
214 The expression of piRNAs is known to display tissue specificity and piRNA expression patterns
215 could have important consequences for the ability of EVE-derived piRNAs to target cognate
216 viruses. To determine the tissue specificity of piRNAs produced from ENS and to determine the
217 tissues in which the ping-pong cycle is active in *D. citri*, we sequenced sRNAs purified from
218 dissected *D. citri* guts, heads, ovaries, testis, and hemolymph using insects collected from CRF-
219 CA. We found that while DcDV-specific piRNAs derived from ENS were present in all tissues
220 analyzed, their expression was significantly higher in *D. citri* guts than in any other tissue (Fig.
221 3d). The ping-pong cycle is restricted to germline tissues in *D. melanogaster* (1), however a
222 comprehensive analysis of somatic sRNAs mapping to TEs genome wide in 20 arthropod species
223 in combination with ancestral state reconstruction indicates that somatic ping-pong amplification
224 is widespread throughout arthropods despite having been independently lost in some species,
225 including *D. melanogaster* (35). To determine the tissues in which the ping-pong cycle is active
226 in *D. citri* we mapped sRNAs from CRF-CA *D. citri* guts, heads, hemolymph, ovaries, and testis
227 to all TEs identified within the *D. citri* genome and analyzed the mapped 27-32 nt sRNAs for the
228 presence of ping-pong signatures. We found evidence for ping-pong amplification of TE-derived
229 piRNAs in all tissues examined (Fig. S3).
230

231

232 *DcDV is targeted by ping-pong-dependent piRNAs in the absence of a DcDV-derived EVE*

233 We previously found that *D. citri* from CRF-CA are resistant to infection with DcDV (34). In
234 contrast, the virus is maintained as a persistent, maternally transmitted infection in *D. citri* from
235 CRF-TW (34). Thus, to understand the sRNA-based response to DcDV infection in *D. citri*, we
236 sequenced sRNAs from DcDV-infected *D. citri* from CRF-TW. These results revealed a major
237 population of 21 nt sRNAs, indicative of an siRNA-based response (Fig. 4a). Unexpectedly, we
238 also observed a smaller peak within the piRNA size range centered at 29 nt and we obtained
239 99.5% coverage of transcribed regions of the DcDV genome by mapping only 27-32 nt sRNAs
240 (Fig. 4a & b). We found that complementary 27-32 nt sRNAs mapping to opposite strands
241 throughout the DcDV genome possessed 5' ends separated by 10 nt more often than expected by
242 chance, an indication of ping-pong amplification (Z-score = 4.05±0.08)(Fig. 4c). Moreover, we
243 detected the 1U and 10A biases typical of ping-pong amplification in 27-32 nt sRNAs mapping
244 antisense and sense to the canonical DcDV transcripts, respectively (Fig. 4d & e).

245

246 We found that ENS is not present in the genome of CRF-TW *D. citri* (Fig. 2a), thus our
247 observation of ping-pong-dependent DcDV-derived piRNAs in these insects suggests that DcDV
248 is targeted by piRNAs independent of ENS. However, because this result was obtained by
249 performing PCR using primers flanking ENS, it is possible that ENS is present in CRF-TW *D.*
250 *citri,* but resides in different genomic context than seen in other *D. citri* populations. ENS is not
251 identical to the corresponding region of the DcDV genome at the nucleotide level (Fig. 1c).
252 Thus, some piRNAs derived from ENS do not perfectly map to DcDV, providing a means with
253 which to distinguish some ENS-derived piRNAs from DcDV-derived piRNAs. To rule out the

254 possibility that ENS is present in the genome of CRF-TW *D. citri*, we mapped 27-32 nt sRNAs
255 from CRF-TW *D. citri* to the DcDV genome without allowing any mismatches. When the
256 unmapped reads from this analysis were mapped to ENS without allowing any mismatches, no
257 reads mapped. In contrast, we obtained an average of 55.8% coverage of the ENS sequence when
258 the same analysis was performed using sRNAs from CRF-CA *D. citri* (data not shown). This
259 result indicates that ENS is indeed not present in the genome of CRF-TW *D. citri* and that the
260 targeting of DcDV by ping-pong-dependent vpiRNAs in these insects is independent of ENS-
261 derived piRNAs.
262
263 We cannot exclude the possibility that the genome of CRF-TW *D. citri* harbors a different
264 piRNA-producing DcDV-derived EVE. Because DcDV is maternally transmitted to 100% of the
265 progeny of CRF-TW females (34), it is not possible to determine the repertoire of endogenous
266 piRNAs in these insects in the absence of DcDV infection. Thus, we analyzed the sRNAs present
267 in *D. citri* from CRF-Uru, as these insects lack ENS and are not infected with DcDV (Fig. 2a).
268 We found that no DcDV-specific piRNAs are produced in uninfected CRF-Uru *D. citri*,
269 indicating that these insects do not harbor a piRNA-producing DcDV-derived EVE (Fig. S2d).
270 The absence of DcDV-specific piRNAs was not due to a lack of piRNAs in general, as we
271 detected abundant ping-pong-dependent TE-derived piRNAs in these insects (Fig. S4). In
272 contrast to *D. citri* from CRF-CA which are resistant to DcDV infection, we found that *D. citri*
273 from CRF-Uru were susceptible to DcDV infection by both oral acquisition and intrathoracic
274 injection (Fig. S5a & b)(34). Moreover, the progeny of CRF-Uru *D. citri* infected with DcDV
275 were also infected with the virus (Fig. S5c). Analysis of sRNAs from the progeny of CRF-Uru
276 *D. citri* that had been infected with DcDV by intrathoracic injection revealed the presence of

277 ping-pong-dependent DcDV-derived piRNAs (ping-pong Z-score = 2.55) (Fig. S6). Together,
278 these results demonstrate that DcDV is targeted by ping-pong-dependent vpiRNAs in two
279 populations of *D. citri* independent of EVE-derived piRNAs.
280
281
282 *Other viruses of D. citri are not targeted by piRNAs*
283 Besides DcDV, there are five other viruses known to infect *D. citri*: Diaphorina citri reovirus
284 (double-stranded RNA), Diaphorina citri picrona-like virus (plus strand single-stranded RNA; +
285 ssRNA), Diaphorina citri bunyavirus (minus strand single-stranded RNA), Diaphorina citri-
286 associated c virus (+ ssRNA), and Diaphorina citri flavi-like virus (+ ssRNA)(28,36–38). To
287 determine whether piRNAs are part of a general response to viruses in *D. citri*, we mapped
288 sRNAs from publicly available sRNA libraries known to be derived from insects infected with
289 one or more of each of these viruses to the corresponding viral genomes (with the exception of
290 Diaphorina citri-associated c virus for which no such sRNA library exists) (28,36). As expected,
291 we detected a prominent peak at 21 nt for each virus, indicating a siRNA-based response (Fig.
292 S7). While virus-derived sRNAs within the piRNA size range were present during infection with
293 all four viruses, there were no peaks above background levels within the piRNA size range and
294 27-32 nt reads lacked signatures typical of primary or ping-pong-dependent piRNAs (Fig. S7).
295 These results indicate that viruses in general are not targeted by piRNAs in *D. citri* and suggest
296 that the targeting of DcDV by the piRNA pathway in *D. citri* may be due to differences in the
297 infection cycles between RNA and DNA viruses.
298
299

300 *A recombinant CrPV-based reporter virus carrying DcDV sequence is not targeted by piRNAs in*
301 D. citri *that produce primary piRNAs from a DcDV-derived EVE*
302 As has been observed for RNA viruses in mosquitoes, our results suggest that DcDV is targeted
303 by ping-pong-dependent vpiRNAs due to the *de novo* production of vpiRNAs from exogenous
304 DcDV RNA. If EVE-derived piRNAs were to prime the ping-pong cycle in this context, it would
305 be difficult to distinguish priming driven by EVE-derived piRNAs from priming driven by virus-
306 derived vpiRNAs. Because our results suggest that RNA viruses are not targeted by vpiRNAs in
307 *D. citri*, we sought to construct a recombinant RNA virus harboring DcDV-derived EVE
308 sequence in order to study potential priming of the ping-pong cycle by EVE-derived piRNAs
309 without the background of vpiRNAs produced directly from viral RNA. For this purpose, we
310 used Cricket paralysis virus (CrPV), a dicistrovirus with a + ssRNA genome that was originally
311 isolated from field crickets (39). Due to the broad experimental host range of CrPV, including *D.*
312 *citri* (E Matsumura, personal communication, October 2018), and the availability of an infectious
313 clone, CrPV is often used to study antiviral mechanisms in insects (40,41).
314
315 To directly test the hypothesis that EVE-derived primary piRNAs can prime ping-pong
316 amplification during infection with viruses sharing complementary sequence, we inserted 57 nt
317 from the DcDV genome (from a region corresponding to ENS) into the CrPV genome between
318 the 1A and 2B coding sequences flanked by duplicate copies of the 1A cleavage site (Fig. 5a,
319 Fig. S8). The recombinant DcDV sequence was inserted into the CrPV genome such that ENS-
320 derived primary piRNAs mapping antisense to DcDV transcripts would map to the negative
321 strand of the recombinant CrPV-DcDV genome. Thus, in both cases, ENS-derived primary
322 piRNAs are complementary to the viral coding sequence and in analogy to the mechanisms of

323 ping-pong amplification in the context of TEs, should be theoretically capable of directing
324 cleavage of viral mRNAs. This recombinant virus is designated CrPV-DcDV. While the 57 nt
325 inserted region is relatively small in the context of the ~9 kb CrPV genome, piRNA target sites
326 corresponding to single piRNAs are known to be sufficient for induction of ping-pong
327 amplification and a single piRNA produced from a PCLV-derived EVE was proposed to mediate
328 a piRNA-based antiviral response in Aag2 cells (16,42–44). Additional analysis of the sRNA
329 mapping data from DcDV-uninfected CRF-CA *D. citri* shown in Figure 3a indicates that this 57
330 nt region gives rise to an average of 171.7 27-32 nt sRNAs made up of an average of 30.7 unique
331 sequences in the three sRNA libraries analyzed (data not shown).
332
333 We found that the virions produced following transfection of S2 cells with *in vitro* transcribed
334 CrPV-DcDV RNA were indistinguishable from those produced following transfection with wild-
335 type CrPV RNA (Fig. 5b & c) and RT-PCR of RNA purified from CrPV-DcDV virions indicated
336 that the recombinant sequence was retained (Fig. 5d). Finally, we found that infection of S2 cells
337 with either wild-type CrPV or CrPV-DcDV virions resulted in the cytopathic effects
338 characteristic of CrPV infection (Fig. 5e & f).
339
340 To determine whether ENS-derived piRNAs mediate a response to CrPV-DcDV in *D. citri*, we
341 inoculated CRF-CA *D. citri* with 1000 tissue culture infectious dose 50% ($TCID_{50}$) units of wild-
342 type CrPV or CrPV-DcDV per insect by intrathoracic injection. We observed a significant
343 difference in viral RNA levels for wild-type CrPV and CrPV-DcDV three days post injection
344 (Fig. 6a). By five days post injection, CrPV-DcDV RNA levels remained lower than those of
345 wild-type CrPV, but the difference was not statistically significant (Fig. 6a). RT-PCR results

346 indicated the recombinant DcDV sequence was retained throughout the course of infection with
347 CrPV-DcDV (Fig. S9). To determine if CrPV-DcDV was targeted by piRNAs, we sequenced
348 sRNAs from insects collected five days post infection. During infection with either virus, we
349 observed an abundance of 21 nt sRNAs of both polarities and these 21 nt sRNAs mapped to the
350 entire sequence of both viral genomes, indicating the presence of an siRNA-based response (Fig.
351 6b & c, Fig. S10 a & b). The recombinant DcDV sequence within the CrPV-DcDV genome was
352 a target of this siRNA-based response, as we observed abundant 21 nt sRNAs derived from this
353 portion of the recombinant virus genome in *D. citri* infected with CrPV-DcDV (Fig. 6d). When
354 sRNAs from *D. citri* infected with wild-type CrPV were mapped to the same DcDV-derived
355 sequence, a peak within the piRNA-size range was observed due to the production of piRNAs
356 from ENS, but no 21 nt peak was seen (Fig. 6e)
357
358 To evaluate whether CrPV-DcDV was targeted by ping-pong-dependent piRNAs, we analyzed
359 27-32 nt sRNAs separately depending on whether they mapped to the recombinant DcDV
360 sequence within the CrPV-DcDV genome or to the rest of the CrPV genome. We detected ENS-
361 derived antisense piRNAs mapping to the recombinant DcDV sequence in *D. citri* infected with
362 either wild-type CrPV or CrPV-DcDV (Fig. 6d-f), however, there was no evidence for the
363 production of ping-pong-dependent secondary piRNAs from this region during infection with
364 CrPV-DcDV (ping-pong Z-score = 0.28±0.18)(Fig. 6g & h). Ping-pong signatures were also not
365 observed for the non-recombinant portion of the CrPV-DcDV genome or for wild-type CrPV
366 (Fig. S10 c & d). These results indicate that CrPV-DcDV is not targeted by piRNAs during
367 infection initiated by intrathoracic injection in *D. citri* despite the presence of ENS-derived
368 piRNAs identical to the recombinant DcDV sequence.

369

370 Because the expression of ENS-derived piRNAs is highest in the gut (Fig. 3d) we wanted to
371 examine whether a piRNA-based response to CrPV-DcDV could be detected during infection
372 initiated by oral acquisition. Thus, we orally inoculated *D. citri* insects from CRF-CA with wild-
373 type CrPV or CrPV-DcDV by allowing the insects to feed for 96 hours on an artificial diet
374 solution containing 10^9 $TCID_{50}$ units/mL of wild-type CrPV or CrPV-DcDV. Following the 96
375 hour feeding period, the insects were transferred to *Citrus macrophylla* plants and viral RNA
376 levels were evaluated every three days by RT-qPCR. Viral RNA levels in the insects were nearly
377 identical for wild-type CrPV and CrPV-DcDV immediately following their removal from the
378 virus-containing artificial diet and average viral RNA levels increased in the days following
379 transfer of the insects to plants, although there was a substantial amount of variation between
380 biological replicates (Fig. 7a). As with infections initiated by intrathracic injection, RT-PCR
381 results indicated the recombinant DcDV sequence was retained throughout the course of oral
382 infection with CrPV-DcDV (Fig. S9b). These results suggest that while wild-type CrPV and
383 CrPV-DcDV are both infectious in *D. citri* following oral acquisition, individual insects
384 displayed a variety of infection outcomes. Similar results have previously been reported for
385 orally acquired infections with other RNA viruses in insects (45). We sequenced sRNAs from
386 insects collected nine days after transfer of the insects from the virus-containing artificial diet to
387 plants. We analyzed this data as described above for infections initiated by injection and found
388 that while both wild-type CrPV and CrPV-DcDV were targeted by siRNAs, neither virus was
389 targeted by ping-pong-dependent vpiRNAs during oral infection (ping-pong Z-score for 27-32 nt
390 sRNAs mapping to the recombinant DcDV sequence during CrPV-DcDV infection = 1.04 ±
391 0.67)(Fig. 7b-h). Together, our results indicate that despite the presence of endogenous primary

392 piRNAs complementary to a portion of the recombinant viral genome, CrPV-DcDV is not a
393 target of the piRNA pathway in *D. citri* following infection initiated either orally or by
394 intrathoracic injection.
395
396
397 **Discussion**
398 Primary vpiRNAs were first reported for American nodavirus, Drosophila C virus, Drosophila
399 tetravirus, Nora virus, Drosophila X virus, and Drosophila birnavirus in a *D. melanogaster*
400 ovarian somatic sheet (OSS) cell line (46). However, these conclusions were based only the size
401 of the sRNAs and nucleotide biases for all virus-specific sRNAs pooled into a single analysis. A
402 subsequent study using whole flies found no evidence for the production of primary or secondary
403 vpiRNAs during infection with Drosophila C virus, Drosophila X virus, Sindbis virus, Nora
404 virus, Drosophila A virus, Drosophila melanogaster sigma virus, invertebrate iridescent virus 6,
405 or Flock house virus (14). To date, ping-pong dependent vpiRNAs have only been found in a
406 small number of mosquito species and cells lines that express an expanded group of Piwi-family
407 Argonaute proteins.
408
409 Aedes albopictus densovirus 1 (AalDV-1) was recently shown to be a target of both siRNAs and
410 ping-pong-dependent vpiRNAs during persistent infection in *Ae. aegypti*-derived Aag2 cells,
411 representing the first report of vpiRNAs targeting a DNA virus and the first report of siRNAs
412 targeting a ssDNA virus in insects (47). Densoviruses replicate exclusively in the nucleus (48)
413 and interestingly, analysis of nuclear and cytoplasmic sRNA fractions of AalDV-1-infected Aag2
414 cells revealed that while AalDV-1-derived primary piRNAs are found in both the nucleus and
415 cytoplasm, ping-pong-dependent piRNAs targeting AalDV-1 are found only in the cytoplasm

416 (47). Ago3 and Piwi5 are the sole Piwi family proteins required for the production of vpiRNAs
417 during infection of Aag2 cells with Sindbis virus (9). As these proteins are expressed only in the
418 cytoplasm, the observation that AalDV-1-derived primary piRNAs are present in the nucleus
419 during infection of Aag2 cells suggests that a different pathway may generate primary vpiRNAs
420 from viral RNA in the nucleus (47). Such a pathway would be distinct from the cytoplasmic
421 Ago3/Piwi5 pathway that is responsible for production of vpiRNAs from Sindbis virus and
422 might rely on a piRNA biogenesis factor expressed in the nucleus, such as zucchini endonuclease
423 (47).
424
425 How viruses are marked as substrates for piRNA biogenesis remains unclear. In Aag2 cells, viral
426 and TE RNAs are processed into piRNAs by distinct sets of Piwi-family Argonaute proteins, but
427 it is unknown how viral RNA is recognized by these proteins (9). Moreover, as discussed above,
428 the presence of AalDV-1-derived primary piRNAs in the nucleus of Aag2 cells suggests that a
429 different pathway for the production of primary vpiRNAs may operate in the nucleus (47).
430 Several previous reports have speculated that primary piRNAs derived from EVEs may facilitate
431 the targeting of cognate viruses in a manner analogous to the targeting of TEs by the canonical
432 ping-pong cycle (15–17). Theoretically, such a targeting mechanism could facilitate the
433 production of vpiRNAs in species that do not express the expanded group of Piwi-family
434 Argonaute proteins expressed by some mosquito species and mosquito-derived cell lines.
435
436 The *D. citri* genome contains a DcDV-derived EVE (denoted ENS) with 86% nucleotide identity
437 to the corresponding portion of the DcDV genome and here we found that ENS serves as a
438 source for DcDV-specific primary piRNAs in geographically distinct *D. citri* populations.

439 However, we found that ENS is not present in all *D. citri* populations and that *D. citri* insects
440 lacking ENS do not produce endogenous DcDV-specific piRNAs. While more comprehensive
441 analyses are needed, comparing our results with the known geographic distribution of *D. citri*
442 lineages suggests that ENS may be present in lineage B *D. citri* and absent in lineage A *D. citri.*
443 We previously established a colony of *D. citri* insects harboring DcDV as a persistent infection
444 (CRF-TW)(34) and here we found that insects from this colony do not possess ENS. Sequencing
445 of sRNAs from *D. citri* insects from CRF-TW revealed the production of DcDV-derived ping-
446 pong-dependent vpiRNAs. Similar results were obtained by sequencing sRNAs from DcDV-
447 infected CRF-Uru *D. citri*, another *D. citri* population which does produce any endogenous
448 DcDV-specific piRNAs. These were surprising results, as the *D. citri* repertoire of Piwi-family
449 Argonaute proteins is limited to homologs of those expressed in *D. melanogaster,* which have so
450 far not been associated with the production of vpiRNAs (29,49). The production of vpiRNAs in
451 *D. citri* seems to display some virus specificity as vpiRNAs were not seen during infection with
452 Diaphorina citri reovirus, Diaphorina citri picrona-like virus, Diaphorina citri bunyavirus,
453 Diaphorina citri flavi-like virus, or CrPV. Notably, these viruses are all RNA viruses that
454 replicate in the cytoplasm, but DcDV is a DNA virus that replicates in the nucleus. While the
455 production of ping-pong-dependent DcDV-derived vpiRNAs is noteworthy, we cannot determine
456 whether this process plays an antiviral role based on the present data.
457
458 Densoviruses are known to replicate exclusively in the nucleus (48). Given that vpiRNAs are not
459 produced during infection of *D. citri* with positive strand RNA viruses, a negative strand RNA
460 virus, or a dsRNA virus, future analysis should determine whether DcDV RNA is processed into
461 primary vpiRNAs in the nucleus as was suggested for AalDV-1 (47). Interestingly, abundant

462 virus-derived sRNAs within the piRNA size range are also observed during infection of *Myzus*
463 *persicae* with Myzus persicae densovirus and during infection of *Culex pipiens molestus* with
464 Mosquito densovirus, however, these sRNAs were not analyzed for the presence of piRNA
465 signatures (50,51).
466
467 In addition to ping-pong dependent vpiRNAs, our results show that DcDV is also targeted by 21
468 nt siRNAs. RNA interference mediated by siRNAs is the primary antiviral defense mechanism
469 against RNA viruses in insects and requires cleavage of dsRNA substrates by the RNase III
470 enzyme dicer (52). RNA interference mediated by siRNAs is also known to target dsDNA
471 viruses and a ssDNA virus was recently shown to be targeted by siRNAs in Aag2 cells (47,53–
472 56). Previous reports have suggested that overlapping viral transcripts represent the dsRNA
473 substrate that becomes processed into siRNAs during infection with DNA viruses (47,53–55).
474 Indeed, we previously found that transcriptional readthrough during transcription of the
475 ambisense DcDV genome leads to the production of nearly genome length complementary
476 transcripts (34). These transcripts may anneal to form the dsRNA required for production of
477 siRNAs and may constitute the sense and antisense precursors required for the ping-pong cycle.
478
479 Although our results indicate that EVE-derived piRNAs are not required for the production of
480 ping-pong-dependent DcDV-derived vpiRNAs, we cannot exclude the possibility that EVE-
481 derived piRNAs could serve as an additional pool of piRNAs that could also contribute to
482 priming of the ping-pong cycle. Thus, to investigate whether EVE-derived piRNAs can prime
483 ping-pong amplification during infection with a virus sharing complementary sequence in *D.*
484 *citri*, we constructed CrPV-DcDV, a recombinant CrPV harboring 57 nt derived from the DcDV

485 genome and corresponding to ENS. Despite the presence of endogenous primary piRNAs
486 derived from ENS perfectly mapping to the recombinant region of CrPV-DcDV, we did not
487 observe induction of the ping-pong cycle during infection with this virus initiated either by oral
488 acquisition or intrathoracic injection of virions. The lack of ping-pong cycle activation during
489 infection with CrPV-DcDV is unlikely to be due to an absence of ping-pong activity in the
490 tissues infected by CrPV-DcDV, as we detected abundant TE-derived ping-pong-dependent
491 piRNAs in dissected *D. citri* guts, heads, hemolymph, ovaries, and testis. These results indicate
492 that the presence of EVE-derived primary piRNAs sharing perfect nucleotide identity with an
493 exogenous virus is not in itself sufficient to mark that virus as a target for ping-pong
494 amplification in *D. citri*. We note that our approach has some limitations. Host-responses to virus
495 infection are shaped by coevolution and there is a possibility that CrPV-DcDV was not targeted
496 by ping-pong-derived piRNAs due to differences in the infection cycles between DcDV (a
497 naturally infecting DNA virus) and CrPV (a non-naturally infecting RNA virus). For example,
498 these may include potential differences in the spatial or physical accessibility of viral RNA to
499 piRNA biogenesis factors. Indeed, our results suggest that RNA viruses are not targeted by
500 piRNAs in *D. citri*. Thus, it is possible that even if EVE-derived piRNAs could prime ping-pong
501 amplification during infection with a cognate DNA virus, cognate RNA viruses may not be
502 susceptible to such targeting by the piRNA pathway. Future research aiming to uncover the
503 mechanisms of vpiRNA production during infection with DNA viruses such as DcDV and
504 AalDV-1 will help to resolve these questions.
505
506 We previously reported that while DcDV persistently infects CRF-TW *D. citri,* CRF-CA *D. citri*
507 are not susceptible to DcDV infection by oral acquisition or intrathoracic injection (34). We

508 report here that CRF-Uru *D. citri* can be persistently infected with DcDV by either of these
509 routes of infection. As discussed above, CRF-TW and CRF-Uru *D. citri* are likely to be of
510 lineage A, while CRF-CA *D. citri* are likely to be of lineage B. At present, no genotypic or
511 phenotypic differences between the two lineages have been described except for nucleotide
512 variation within the mitochondrial cytochrome oxidase subunit I gene (30–32). Genetic
513 background is known to contribute to susceptibility to virus infection within different
514 populations of the same species (57). In the case of Densoviruses, a single arginine residue at
515 position 188 within a mucin-like glycoprotein expressed in the midgut epithelium of *Bombyx*
516 *mori* strains susceptible to infection with Bombyx mori densovirus (BmDV) was substituted by
517 other amino acids in resistant strains (58). Further analysis showed that this mucin-like
518 glycoprotein was the cellular receptor for BmDV in the midgut epithelium and that the arginine
519 residue at position 188 was required for BmDV binding to this protein (58). EVEs have been
520 suggested to play a role in susceptibility to virus infection through a variety of potential
521 mechanisms (16,59–61). Our results suggest that ENS is present in lineage B *D. citri,* but absent
522 in lineage A *D. citri* and that the presence of ENS correlates with resistance to DcDV infection in
523 the populations tested. However, other genetic differences between the two lineages are likely to
524 exist and we cannot assess whether ENS may be involved with resistance to DcDV infection
525 based on the present data. Nevertheless, our data highlight both a genetic and phenotypic
526 difference between the two *D. citri* lineages. Other such differences between the lineages and
527 how they may contribute to susceptibility to DcDV infections should be the subject of future
528 investigations.
529
530

531 **Materials and methods**

532 *Maintenance of* D. citri *insects*

533 *D. citri* insects were reared on *C. macrophylla* plants in mesh cages (BugDorm, Taichung,
534 Taiwan) at 25 ± 2 °C under a 14:10 hour (light:dark) photoperiod and 60-70% relative humidity
535 at the University of California Davis CRF (62). *D. citri* insects from Uruguay and Taiwan were
536 imported under USDA APHIS-PPQ permit P526P-17-02906 by shipping adults and nymphs on
537 *C. macrophylla* or *Murraya paniculata* cuttings, respectively.

538

539

540 *sRNA sequencing and analysis*

541 For total insect sRNA sequencing of CRF-CA, CRF-TW, and CRF-Uru *D. citri*, total RNA was
542 extracted from groups of 50 adult *D. citri* insects using TRIzol reagent (Invitrogen, Carlsbad,
543 California) according to the manufacturer's instructions, but omitting the 75% ethanol wash. For
544 CRF-CA *D. citri* infected with wild-type CrPV or CrPV-DcDV by oral acquisition or injection
545 and for the progeny CRF-Uru *D. citri* infected with DcDV by injection, sRNAs were extracted
546 from groups of 25 adult *D. citri* insects using TRIzol reagent according to the manufacturer's
547 instructions, but omitting the 75% ethanol wash. For organ specific sRNA sequencing multiple
548 dissected organs were combined to be processed as one sample: 200 for gut, head, or hemolymph
549 and 100 for ovary or testis. RNA extraction was done from pooled organ samples with the
550 Direct-zol RNA Miniprep kit (Zymo Research, Irvine, California) following the manufacturer's
551 instructions. Three pooled samples of each organ were used. For all sRNA sequencing, total
552 RNA was sent to Beijing Genomics Institute for library prep with the TruSeq small RNA sample
553 preparation kit (Illumina, San Diego, California) and sequencing by 50 bp single end sequencing

554 with the BGISEQ-500 platform. For all sRNA reads, adaptor sequences were removed with Trim
555 Galore version 0.4.4 (63) and the fastxtoolkit was used to remove reads containing bases with
556 quality score <20 (64). The remaining reads were mapped to viral genomes, insect genomes, or
557 TEs using BowTie version 1.2.1.1 (65) using the default settings with the following exceptions: -
558 n 1 -l 20. When only perfectly mapped reads were considered (specifically indicated in the text),
559 the –v 0 option was used instead of the –n 1 –l 20 options. Sequence logos were produced using
560 WebLogo 3 (66). Ping-pong Z-scores were calculated with signature.py (67). All sRNA
561 sequence data generated for this study was deposited to the NCBI SRA database (BioProject
562 accession no. PRJNA629895).
563
564
565 *Analysis of EVEs*
566 DcDV-derived EVEs were identified in *D. citri* genomic scaffold 2850 by BLASTx using as
567 queries the deduced amino acid sequences of DcDV structural and non-structural proteins
568 (GenBank accession no.'s YP_009256210, YP_009256211, YP_009256212, and
569 YP_009256213). DcDV-derived EVEs were then inspected by aligning the DcDV genome
570 (GenBank accession no. NC_030296) to the nucleotide sequence of *D. citri* genomic scaffold
571 2850 using ClustalW (68). Deduced amino acid and nucleotide identities shared between ENS
572 and DcDV and between EITR and DcDV were calculated as p-distance with complete deletion
573 of gaps from ClustalW alignments using MEGA 7 (69).
574
575
576 *Identification of TEs*

577 We identified TEs present within the *D. citri* genome using RepeatMasker version 4.0.6 (70)
578 with the Metazoa library. In addition, to identify TEs lacking homology to previously annotated
579 TEs, we used RepeatModeler version 1.0.8 (71) to produce a *de novo* hidden Markov model for
580 TEs within the *D. citri* genome which was subsequently used as input for a second analysis using
581 RepeatMasker. All TE identification was performed using TEAnnotator.py as previously
582 described to produce a single strand-specific .fasta file containing the sequences of all TEs >100
583 nt identified in the *D. citri* genome (35).
584
585
586 *Nucleic acid extraction, reverse transcription, PCR, exonuclease digestion, and Sanger*
587 *sequencing*
588 Unless otherwise specified, DNA was extracted from groups of 25 homogenized *D. citri* insects
589 by phenol/chloroform extraction followed by ethanol precipitation. For PCR using exonuclease
590 treated DNA, 1 µg CRF-CA *D. citri* genomic DNA, an intact plasmid containing the ENS
591 amplicon, or the same plasmid digested overnight with HindIII-HF and SbfI-HF (New England
592 Biolabs, Ipswich, Massachusetts) were digested overnight with 10 units of Plasmid-Safe ATP-
593 Dependent DNAse (Lucigen, Middleton, Wisconsin) according to the manufacturer's
594 instructions. For analysis of ENS transcription, RNA was extracted from groups of 25
595 homogenized CRF-CA *D. citri* using TRIzol reagent according to the manufacturer's
596 instructions. RNA was treated twice with the Qiagen RNAse-free DNAse set (Qiagen, Hilden,
597 Germany) to remove DNA. cDNA was prepared from 50 ng RNA with SuperScript IV reverse
598 transcriptase (Invitrogen, Carlsbad, California) using specific primers (see the legend for Fig. 1)
599 according to the manufacturer’s instructions. To assess retention of the insert in CrPV-DcDV,

600 RNA was extracted from CrPV or CrPV-DcDV virions using TRIzol LS (Invitrogen, Carlsbad,
601 California) according to the manufacturer's instructions. cDNA was prepared from 50 ng virion
602 RNA using the Applied Biosystems high capacity reverse transcription kit (Applied Biosystems,
603 Foster City, California) using random primers according to the manufacturer's instructions and
604 cDNA was diluted 1:10 prior to PCR. Primer sequences are given in supplementary Table 1 and
605 their use is described in the appropriate figure legends. All PCR reactions were performed with
606 CloneAmp HiFi PCR premix (Takara Bio, Mountain View, California) using 25 ng of DNA, 1 µl
607 of cDNA (for determination of ENS transcription), or 3 µl diluted cDNA (for determination of
608 CrPV-DcDV insertion retention). Detailed PCR protocols are available upon request. PCR
609 products were analyzed by 0.8% or 2% agarose gel electrophoresis and visualized under UV
610 after staining with SYBR safe DNA gel stain (Invitrogen, Carlsbad, California). PCR products
611 were purified with the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, California)
612 and sequenced by Quintarabio using the Sanger method.
613
614
615 *Southern blotting*
616 5 µg undigested CRF-CA *D citri* DNA or 5 µg CRF-CA *D. citri* DNA digested overnight with
617 PstI-HF and HindIII-HF (New England Biolabs, Ibswich, Massachusetts) was electrophoresed in
618 a 0.8% agarose/0.5x TAE gel and the gel was prepared for transfer by incubation in 0.25 M HCl
619 for 30 minutes, then 0.5 M NaCl, 0.5 M NaOH for 30 minutes, and then 1.5 M NaCl, 0.5 M Tris-
620 HCl pH 7 for 30 minutes. DNA was then transferred to an Amersham Hybond Nx membrane
621 (GE Life Sciences, Boston, Massachusetts) for 24 hours in 20X SSC using the capillary method.

622 DNA was cross-linked to the membrane by exposure to 120,000 µjoles of UV light using a UV
623 Stratalinker 2400 (Stratagene, San Diego, California).
624
625 A PCR product corresponding to ENS was prepared by PCR with primers 27 and 3 using CRF-
626 CA *D. citri* DNA as a template. This PCR product was then used a template to prepare a
627 radioactive RNA probe by *in vitro* transcription using the MAXIscript T7 Transcription Kit
628 (Invitrogen, Carlsbad, California) with 32P-UTP according to the manufacturer's instructions.
629
630 The membrane with cross-linked DNA was prehybridized by incubation in Ambion ULTRAhyb
631 Ultrasensitive hybridization buffer (Invitrogen, Carlsbad, California) for 2 hours at 42 °C in a
632 rotary incubation oven. Following prehybridization, 2 x 10^6 CPM/mL radioactive probe was
633 added and hybridization proceeded for 16 hours at 42 °C in a rotary incubation oven. The
634 membrane was then washed at 42 °C for 15 minutes each with 2X SSC/0.1% SDS, 0.2X
635 SSC/0.1% SDS, and 0.1X SSC/0.1% SDS.
636
637
638 *DcDV infections and DcDV qPCR*
639 DcDV virions were partially purified from 0.5 g of CRF-TW *D. citri* as previously described
640 (34). The number of DcDV genome copies in the virion preparation was determined by qPCR
641 with primers 29 and 30 (see table S1) as described previously (34). For infections by
642 intrathoracic injection, the virion preparations were diluted to 8.2 x 10^5 genome copies/µl in 100
643 mM Tris-HCl, pH 7.5. Adult CRF-Uru *D. citri* were anesthetized on ice for approximately 15
644 minutes and intrathoracicly injected with approximately 200 nl of the diluted virion preparation

645 by manual injection using a syringe fitted to a 34 gauge stainless steel needle (Hamilton, Reno,
646 Nevada). Following injection, insects were maintained on *C. macrophylla* plants. For infections
647 by oral acquisition, the virion preparations were diluted to 8.2 x 10^5 genome copies/µl in 15%
648 sucrose, 0.1% green food coloring, 0.4% yellow food coloring prepared in 100 mM Tris-HCl, pH
649 7.5. Groups of 30 adult CRF-Uru *D. citri* were fed on the virus-containing sucrose solution for
650 96 hours by membrane feeding as described previously (34). Following the feeding period,
651 insects were maintained on *C. macrophylla* plants. Day 0 is designated as the day insects were
652 transferred from the virus containing artificial diet to plants. For both types of infection, adult
653 insects (>1 day old) were used and neither age nor gender of the insects was determined prior to
654 infection.
655
656 For both types of infection, five pools of three insects were collected every two days and DNA
657 was purified from each pool by phenol/chloroform extraction followed by ethanol precipitation
658 as described previously (34). All insects had been collected or were removed from the plants by
659 17 days post infection. During the maintenance period on plants, the DcDV-injected and -fed
660 insects laid eggs on the plants. Following removal of the DcDV-injected or –fed insects, the
661 plants were maintained to allow development of the progeny. The progeny were removed on the
662 day of emergency to the adult stage and DNA was extracted from 10 individual progeny insects
663 from each experiment as described above. A pool of 25 progeny of the DcDV-injected *D. citri*
664 was collected at the same time for sRNA purification. The concentration of DcDV genome
665 copies in the DNA pools or in DNA extracted from individual progeny insects was assessed by
666 qPCR with primers 29 and 30 (see Fig. S1) using 25 ng DNA as described previously (34).
667 Technical triplicates were used for all qPCR reactions.

668

669

670 *Cloning of CrPV-DcDV*

671 Complementary lyophilized oligonucleotides containing the DcDV sequence to be inserted
672 flanked upstream by 15 nt corresponding to the 3' end of the CrPV 1A nucleotide sequence and
673 downstream by 15 nt corresponding to the 5' end of the CrPV 2B nucleotide sequence were
674 suspended to a concentration of 100 µM in 100 mM potassium acetate, 300 µM HEPES, pH 7.5
675 (oligo sequences: 5'-
676 TCTAATCCTGGTCCTGCAGACCGTTCACCTTCTCCAGGACCTTCTACTGCATATCGCT
677 ATTGTAGCGAGGAAGTGCAATCGCGCCCC and 5' –
678 GGGGCGCGATTGCACTTCCTCGCTACAATAGCGATATGCAGTAGAAGGTCCTGGAG
679 AAGGTGAACGGTCTGCAGGACCAGGATTAGA). To anneal these oligonucleotides, the
680 oligonucleotides were mixed in equimolar ratios, incubated at 94 °C for 3 minutes, and allowed
681 to cool slowly to room temperature to form the duplex oligo. pCrPV-3 (a gift from Shou-Wei
682 Ding) (72) was linerized by inverse PCR with primers 13 and 14 and the duplex oligo was
683 ligated to the linearized plasmid in a 2:1 insert:vector molar ratio using NEBuilder Hifi
684 Assembly Master Mix (New England Biolabs, Ipswich, Massachusetts) according to the
685 manufacturer's instructions. The resulting plasmid (designated pCrPV-1A-DcDV) was
686 transformed into chemically competent *E. coli* DH5α and transformed cells were grown on luria
687 bertani (LB) agar plates containing 100 µg/mL ampicillin at 28 °C for 16 hours. Insert presence
688 was verified by colony PCR with primers 15 and 17 and by Sanger sequencing of colony PCR
689 products. Colonies containing the desired plasmid were used to inoculate 5 mL LB broth
690 containing 100 µg/mL ampicillin and the cultures were incubated at 28 °C for 16 hours with 220

691 rpm shaking. Plasmids were purified from overnight cultures with the QIAprep Spin Miniprep
692 Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.
693
694 To duplicate the 1A cleavage site on the 3' end of the recombinant DcDV sequence, pCrPV-1A-
695 DcDV was linearized by inverse PCR with primers 17 and 18 and the linearized plasmid was
696 circularized by blunt end ligation with T4 DNA ligase (New England Biolabs, Ipswich,
697 Massachusetts) according to the manufacturer's instructions to create pCrPV-1A-DcDV-1A. The
698 ligation product was transformed into chemically competent *E. coli* DH5α and plasmids were
699 purified as described.
700
701
702 *Transfection and infection of S2 Cells and virion purification*
703 pCrPV-3 and pCrPV-1A-DcDV-1A were linearized by digestion with BamHI-HF (New England
704 Biolabs, Ipswich, Massachusetts). Wild-type and CrPV-DcDV RNA was prepared by *in vitro*
705 transcription of 1 µg linearized plasmid using the mMESSAGE mMACHINE T7 transcription
706 kit (Invitrogen, Carlsbad, California) according to the manufacturer's instructions. 24 hours prior
707 to transfection, 5 x 10^6 S2 cells were seeded in 2 mL Schneider's Drosophila media (Thermo
708 Fisher, Waltham, Massachusetts) supplemented with 10% heat-inactivated fetal bovine serum in
709 9.6 cm^2 wells of a tissue culture plate and incubated at 26 °C. The following day, cells were
710 transfected with wild-type CrPV or CrPV-DcDV RNA using the TransMessenger transfection
711 reagent (Qiagen, Hilden, Germany) according to the manufacturer's instructions and the
712 transfected cells were incubated at 26 °C. After 72 hours, 1 mL of transfected cells was
713 transferred to a 50 mL tissue culture flask containing 3 x 10^6 S2 cells/mL that had been passed

714 24 hours prior in 6 mL Schneider's Drosophila media supplemented with 10% heat-inactivated
715 fetal bovine serum. After incubation at 26 °C for 96 hours, virions were purified from the
716 infected S2 cells by centrifugation at 8,000 RPM for 10 minutes at 4 °C using a Beckman GSA
717 rotor (Beckman Coulter, Brea, California). The supernatant was then collected and centrifuged at
718 8,000 RPM for 30 minutes at 4 °C using a Beckman GSA rotor. The supernatant was collected
719 and centrifuged through a 15% sucrose cushion (prepared in 10 mM Tris-HCl, pH 7.5) at 45,000
720 RPM for 45 minutes at 11 °C in a Beckman 70.1 Ti rotor (Beckman Coulter, Brea, California).
721 The pellet was resuspended in 10 mM Tris-HCl, pH 7.5 and then filtered through a 0.22 µm
722 filter.
723
724 The titer of purified CrPV-DcDV or CrPV virions was measured by end-point dilution. Briefly, 4
725 x 10^5 S2 cells were seeded in 500 µL Schneider's Drosophila media supplemented with 10%
726 heat-inactivated fetal bovine serum in each well of 24 well plates. Cells were incubated for 24
727 hours at 26°C. After 24 hours cells were infected with tenfold serial dilutions of purified virions
728 in 100 µL volumes. Dilutions over the range of 10^{-5} to 10^{-10} were used and six individual wells
729 were used for each dilution. After 72 hours, cells were examined for the presence cytopathic
730 effects and the titer of the undiluted virus stock was calculated from these results using the Reed
731 and Muench method.
732
733 For infection of S2 cells using purified virions, 5 x 10^6 S2 cells were seeded in 2 mL Schneider's
734 Drosophila media supplemented with 10% heat-inactivated fetal bovine serum in 9.6 cm^2 wells
735 of a tissue culture plate and incubated at 26 °C. After 24 hours, 100 $TCID_{50}$ units of wild-type

736 CrPV or CrPV-DcDV virions suspended in 100 µL Schneider's Drosophila media were added to
737 the wells.
738
739
740 *Microscopy*
741 Wild-type CrPV or CrPV-DcDV virions (see above) were further purified by centrifugation
742 through a 1.2 g/cm^3 to 1.6 g/cm^3 cesium chloride density gradient at 50,000 RPM for 4 hours at
743 11 °C in a Beckman SW 65 Ti rotor (Beckman Coulter, Brea, California). Visible "virus bands"
744 in the gradients were collected using a Hamilton syringe and diluted 1:10 in 10 mM Tris-HCl,
745 pH 7.5. Diluted virions were then centrifuged at 75,000 RPM for 30 minutes at 11 °C in a
746 Beckman TLA 120 rotor (Beckman Coulter, Brea, California). The pellet was resuspended in 10
747 mM Tris-HCl, pH 7.5 and negatively stained with uranyl formate. Negatively stained virion
748 samples were observed by transmission electron microscopy using a JEOL 1230 electron
749 microscope (JEOL, Tokyo, Japan) operating at 100 kV. Mock-infected S2 cells and S2 cells
750 infected with wild-type CrPV or CrPV-DcDV were observed by bright-field microscopy using a
751 Leica DM5000B microscope (Leica, Wetzlar, Germany).
752
753
754 *CrPV infections in* D. citri
755 Adult CRF-CA *D. citri* were infected with purified CrPV or CrPV-DcDV virions by
756 intrathoracic injection or membrane feeding exactly as described above for DcDV infections. For
757 injections, purified virions were diluted to 5000 $TCID_{50}$ units/µl and insects were injected with
758 approximately 200 nl each. Seven pools of three insects each were collected on days 0, 3, 6, and

759 9 for RNA extraction. Three additional pools of 25 insects each were collected on day 9 for
760 sRNA extraction. For membrane feeding, purified virions were diluted to 10^9 $TCID_{50}$ units/µl.
761 Five pools of three insects each were collected on days 0, 3, and 5 for RNA extraction. Three
762 additional pools of 25 insects each were collected on day 5 for sRNA extraction.
763 To determine viral RNA levels, RNA was extracted from the pools of three insects using TRIzol
764 reagent according to the manufacturer's instructions. RNA pellets were resuspended in 25 µl
765 water and cDNA was prepared from 1 µl RNA (approximately 200-400 ng RNA) with the high
766 capacity cDNA reverse transcription kit using random primers according to the manufacturer's
767 instructions and cDNA was diluted 1:10. Wild-type CrPV and CrPV-DcDV RNA levels were
768 determined using 4.5 µL diluted cDNA for qPCR reactions with the SsoAdvanced Universal
769 SYBR Green Supermix (BioRad, Hercules, California) in 10 µL reactions. Primers 19 and 20
770 were used for CrPV RNA and primers 21 and 22 were used for *D. citri* actin. All primers were
771 used at a final concentration of 0.25 nM. The qPCR reaction conditions consisted of an initial
772 denaturation at 98 °C for 2 min followed by 40 cycles at 95 °C for 10 s and 60 °C for 20 s.
773 CrPV RNA levels were normalized based on the expression of *D. citri* actin using the $2^{-\Delta\Delta Ct}$
774 method (73). Technical triplicates were used for al qPCR reactions. Diluted cDNA from each
775 day was pooled to assess retention of the insert in CrPV-DcDV as described above.
776
777

778 **Acknowledgements**

779 We thank E. Matsumura for assistance and protocols for CrPV infections. We thank colleagues
780 who provided us *D. citri* samples: W.O. Dawson, H. Wuriyanghan, H.-H Yeh, Y. Cen, A. M.
781 Khan, M. A. Machado, T. da Silva, D. M. Galdeano, D. Jenkins, C. Higashi, A. Chow, D.

782 Morgan, K. Pelz-Stelinski, K. Godfrey, B. Baker, G. Simmons, M. Keremane, and J. Buenahora.
783 We thank S.W.-Ding for providing us with pCrPV-3. We also thank M-C. Saleh for critical
784 reading of the manuscript. The U.S. Department of Agriculture provided funding to Bryce W.
785 Falk under grant numbers 13-002NU-781 and 2015-70016-23011. This material is based upon
786 work supported by the National Science Foundation Graduate Research Fellowship Program
787 under Grant No. 1650042.

788

789

790 **Competing interests**

791 The authors declare that they have no competing interests.

792

793

794 **References**

795 1. Czech B, Hannon GJ. One loop to rule them all: the ping-pong cycle and piRNA-guided
796 silencing. Trends Biochem Sci. 2016;41(4):324–37.

797 2. Dietrich I, Jansen S, Fall G, Lorenzen S, Rudolf M, Huber K, et al. RNA interference
798 restricts Rift Valley Fever virus in multiple insect systems. MSphere. 2017;2(3):e00090-
799 17.

800 3. Varjak M, Maringer K, Watson M, Sreenu VB, Fredericks AC, Pondeville E, et al. Aedes
801 aegypti Piwi4 is a noncanonical PIWI protein involved in antiviral responses. MSphere.
802 2017;2(3):e00144-17.

803 4. Dietrich I, Shi X, McFarlane M, Watson M, Blomström A-L, Skelton JK, et al. The

804 antiviral RNAi response in vector and non-vector cells against orthobunyaviruses. PLoS
805 Negl Trop Dis. 2017;11(1):e0005272.

806 5. Schnettler E, Donald CL, Human S, Watson M, Siu RWC, McFarlane M, et al.
807 Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus
808 production in mosquito cells. J Gen Virol. 2013;94:1680-89.

809 6. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete
810 small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell.
811 2007;128(6):1089–103.

812 7. Vodovar N, Bronkhorst AW, Van Cleef KWR, Miesen P, Blanc H, Van Rij RP, et al.
813 Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS One.
814 2012;7(1):e30861.

815 8. Leger P, Lara E, Jagla B, Sismeiro O, Mansuroglu Z, Coppée JY, et al. Dicer-2-and Piwi-
816 mediated RNA interference in Rift Valley fever virus-infected mosquito cells. J Virol.
817 2013;87(3):1631–48.

818 9. Miesen P, Girardi E, van Rij RP. Distinct sets of PIWI proteins produce arbovirus and
819 transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res.
820 2015;43(13):6545–56.

821 10. Miesen P, Ivens A, Buck AH, van Rij RP. Small RNA profiling in dengue virus 2-infected
822 Aedes mosquito cells reveals viral piRNAs and novel host miRNAs. PLoS Negl Trop Dis.
823 2016;10(2):e0004452.

824 11. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. Production of virus-

825 derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. PLoS
826 Pathog. 2012;8(1):e1002470.

827 12. Goic B, Stapleford KA, Frangeul L, Doucet AJ, Gausson V, Blanc H, et al. Virus-derived
828 DNA drives mosquito vector tolerance to arboviral infection. Nat Commun.
829 2016;7:12410.

830 13. Campbell CL, Black WC, Hess AM, Foy BD. Comparative genomics of small RNA
831 regulatory pathway components in vector mosquitoes. BMC Genomics. 2008;9(1):425.

832 14. Petit M, Mongelli V, Frangeul L, Blanc H, Jiggins F, Saleh M-C. piRNA pathway is not
833 required for antiviral defense in Drosophila melanogaster. Proc Natl Acad Sci.
834 2016;113(29):E4218–27.

835 15. ter Horst AM, Nigg JC, Dekker FM, Falk BW. Endogenous Viral Elements Are
836 Widespread in Arthropod Genomes and Commonly Give Rise to PIWI-Interacting RNAs.
837 J Virol. 2019;93(6):e02124-18.

838 16. Whitfield ZJ, Dolan PT, Kunitomi M, Tassetto M, Seetin MG, Oh S, et al. The diversity,
839 structure, and function of heritable adaptive immunity sequences in the Aedes aegypti
840 genome. Curr Biol. 2017;27(22):3511–9.

841 17. Palatini U, Miesen P, Carballar-Lejarazu R, Ometto L, Rizzo E, Tu Z, et al. Comparative
842 genomics shows that viral integrations are abundant and express piRNAs in the arboviral
843 vectors Aedes aegypti and Aedes albopictus. BMC Genomics. 2017;18(1):512.

844 18. Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. PLoS Genet.
845 2010;6(11):e1001191.

19. François S, Filloux D, Roumagnac P, Bigot D, Gayral P, Martin DP, et al. Discovery of parvovirus-related sequences in an unexpected broad range of animals. Sci Rep. 2016;6:30880.

20. Suzuki Y, Frangeul L, Dickson LB, Blanc H, Verdier Y, Vinh J, et al. Uncovering the repertoire of endogenous flaviviral elements in Aedes mosquito genomes. J Virol. 2017;91(15):e00571-17.

21. Tassetto M, Kunitomi M, Whitfield ZJ, Dolan PT, Sánchez-Vargas I, Garcia-Knight M, et al. Control of RNA viruses in mosquito cells through the acquisition of vDNA and endogenous viral elements. Elife. 2019;8:e41244.

22. Grafton-Cardwell EE, Stelinski LL, Stansly PA. Biology and management of Asian citrus psyllid, vector of the huanglongbing pathogens. Annu Rev Entomol. 2013;58:413–32.

23. Singerman A, Rogers ME. The economic challenges of dealing with citrus greening: The case of Florida. J Integr Pest Manag. 2020;11(1):3.

24. Tiwari S, Mann RS, Rogers ME, Stelinski LL. Insecticide resistance in field populations of Asian citrus psyllid in Florida. Pest Manag Sci. 2011;67(10):1258–68.

25. Juan-Blasco M, Qureshi JA, Urbaneja A, Stansly PA. Predatory mite, Amblyseius swirskii (Acari: Phytoseiidae), for biological control of Asian citrus psyllid, Diaphorina citri (Hemiptera: Psyllidae). Florida Entomol. 2012;95(3):543–51.

26. Étienne J, Quilici S, Marival D, Franck A. Biological control of Diaphorina citri (Hemiptera: Psyllidae) in Guadeloupe by imported Tamarixia radiata (hymenoptera: Eulophidae). Fruits. 2001;56(5):307–15.

867 27. Hajeri S, Killiny N, El-Mohtar C, Dawson WO, Gowda S. Citrus tristeza virus-based
868 RNAi in citrus plants induces gene silencing in Diaphorina citri, a phloem-sap sucking
869 insect vector of citrus greening disease (Huanglongbing). J Biotechnol. 2014;176:42–9.

870 28. Nouri S, Salem N, Nigg JC, Falk BW. Diverse array of new viral sequences identified in
871 worldwide populations of the Asian citrus psyllid (Diaphorina citri) using viral
872 metagenomics. J Virol. 2016;90(5):2434–45.

873 29. Saha S, Hosmani PS, Villalobos-Ayala K, Miller S, Shippy T, Flores M, et al. Improved
874 annotation of the insect vector of citrus greening disease: biocuration by a diverse
875 genomics community. Database. 2017;2017:bax032.

876 30. Luo Y, Agnarsson I. Global mt DNA genetic structure and hypothesized invasion history
877 of a major pest of citrus, Diaphorina citri (Hemiptera: Liviidae). Ecol Evol.
878 2018;8(1):257–65.

879 31. De León JH, Sétamou M, Gastaminza GA, Buenahora J, Cáceres S, Yamamoto PT, et al.
880 Two separate introductions of Asian citrus psyllid populations found in the American
881 continents. Ann Entomol Soc Am. 2011;104(6):1392–8.

882 32. Boykin LM, De Barro P, Hall DG, Hunter WB, McKenzie CL, Powell CA, et al.
883 Overview of worldwide diversity of Diaphorina citri Kuwayama mitochondrial
884 cytochrome oxidase 1 haplotypes: two Old World lineages and a New World invasion.
885 Bull Entomol Res. 2012;102(5):573–82.

886 33. Goic B, Vodovar N, Mondotte JA, Monot C, Frangeul L, Blanc H, et al. RNA-mediated
887 interference and reverse transcription control the persistence of RNA viruses in the insect

888 model Drosophila. Nat Immunol. 2013;14(4):396.

889 34. Nigg JC, Falk BW. Diaphorina citri densovirus is a persistently infecting virus with a
890 hybrid genome organization and unique transcription strategy. J Gen Virol. 2020;101:226-
891 239.

892 35. Lewis SH, Quarles KA, Yang Y, Tanguy M, Frézal L, Smith SA, et al. Pan-arthropod
893 analysis reveals somatic piRNAs as an ancestral defence against transposable elements.
894 Nat Ecol Evol. 2018;2(1):174.

895 36. Matsumura EE, Nerva L, Nigg JC, Falk BW, Nouri S. Complete Genome Sequence of the
896 Largest Known Flavi-Like Virus, Diaphorina citri flavi-like virus, a Novel Virus of the
897 Asian Citrus Psyllid, Diaphorina citri. Genome Announc. 2016;4(5):e00946-16.

898 37. Marutani-Hert M, Hunter WB, Katsar CS, Sinisterra XH, Hall DG, Powell CA. Reovirus-
899 like sequences isolated from adult Asian citrus psyllid,(Hemiptera: Psyllidae: Diaphorina
900 citri). Florida Entomol. 2009;92(2):314–21.

901 38. Chen Q, Godfrey K, Liu J, Mao Q, Kuo Y-W, Falk BW. A nonstructural protein
902 responsible for viral spread of a novel insect reovirus provides a safe channel for
903 biparental virus transmission to progeny. J Virol. 2019;93(15):e00702-19.

904 39. Reinganum C, O'Loughlin G., Hogan TW. A nonoccluded virus of the field crickets
905 Teleogryllus océaniens and T.commodus (Orthoptera: Gryllidae). J Invertebr Pathol.
906 1970;16:216–20.

907 40. Nayak A, Berry B, Tassetto M, Kunitomi M, Acevedo A, Deng C, et al. Cricket paralysis
908 virus antagonizes Argonaute 2 to modulate antiviral defense in Drosophila. Nat Struct Mol

909 Biol. 2010;17(5):547.

910 41. Costa A, Jan E, Sarnow P, Schneider D. The Imd pathway is involved in antiviral immune
911 responses in Drosophila. PLoS One. 2009;4(10):e7436.

912 42. Barckmann B, Pierson S, Dufourt J, Papin C, Armenise C, Port F, et al. Aubergine iCLIP
913 reveals piRNA-dependent decay of mRNAs involved in germ cell development in the
914 early embryo. Cell Rep. 2015;12(7):1205–16.

915 43. Bagijn MP, Goldstein LD, Sapetschnig A, Weick E-M, Bouasker S, Lehrbach NJ, et al.
916 Function, targets, and evolution of Caenorhabditis elegans piRNAs. Science (80-).
917 2012;337(6094):574–8.

918 44. Mohn F, Handler D, Brennecke J. piRNA-guided slicing specifies transcripts for
919 Zucchini-dependent, phased piRNA biogenesis. Science. 2015;348(6236):812–7.

920 45. Mondotte JA, Gausson V, Frangeul L, Blanc H, Lambrechts L, Saleh M-C. Immune
921 priming and clearance of orally acquired RNA viruses in Drosophila. Nat Microbiol.
922 2018;3(12):1394.

923 46. Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li W-X, et al. Virus discovery by deep sequencing
924 and assembly of virus-derived small silencing RNAs. Proc Natl Acad Sci.
925 2010;107(4):1606–11.

926 47. Parry R, Bishop C, De Hayr L, Asgari S. Density-dependent enhanced replication of a
927 densovirus in Wolbachia-infected Aedes cells is associated with production of piRNAs
928 and higher virus-derived siRNAs. Virology. 2019;528:89–100.

929 48. Tijssen P, Bando H, Li Y, Jousset F, Zadori Z, Fediere G, et al. Evolution of densoviruses.

930 Parvoviruses. 2005;5:55–60.

931 49. Taning CNT, Andrade EC, Hunter WB, Christiaens O, Smagghe G. Asian citrus psyllid
932 RNAi pathway–RNAi evidence. Sci Rep. 2016;6:38082.

933 50. Pinheiro P V, Wilson JR, Xu Y, Zheng Y, Rebelo AR, Fattah-Hosseini S, et al. Plant
934 Viruses Transmitted in Two Different Modes Produce Differing Effects on Small RNA-
935 Mediated Processes in Their Aphid Vector. Phytobiomes J. 2019;3(1):71–81.

936 51. Ma M, Huang Y, Gong Z, Zhuang L, Li C, Yang H, et al. Discovery of DNA viruses in
937 wild-caught mosquitoes using small RNA high throughput sequencing. PLoS One.
938 2011;6(9):e24758.

939 52. Bronkhorst AW, van Rij RP. The long and short of antiviral defense: small RNA-based
940 immunity in insects. Curr Opin Virol. 2014;7:19–28.

941 53. Bronkhorst AW, van Cleef KWR, Vodovar N, İnce İA, Blanc H, Vlak JM, et al. The DNA
942 virus Invertebrate iridescent virus 6 is a target of the Drosophila RNAi machinery. Proc
943 Natl Acad Sci. 2012;109(51):E3604–13.

944 54. Jayachandran B, Hussain M, Asgari S. RNA interference as a cellular defense mechanism
945 against the DNA virus baculovirus. J Virol. 2012;86(24):13729–34.

946 55. Karamipour N, Fathipour Y, Talebi AA, Asgari S, Mehrabadi M. Small interfering RNA
947 pathway contributes to antiviral immunity in Spodoptera frugiperda (Sf9) cells following
948 Autographa californica multiple nucleopolyhedrovirus infection. Insect Biochem Mol
949 Biol. 2018;101:24–31.

950 56. Kemp C, Mueller S, Goto A, Barbier V, Paro S, Bonnay F, et al. Broad RNA

951 interference–mediated antiviral immunity and virus-specific inducible responses in
952 Drosophila. J Immunol. 2013;190(2):650–8.

953 57. Palmer WH, Varghese FS, Van Rij RP. Natural variation in resistance to virus infection in
954 dipteran insects. Viruses. 2018;10(3):118.

955 58. Ito K, Kidokoro K, Katsuma S, Sezutsu H, Uchino K, Kobayashi I, et al. A single amino
956 acid substitution in the Bombyx-specific mucin-like membrane protein causes resistance
957 to Bombyx mori densovirus. Sci Rep. 2018;8(1):7430.

958 59. Klenerman P, Hengartner H, Zinkernagel RM. A non-retroviral RNA virus persists in
959 DNA form. Nature. 1997;390(6657):298.

960 60. Fujino K, Horie M, Honda T, Merriman DK, Tomonaga K. Inhibition of Borna disease
961 virus replication by an endogenous bornavirus-like element in the ground squirrel genome.
962 Proc Natl Acad Sci. 2014;111(36):13175–80.

963 61. Maori E, Tanne E, Sela I. Reciprocal sequence exchange between non-retro viruses and
964 hosts leading to the appearance of new host phenotypes. Virology. 2007;362(2):342–9.

965 62. Galdeano DM, Breton MC, Lopes JRS, Falk BW, Machado MA. Oral delivery of double-
966 stranded RNAs induces mortality in nymphs and adults of the Asian citrus psyllid,
967 Diaphorina citri. PLoS One. 2017;12(3):e0171847.

968 63. Krueger F. Trim galore. A wrapper tool around Cutadapt FastQC to consistently apply
969 Qual Adapt trimming to FastQ files. 2015;

970 64. Gordon A, Hannon GJ. Fastx-toolkit. FASTQ/A short-reads preprocessing tools
971 (unpublished) http://hannonlab cshl edu/fastx_toolkit. 2010;5.

972 65. Langmead B. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinforma.
973 2010;32(1):11–7.

974 66. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator,
975 Genome Research. WebLogo a Seq logo Gener Genome Res. 2004;14(6).

976 67. Antoniewski C. Computing siRNA and piRNA overlap signatures. In: Animal Endo-
977 SiRNAs. Springer; 2014. p. 135–46.

978 68. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al.
979 Clustal W and Clustal X version 2.0. bioinformatics. 2007;23(21):2947–8.

980 69. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis
981 version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.

982 70. Smit A, Hubley R, Green P. RepeatMasker Open-4.0 [Internet]. 2013-2015. Available
983 from: http://www.repeatmasker.org

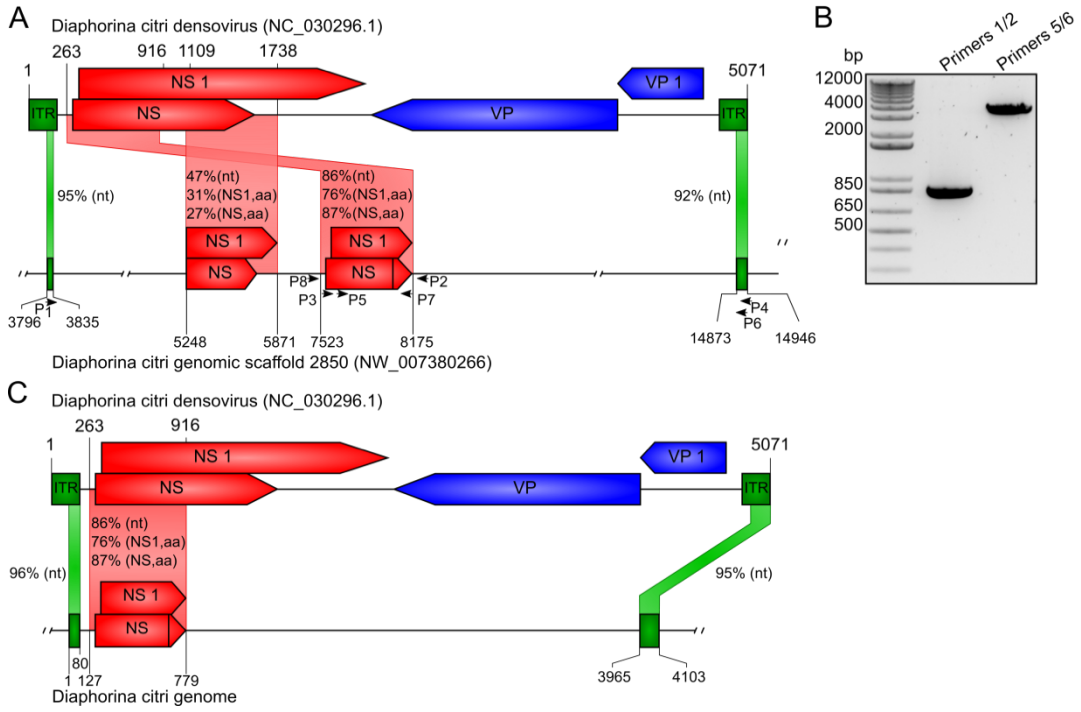
984 71. Smit A, Hubley R. RepeatModeler Open-1.0 [Internet]. 2008. Available from:
985 http://www.repeatmasker.org/RepeatModeler/

986 72. Kerr CH, Wang QS, Keatings K, Khong A, Allan D, Yip CK, et al. The 5′ untranslated
987 region of a novel infectious molecular clone of the dicistrovirus cricket paralysis virus
988 modulates infection. J Virol. 2015;89(11):5919–34.

989 73. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C T method.
990 Nat Protoc. 2008;3(6):1101.

991

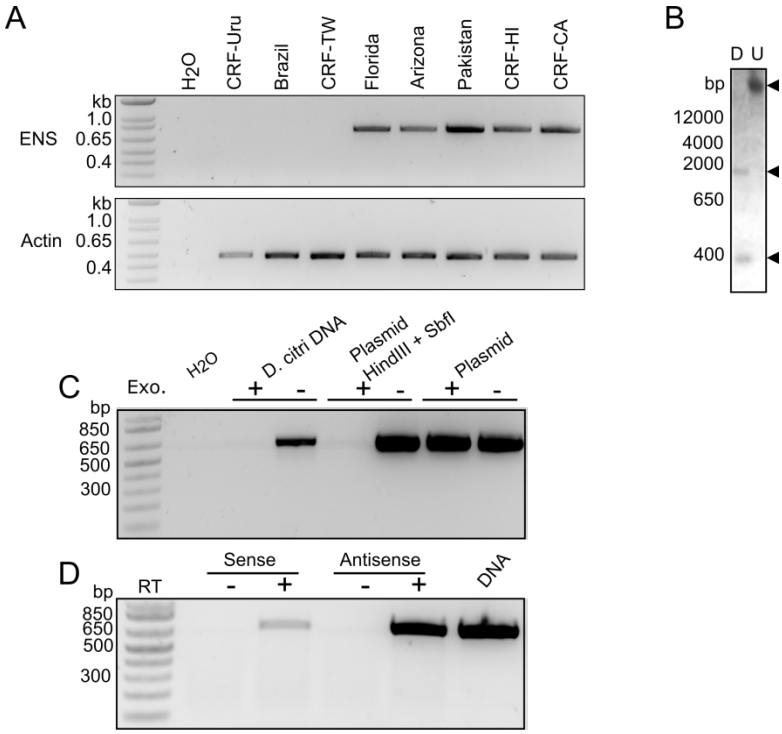
992 **Fig. 1**
993



994
995

996 **Fig. 1.** A DcDV-like EVE is present in the *D. citri* genome. (A) Organization of DcDV-derived
997 EVEs identified in *D. citri* genomic scaffold 2850 by BLASTx followed by manual sequence
998 alignment. The DcDV genome organization is shown on top with the corresponding region of
999 scaffold 2850 on the bottom. Numbers above and below the sequence depictions represent
1000 nucleotide positions. Vertical lines inside shaded boxes represent stop codons. The percent
1001 nucleotide or deduced amino acid (aa) identity between EVEs and corresponding viral genomic
1002 regions/proteins is given. Annealing positions of PCR primers are shown with arrows. (B)
1003 Confirmation of EVE presence by PCR using primers shown in panel A. The product produced
1004 with primers 5/6 was produced by nested PCR using as template a 1:1000 dilution of a PCR
1005 product produced with primers 3/4. (C) The correct organization of the region of the *D. citri*
1006 genome containing ENS and EITR, as confirmed by sequencing of the PCR products shown in
1007 panel B.
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019

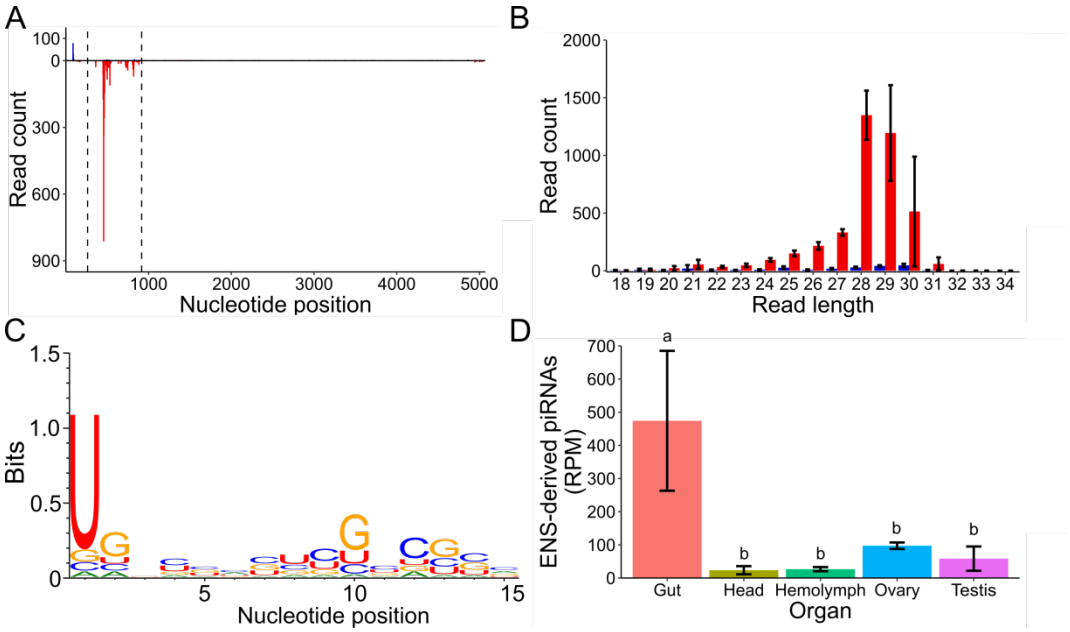
1020 **Fig. 2**
1021



1022
1023

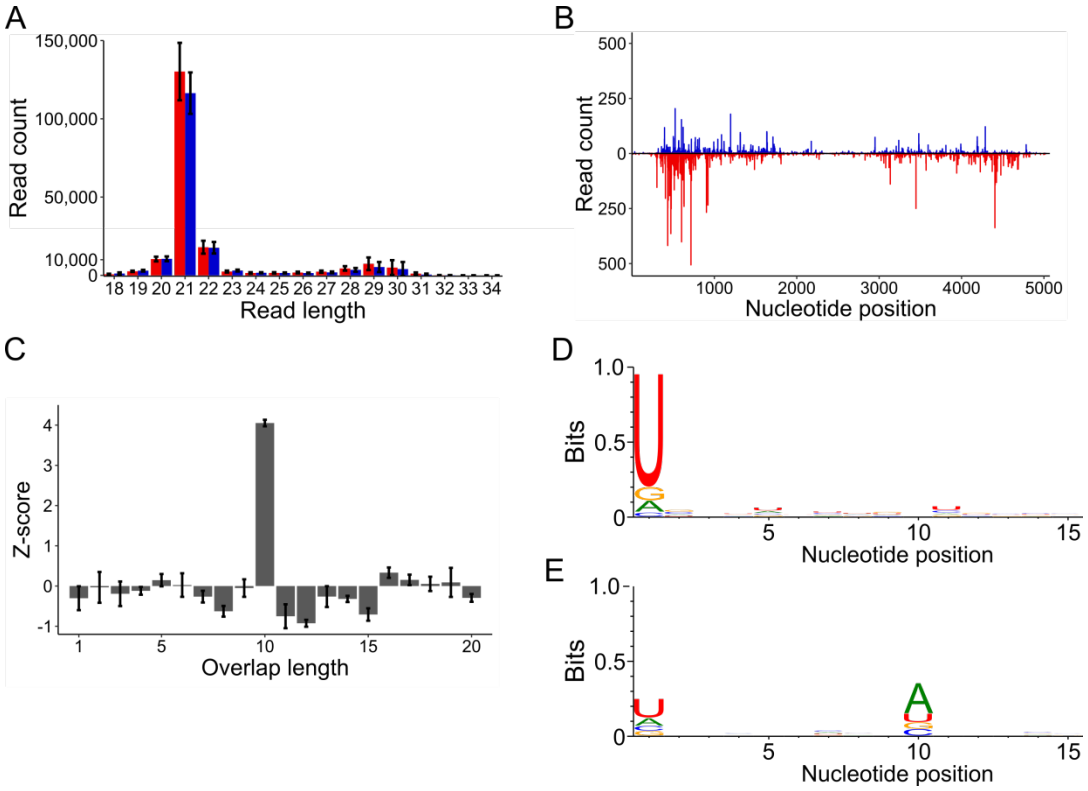
1024 **Fig 2.** ENS is a transcribed EVE that is unevenly distributed among distinct *D. citri* populations.
1025 (A) Upper: PCR products produced using primers flanking ENS (primers 2 and 8). Lower: PCR
1026 products produced using primers specific to *D. citri* actin (primers 9 and 10). (B) Southern blot
1027 of *D. citri* genomic DNA using an RNA probe based on the sequence of ENS; U = undigested, D
1028 = digested with PstI and HindIII. Upper arrow denotes ENS in undigested genomic DNA. Lower
1029 arrows denote cleavage products. (C) PCR products produced using primers 3 and 7 and the
1030 indicated DNA samples. "Plasmid" is a plasmid containing the full ENS sequence and "Plasmid
1031 HindIII + SbfI" is the same plasmid digested with HindIII and SbfI. DNA was left intact or
1032 digested with an exonuclease (Exo.) prior to PCR. (D) Primers 3 or 7 were used to generate
1033 cDNA from antisense or sense transcripts, respectively. cDNAs were used as templates for PCR
1034 using primers 3 and 7. DNA or cDNA prepared without reverse transcriptase (RT) served as
1035 controls.
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047

1048 **Fig. 3**
1049



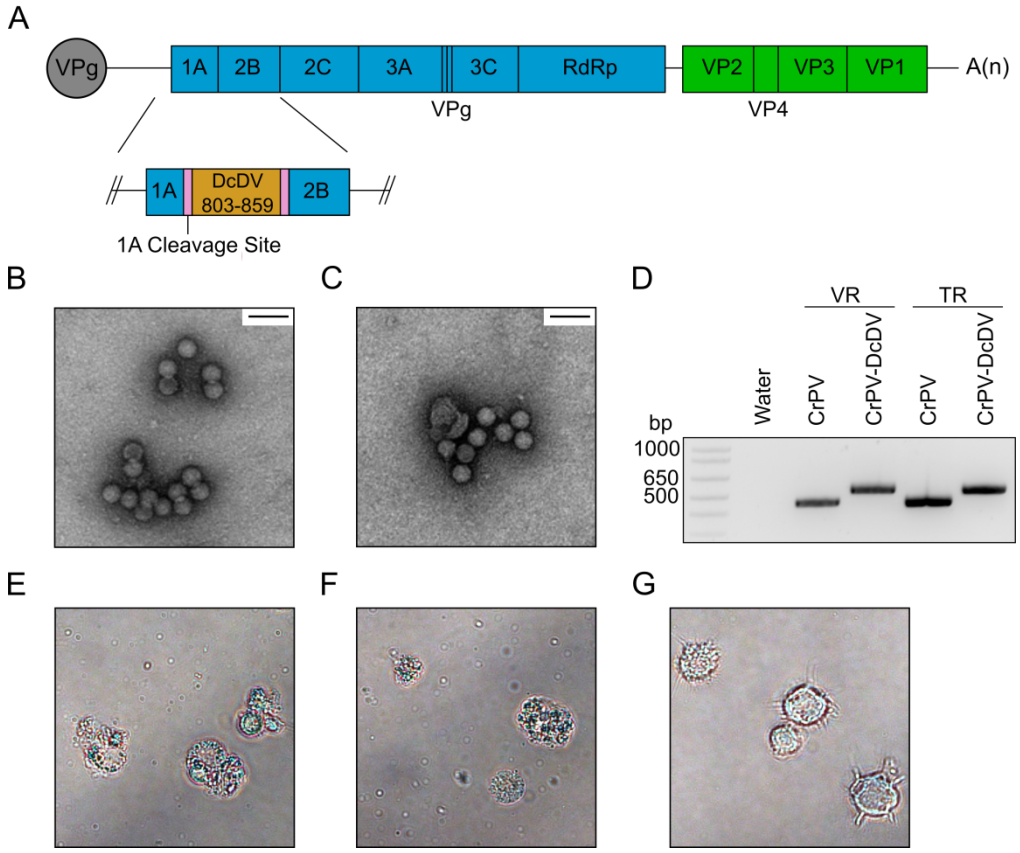
1050
1051
1052 **Fig. 3.** DcDV-specific piRNAs are produced from ENS in *D. citri* from CRF-CA. (A) Positions
1053 of all sRNAs from CRF-CA *D. citri* mapped to the DcDV genome. Dashed lines indicate the
1054 region of the DcDV genome corresponding to ENS. sRNAs are shown as mapped to the genomic
1055 strand containing the coding sequence for the NS proteins. Red = antisense sRNAs, blue = sense
1056 sRNAs. Reads counts are an average of three independent libraries. (B) Length distribution of
1057 sRNAs shown in (A). Red = antisense, blue = sense. Read counts are an average of three
1058 independent libraries. Error bars indicate standard deviation. (C) Sequence logo of sRNAs shown
1059 in (A). Data represents three pooled libraries. (D) Abundance of 27-32 nt sRNAs from CRF-CA
1060 *D. citri* mapping to ENS in the indicated tissue types. Read counts for each library were
1061 normalized using the read counts per million mapped reads method (RPM)(59). Normalized read
1062 counts represent the average of three independent libraries. Error bars indicate standard
1063 deviation. Average RPM values for each tissue were compared by one-way ANOVA and
1064 Turkey's honestly significant difference post-hoc test. Significance is indicated by lowercase
1065 letters and tissues sharing a letter do not have significantly different RPM values ($p < 0.05$).
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078

1079 **Fig. 4**
1080



1081
1082
1083 **Fig. 4.** DcDV is targeted by ping-pong-dependent vpiRNAs in DcDV-infected *D. citri* insects
1084 from CRF-TW. (A) Length distribution of sRNAs mapping to the DcDV genome in DcDV-
1085 infected *D. citri* insects from CRF-TW. To account for the bidirectional transcription strategy of
1086 DcDV, sRNA mapping polarity was assigned from mapping location based on the start and stop
1087 positions of the canonical DcDV transcripts. Red = antisense sRNAs, blue = sense sRNAs. Read
1088 counts are an average of three independent libraries. Error bars indicate standard deviation. (B)
1089 Positions of 27-32 nt sRNAs shown in (A). Red = antisense sRNAs, blue = sense sRNAs. Read
1090 counts are an average of three independent libraries. (C) Z-scores for the indicated overlap
1091 distances between the 5' ends of complementary 27-32 nt sRNAs shown in (A). Z-scores are an
1092 average of three independent libraries. Error bars indicate standard deviation. (D & E) Sequence
1093 logos for the 27-32 nt sRNAs shown in (A). Sequence logos for antisense sRNAs (D) or sense
1094 sRNAs (E) are shown. Data represents three pooled libraries.
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104

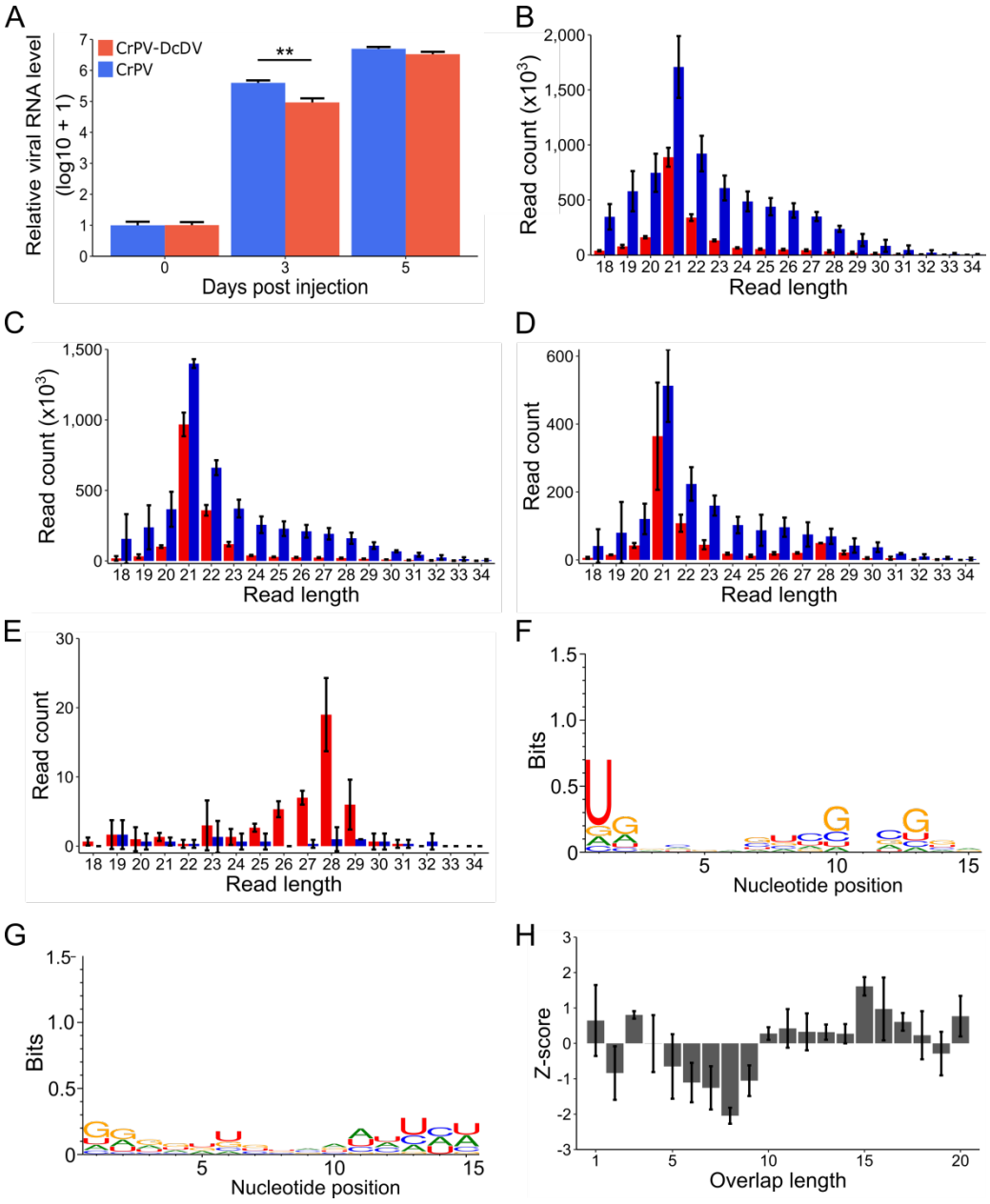
1105 **Fig. 5**
1106



1107
1108

1109 **Fig. 5.** Construction of CrPV-DcDV, a recombinant CrPV mutant containing 57 nt of sequence
1110 from the DcDV genome. (A) Genome organization of CrPV DcDV. Blue rectangles represent
1111 the CrPV non-structural proteins. RdRp = RNA dependent RNA polymerase. VPg = viral protein
1112 genome-linked. Green boxes represent the CrPV structural proteins. The orange rectangle
1113 represents the recombinant DcDV sequence, which corresponds to nucleotides 803-859 from the
1114 DcDV genome. Pink rectangles represent the cleavage site at which the 1A protein is released
1115 from the polyprotein. (B & C) Electron micrographs of wild-type CrPV (B) or CrPV-DcDV (C)
1116 virions purified from S2 cells transfected with viral RNA. 50,000x magnification. Scale bar is 50
1117 nm. (D) RT-PCR products produced using primers flanking the site into which recombinant
1118 DcDV sequence was inserted in CrPV-DcDV (primers 11 and 12). RNA extracted from purified
1119 virions (VR) or *in vitro* transcribed viral RNA (TR) was used as a template. (E-G) Bright-field
1120 microscopy images of S2 cells infected with wild-type CrPV virions (E), CrPV-DcDV virions
1121 (F), or mock infected (G). Images were acquired 72 hours post-infection.
1122
1123
1124
1125
1126
1127
1128

1129 **Fig. 6**
1130

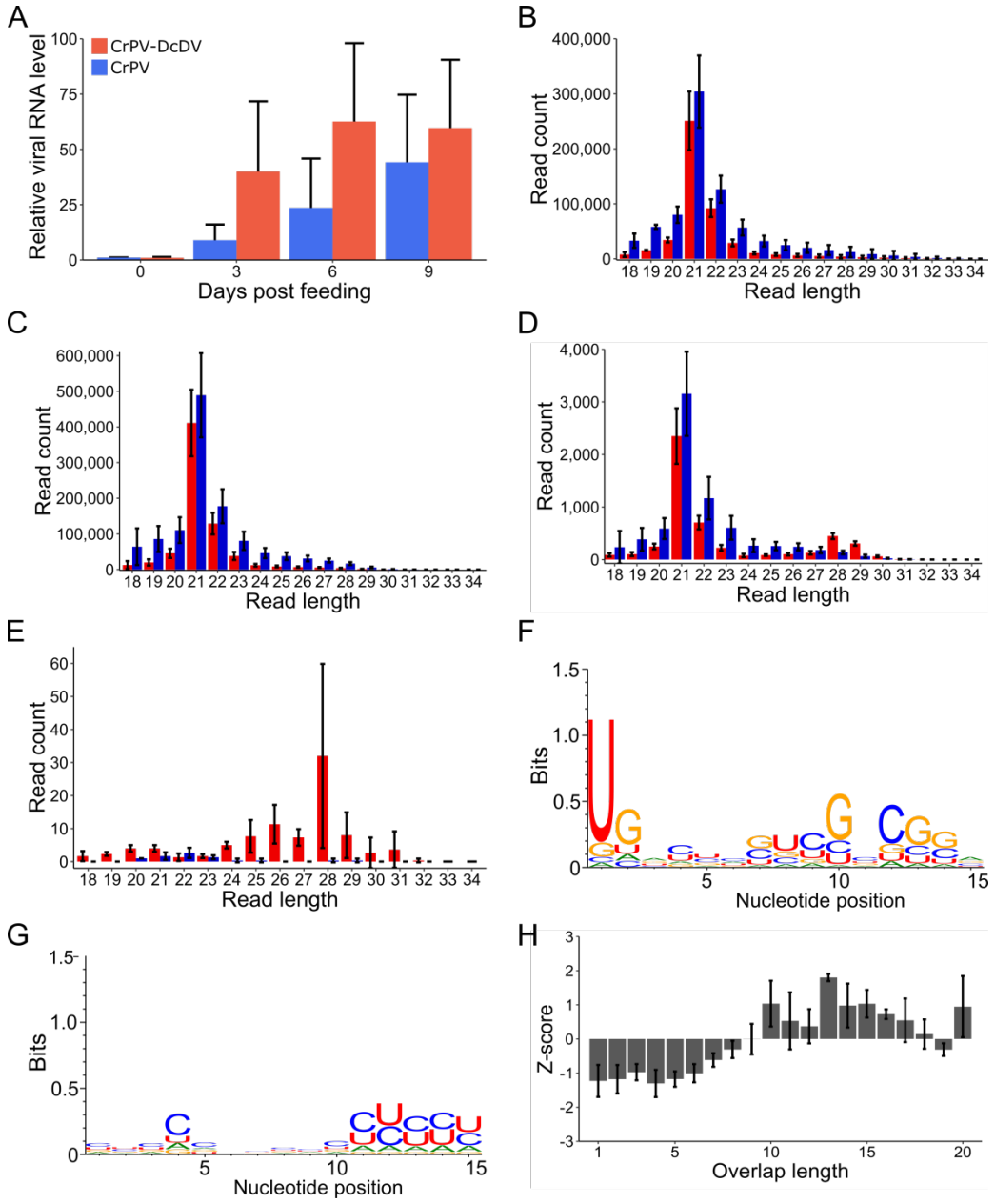


1131
1132

1133 **Fig. 6.** CrPV-DcDV is not targeted by piRNAs during infection initiated by intrathoracic
1134 injection of purified virions. (A) Relative viral RNA levels during infection with wild-type CrPV
1135 or CrPV-DcDV. Infection was initiated by intrathoracic injection of CRF-CA *D. citri* with 1000
1136 $TCID_{50}$ units of purified virions per insect. Viral RNA levels were assessed by RT-qPCR and
1137 normalized based on the expression of actin. The amount of viral RNA present on day 0 was set
1138 as 1 and log10 + 1 levels of viral RNA are shown relative to this value. Bars represent the
1139 average viral RNA level in five pools of three insects. Error bars indicate the standard error of
1140 the mean. ** = $p < 0.01$, two-tailed T-test. (B-F) analysis of sRNA sequencing data of sRNAs
1141 purified from CRF-CA *D. citri* infected with wild-type CrPV or CrPV-DcDV by intrathoracic

1142 injection as described for (A). sRNA was purified from pools of 25 *D. citri* collected 5 days post
1143 injection. (B) Length distribution of sRNAs from wild-type CrPV-infected *D. citri* mapped to the
1144 wild-type CrPV genome. Red = antisense, blue = sense. Read counts are an average of three
1145 independent libraries. Error bars indicate standard deviation. (C) Length distribution of sRNAs
1146 from CrPV-DcDV-infected *D. citri* mapped to the CrPV-DcDV genome. Red = antisense, blue =
1147 sense. Read counts are an average of three independent libraries. Error bars indicate standard
1148 deviation. (D) Length distribution of sRNAs from CrPV-DcDV-infected *D. citri* mapped to the
1149 recombinant DcDV sequence present within the CrPV-DcDV genome. Red = antisense, blue =
1150 sense. Read counts are an average of three independent libraries. Error bars indicate standard
1151 deviation. (E) Length distribution of sRNAs from wild-type CrPV-infected *D. citri* mapped to
1152 the recombinant DcDV sequence present within the CrPV-DcDV genome. Red = antisense, blue
1153 = sense. Read counts are an average of three independent libraries. Error bars indicate standard
1154 deviation. (F & G) Sequence logos for 27-32 nt sRNAs from CrPV-DcDV-infected *D. citri*
1155 mapped to the recombinant DcDV sequence present within the CrPV-DcDV genome. Sequence
1156 logos for antisense sRNAs (F) or sense sRNAs (G) are shown. Data represents three pooled
1157 libraries. (H) Probability of overlap between the 5' ends of complementary 27-32 nt sRNAs
1158 mapping to opposite strands of the recombinant DcDV sequence present within the CrPV-DcDV
1159 genome during infection with CrPV-DcDV. Probabilities are shown for the indicated overlap
1160 distances and represent the average of three independent libraries. Error bars indicate standard
1161 deviation. The average Z-score and the standard deviation of the Z-score for an overlap length of
1162 10 nt is shown
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187

1188 **Fig. 7**
1189



1190
1191

1192 **Fig. 7.** CrPV-DcDV is not targeted by piRNAs during infection initiated by oral acquisition of
1193 purified virions. (A) Relative viral RNA levels during infection with wild-type CrPV or CrPV-
1194 DcDV. Infection was initiated in CRF-CA *D. citri* by oral acquisition. Insects were allowed to
1195 feed for 96 hours on an artificial diet solution containing 10^9 $TCID_{50}$ units/mL of wild-type
1196 CrPV or CrPV-DcDV. Following the feeding period insects were moved to *C. macrophylla*
1197 plants (day 0 post feeding). Viral RNA levels were assessed by RT-qPCR and normalized based
1198 on the expression of actin. The amount of viral RNA present on day 0 was set as 1 and viral
1199 RNA levels are shown relative to this value. Bars represent the average viral RNA level in seven
1200 pools of three insects. Error bars indicate the standard error of the mean. (B-F) analysis of sRNA

1201 sequencing data of sRNAs purified from CRF-CA *D. citri* infected with wild-type CrPV or
1202 CrPV-DcDV by oral acquisition as described for (A). sRNA was purified from pools of 25 *D.*
1203 *citri* collected 9 days post feeding. (B) Length distribution of sRNAs from wild-type CrPV-
1204 infected *D. citri* mapped to the wild-type CrPV genome. Red = antisense, blue = sense. Read
1205 counts are an average of three independent libraries. Error bars indicate standard deviation. (C)
1206 Length distribution of sRNAs from CrPV-DcDV-infected *D. citri* mapped to the CrPV-DcDV
1207 genome. Red = antisense, blue = sense. Read counts are an average of three independent
1208 libraries. Error bars indicate standard deviation. (D) Length distribution of sRNAs from CrPV-
1209 DcDV-infected *D. citri* mapped to the recombinant DcDV sequence present within the CrPV-
1210 DcDV genome. Red = antisense, blue = sense. Read counts are an average of three independent
1211 libraries. Error bars indicate standard deviation. (E) Length distribution of sRNAs from wild-
1212 type CrPV-infected *D. citri* mapped to the recombinant DcDV sequence present within the
1213 CrPV-DcDV genome. Red = antisense, blue = sense. Read counts are an average of three
1214 independent libraries. Error bars indicate standard deviation. (F & G) Sequence logos for 27-32
1215 nt sRNAs from CrPV-DcDV-infected *D. citri* mapped to the recombinant DcDV sequence
1216 present within the CrPV-DcDV genome. Sequence logos for antisense sRNAs (F) or sense
1217 sRNAs (G) are shown. Data represents three pooled libraries. (H) Probability of overlap between
1218 the 5' ends of complementary 27-32 nt sRNAs mapping to opposite strands of the recombinant
1219 DcDV sequence present within the CrPV-DcDV genome during infection with CrPV-DcDV.
1220 Probabilities are shown for the indicated overlap distances and represent the average of three
1221 independent libraries. Error bars indicate standard deviation. The average Z-score and the
1222 standard deviation of the Z-score for an overlap length of 10 nt is shown.
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237

Fig. S1

```
CRF-CA   AAAGCGATAA GTTGTCGTAT AAATAGTGGT GTTGGTAGCA TTGTAACTTT AGTATGGACT GTTCA [ 65]
Florida  AAAGCGATAA GTTGTCGTAT AAATAGTGGT GTTGGTAGCA TTGTAACTTT AGTATGGACT GTTCA [ 65]
Pakistan AAAGCGATAA GTTGTCGTAT AAATAGTGGT GTTGGTAGCA TTGTAACTTT AGTATGGACT GTTCA [ 65]
Arizona  AAAGCGATAA GTTGTCGTAT AAATAGTGGT GTTGGTAGCA TTGTAACTTT AGTATGGACT GTTCA [ 65]
CRF-HI   AAAGCGATAA GTTGTCGTAT AAATAGTGGT GTTGGTAGCA TTGTAACTTT AGTATGGACT GTTCA [ 65]

CRF-CA   GCAGCGAGAA AATTTGAACG TCTTGATTGC GTGTGTTATT TTTGTCTTTT ACCAAAATTG CCAAC [130]
Florida  GCAGCGAGAA AATTTGAACG TCTTGATTGC GTGTGTTATT TTTGTCTTTT ACCAAAATTG CCAAC [130]
Pakistan GCAGCGAGAA AATTTGAACG TCTTGATTGC GTGTGTTATT TTTGTCTTTT ACCAAAATTG CCAAC [130]
Arizona  GCAGCGAGAA AATTTGAACG TCTTGATTGC GTGTGTTATT TTTGTCTTTT ACCAAAATTG CCAAC [130]
CRF-HI   GCAGCGAGAA AATTTGAACG TCTTGATTGC GTGTGTTATT TTTGTCTTTT ACCAAAATTG CCAAC [130]

CRF-CA   ACCAGTAGTG AATTTAGTTC CGACGAGGAT GAGTTCCGAC GAGGATTCGT GTCACTCGGT GATCG [195]
Florida  ACCAGTAGTG AATTTAGTTC CGACGAGGAT GAGTTCCGAC GAGGATTCGT GTCACTCGGT GATCG [195]
Pakistan ACCAGTAGTG AATTTAGTTC CGACGAGGAT GAGTTCCGAC GAGGATTCGT GTCACTCGGT GATCG [195]
Arizona  ACCAGTAGTG AATTTAGTTC CGACGAGGAT GAGTTCCAAC GAGGATTCGT GTCACTCGGT GATCG [195]
CRF-HI   ACCAGTAGTG AATTTAGTTC CGACGAGGAT GAGTTCCGAC GAGGATTCGT GTCACTCGGT GATCG [195]

CRF-CA   AGGAGGATCC TCTGAGTTCT GCTCCAGCGA GTCCGGCGAC GAGTCAACAT GGTCACAACT TAAGC [260]
Florida  AGGAGGATCC TCTGAGTTCT GCTCCAGCGA GTCCGGCGAC GAGTCAACAT GGTCACAACT TAAGC [260]
Pakistan AGGAGGATCC TCTGAGTTCT GCTCCAGCGA GTCCGGCGAC GAGTCAACAT GGTCACAACT TAAGC [260]
Arizona  AGGAGGATCC TCTGAGTTCT GCTCCAGCGA GTCCGGCGAC GAGTCAACAT GGTCACAACT TAAGC [260]
CRF-HI   AGGAGGATCC TCTGAGTTCT GCTCCAGCGA GTCCGGCGAC GAGTCAACAT GGTCACAACT TAAGC [260]

CRF-CA   GATTCCGAGG AGACGAACTG TACCGTCGTG TTGGCACAGA ATCCCCTGCC CGACACACAA GAAGT [325]
Florida  GATTCCGAGG AGACGAACTG TACCGTCGTG TTGGCACAGA ATCCCCTGCC CGACACACAA GAAGT [325]
Pakistan GATTCCGAGG AGACGAACTG TACCGTCGTG TTGGCACAGA ATCCCCTGCC CGACACACAA GAAGT [325]
Arizona  GATTCCGAGG AGACGAACTG TACCGTCGTG TTGGCACAGA ATCCCCTGCC CGACACACAA GAAGT [325]
CRF-HI   GATTCCGAGG AGACGAACTG TACCGTCGTG TTGGCACAGA ATCCCCTGCC CGACACACAA GAAGT [325]

CRF-CA   CGAATCGATG TTTCAAGAAG ACCGGAGCGC ATCAGTACAT TTCGCAGACG ATCCTCCGAC GGGAG [390]
Florida  CGAATCGATG TTTCAAGAAG ACCGGAGCGC ATCAGTACAT TTCGCAGACG ATCCTCCGAC GGGAG [390]
Pakistan CGAATCGATG TTTCAAGAAG ACCGGAGCGC ATCAGTACAT TTCGCAGACG ATCCTCCGAC GGGAG [390]
Arizona  CGAATCGATG TTTCAAGAAG ACCGGAGCGC ATCAGTACAT TTCGCAGACG ATCCTCCGAC GGGAG [390]
CRF-HI   CGAATCGATG TTTCAAGAAG ACCGGAGCGC ATCAGTACAT TTCGCAGACG ATCCTCCGAC GGGAG [390]

CRF-CA   CCCGGAAGAA GAAATCTATT CTGAAACGGG CGACGAAAAA GAAAACGACT CCCAAAGCGA ACACA [455]
Florida  CCCGGAAGAA GAAATCTATT CTGAAACGGG CGACGAAAAA GAAAACGACT CCCAAAGCGA ACACA [455]
Pakistan CCCGGAAGAA GAAATCTATT CTGAAACGGG CGACGAAAAA GAAAACGACT CCCAAAGCGA ACACA [455]
Arizona  CCCGGAAGAA GAAATCTATT CTGAAACGGG CGACGAAAAA GAAAACGACT CCCAAAGCGA ACACA [455]
CRF-HI   CCCGGAAGAA GAAATCTATT CTGAAACGGG CGACGAAAAA GAAAACGACT CCCAAAGCGA ACACA [455]

CRF-CA   GCGAAGAAGA TGCTCACTCT GACTCCTCAG CAACAGACAT CAGTGTGGGA ACGGCAAACC AACAA [520]
Florida  GCGAAGAAGA TGCTCACTCT GACTCCTCAG CAACAGACAT CAGTGTGGGA ACGGCAAACC AACAA [520]
Pakistan GCGAAGAAGA TGCTCACTCT GACTCCTCAG CAACAGACAT CAGTGTGGGA ACGGCAAACC AACAA [520]
Arizona  GCGAAGAAGA TGCTCACTCT GACTCCT-AG CAACAGACAT CAGTGTGGGA ACGGCAAACC AACAA [520]
CRF-HI   GCGAAGAAGA TGCTCACTCT GACTCCTCAG CAACAGACAT CAGTGTGGGA ACGGCAAACC AACAA [520]

CRF-CA   CTTTATGGCA TTCTTGTTCC TCGTTTGAAC GCCCCTGCAG ACCGTTCACC TTCTCCAGGA CCTTC [585]
Florida  CTTTATGGCA TTCTTGTTCC TCGTTTGAAC GCCCCTGCAG ACCGTTCACC TTCTCCAGGA CCTTC [585]
Pakistan CTTTATGGCA TTCTTGTTCC TCGTTTGAAC GCCCCTGCAG ACCGTTCACC TTCTCCAGGA CCTTC [585]
Arizona  CTTTATGGCA TTCTTGTTCC TCGTTTGAAC GCCCCTGCAG ACCGTTCACC TTCTCCAGGA CCTTC [585]
CRF-HI   CTTTATGGCA TTCTTGTTCC TCGTTTGAAC GCCCCTGCAG ACCGTTCACC TTCTCCAGGA CCTTC [585]

CRF-CA   TACTGCATAT CGCTATCGTG GCGAGGAAAA CCCAAAATAT GGACCTAAAA AA-GCAGTGC ATCCG [650]
Florida  TACTGCATAT CGCTATCGTG GCGAGGAAAA CCCAAAATAT GGACCTAAAA AA-GCAGTGC ATCCG [650]
Pakistan TACTGCATAT CGCTATCGTG GCGAGGAAAA CCCAAAATAT GGACCTAAAA AA-GCAGTGC ATCCG [650]
Arizona  TACTGCATAT CGCTATCGTG GCGAGGAAAA CCCAAAATGT GGACCTAAAA AAACAAGGGA TTCCG [650]
CRF-HI   TACTGCATAT CGCTATCGTG GCGAGGAAAA CCCAAAATGT GGACCTAAAA AAACCATGGC TTCCG [650]

CRF-CA   TTGGGGGGAA GGCTTAGCGG T [671]
Florida  TTGGGGGGAA GGCTTAGCGG T [671]
Pakistan TTGGGGGGAA GGCTTAGCGG T [671]
Arizona  TGGGGGGAA- ---------- - [671]
CRF-HI   TTGGGGGAA- ---------- - [671]
```

Fig. S1. Alignment of the sequences of the ENS PCR products shown in Fig. 2. Sequences were aligned using ClustalW and mismatched nucleotides are shaded.

Fig. S2

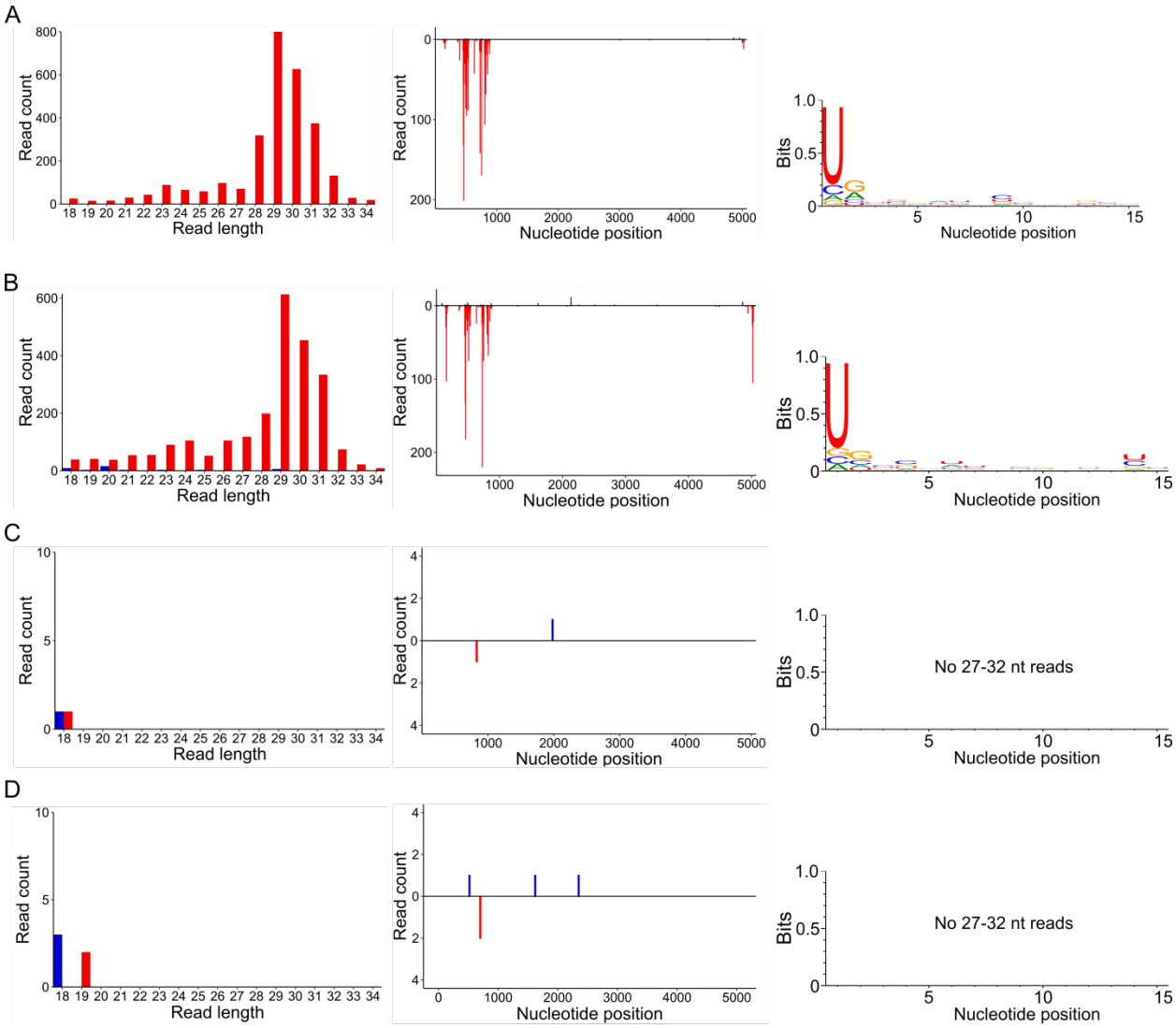


Fig. S2. (A-D) Left: Length distribution of all sRNAs from various *D. citri* populations mapped to the DcDV genome. Red = antisense, blue = sense. Middle: Positions of all sRNAs from various *D. citri* populations mapped to the DcDV genome. sRNAs are shown as mapped to the genomic strand containing the coding sequence for the NS proteins. Red = antisense sRNAs, blue = sense sRNAs. Right: Sequence logo of 27-32 nt sRNAs shown in the left and middle panels. sRNAs were purified from field-collected insects from China (GenBank accession no. SRX1164134)(A), the US state of Florida (GenBank accession no. SRX1164135) (B), Brazil (GenBank accession no. SRX1164127)(C), or laboratory reared insects from CRF-Uru (D). Sequence data for (A-C) are described in (25,26).

Fig. S3

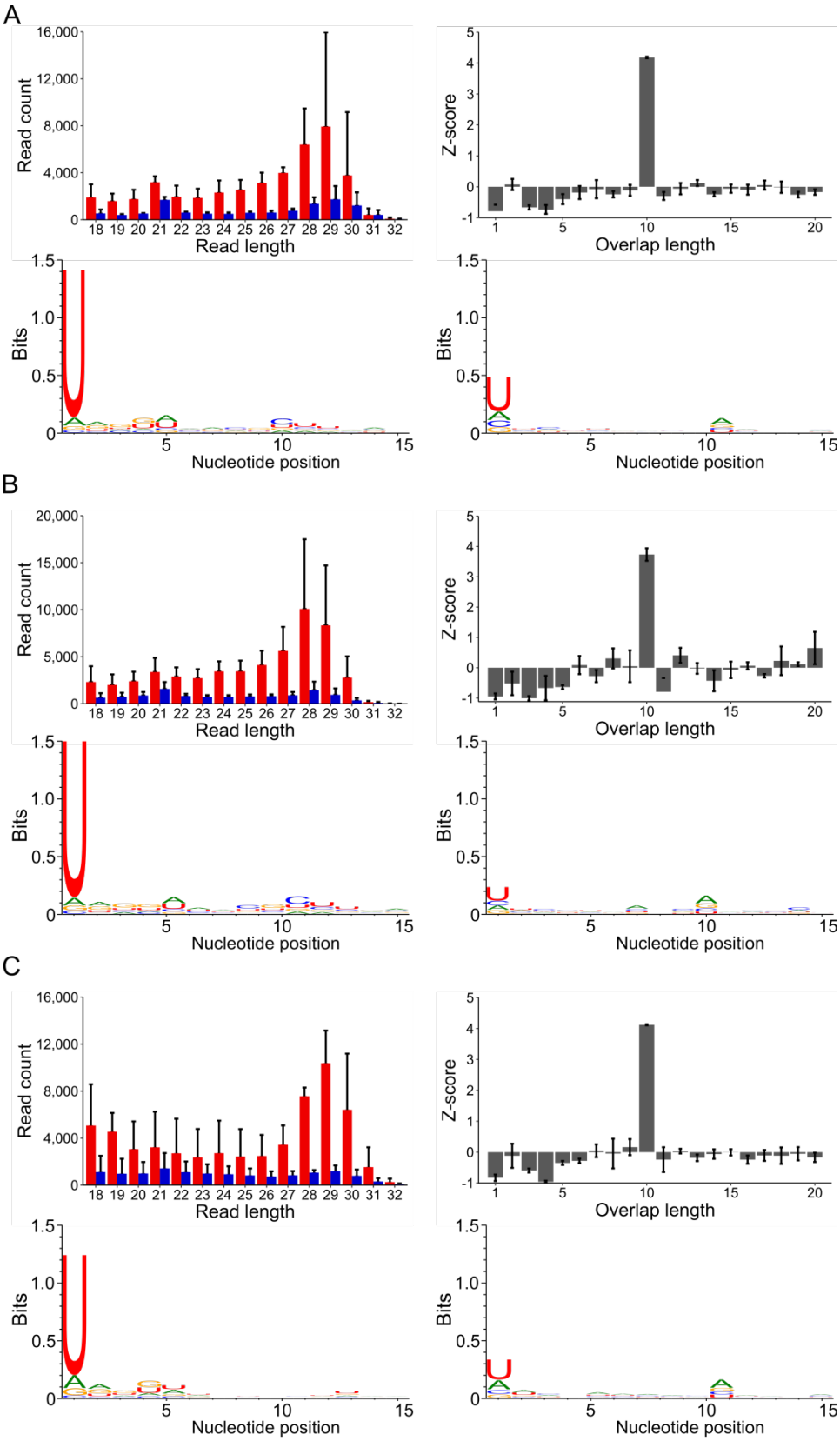

A
Read count
Read length
Z-score
Overlap length
Bits
Nucleotide position
B
Read count
Read length
Z-score
Overlap length
Bits
Nucleotide position
C
Read count
Read length
Z-score
Overlap length
Bits
Nucleotide position

Fig. S3 continued

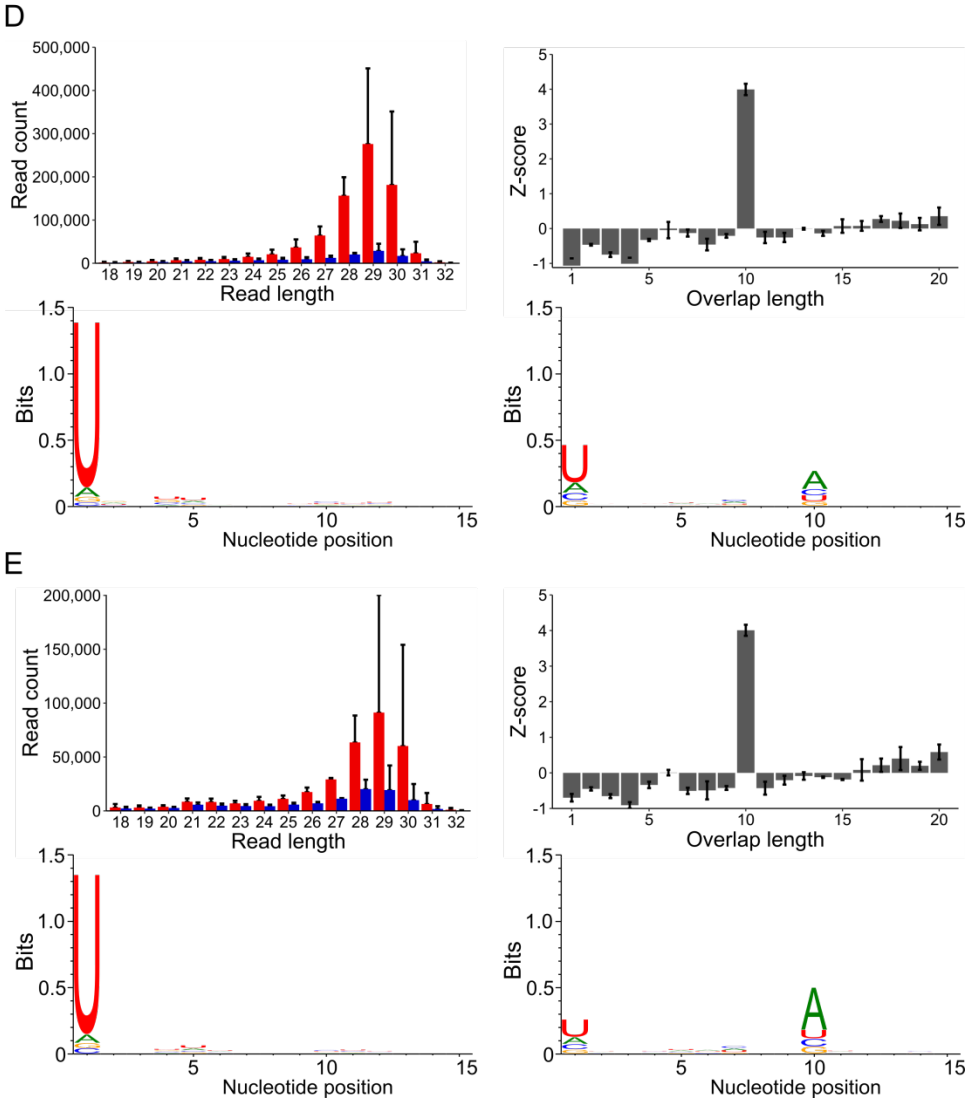


Fig. S3. (A-E) Analysis of sRNAs from dissected CRF-CA *D. citri* tissues. Figures indicate mapping statistics for sRNAs mapping to all TEs identified in the *D. citri* genome. (A) sRNAs from three pools of 200 *D. citri* guts. (B) sRNAs from three pools of 200 *D. citri* heads. (C) sRNAs from three pools of hemolymph from 100 *D. citri* each. (D) sRNAs from three pools of 100 *D. citri* ovaries. (E) sRNAs from three pools of 100 *D. citri* testis. Upper left panels: length distribution of sRNAs mapping to TEs. Red = antisense, blue = sense. Read counts are an average of three independent libraries. Error bars indicate standard deviation. Upper right panels: Z-scores for the indicated overlap distances between the 5' ends of complementary 27-32 nt sRNAs. Z-scores are an average of three independent libraries. Errors bars indicate the standard deviation. Lower panels: Sequence logos for 27-32 nt sRNAs mapping to TEs. Lower left = antisense, Lower right = sense. Data represents three pooled libraries.

Fig. S4

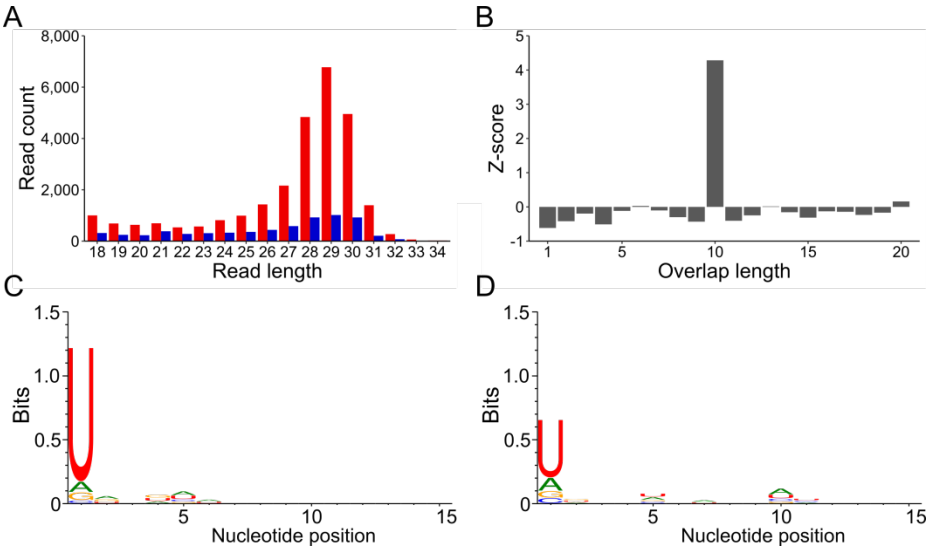


Fig. S3. (A-E) Analysis of sRNAs from CRF-Uru *D. citri*. Figures indicate mapping statistics for sRNAs mapping to all TEs identified in the *D. citri* genome. (A) length distribution of sRNAs mapping to TEs. Red = antisense, blue = sense. (B) Z-scores for the indicated overlap distances between the 5' ends of complementary 27-32 nt sRNAs. (C & D) Sequence logos for 27-32 nt sRNAs mapping to TEs. (C) Antisense sRNAs. (D) Sense sRNAs.

Fig. S5

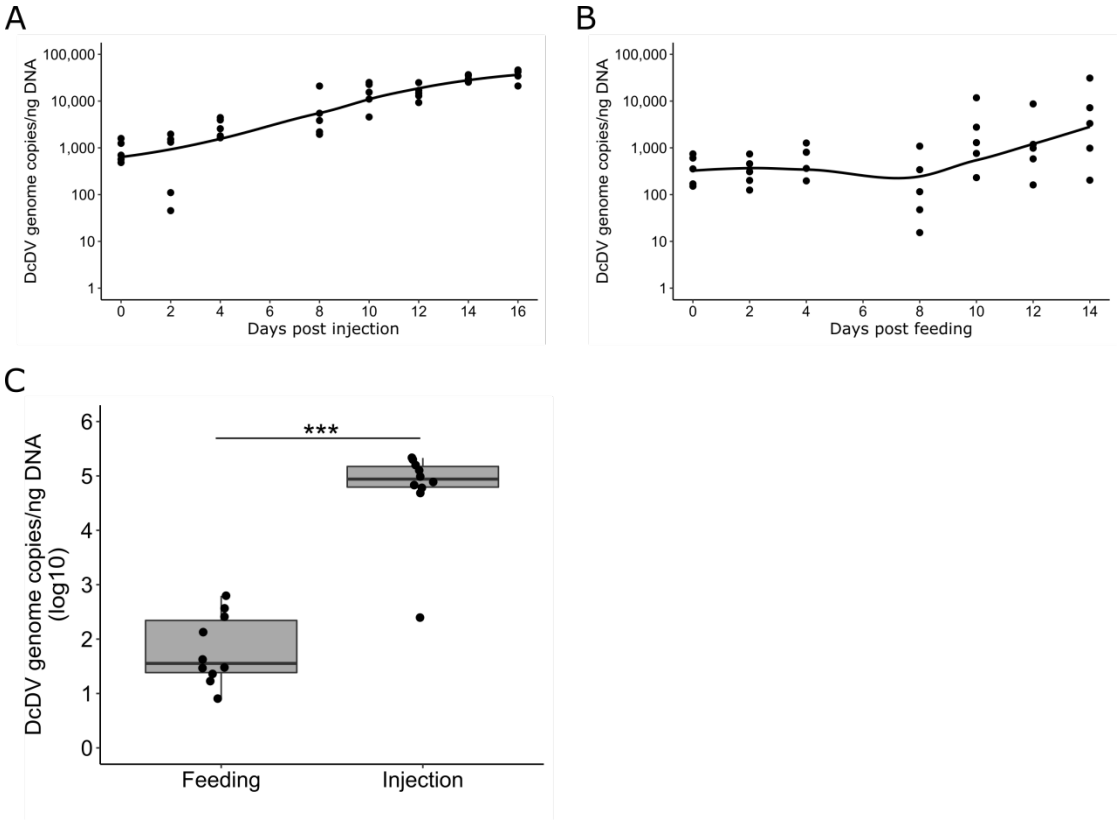


Fig. S4. (A) Titer of DcDV following intrathoracic injection of virions into uninfected *D. citri* insects from CRF-Uru. Each data point represents three pooled insects (n=5). The curve was smoothed using loess regression. (B) Titer of DcDV following feeding of virions to uninfected *D. citri* insects from CRF-Uru. Virions present in a 15% sucrose artificial diet solution were acquired by *D. citri* for 96 hours by membrane feeding. After the feeding period, insects were moved to *Citrus macrophylla* plants and this represents day 0 post feeding. Each data point represents three pooled insects (n=5). The curve was smoothed using loess regression. (C) Titer of DcDV in F1 progeny of the insects shown in (A), designated by injection, and (B), designated by Feeding. n=10 individual insects. Individual insects were collected on the day of emergence to adulthood.

Fig. S6

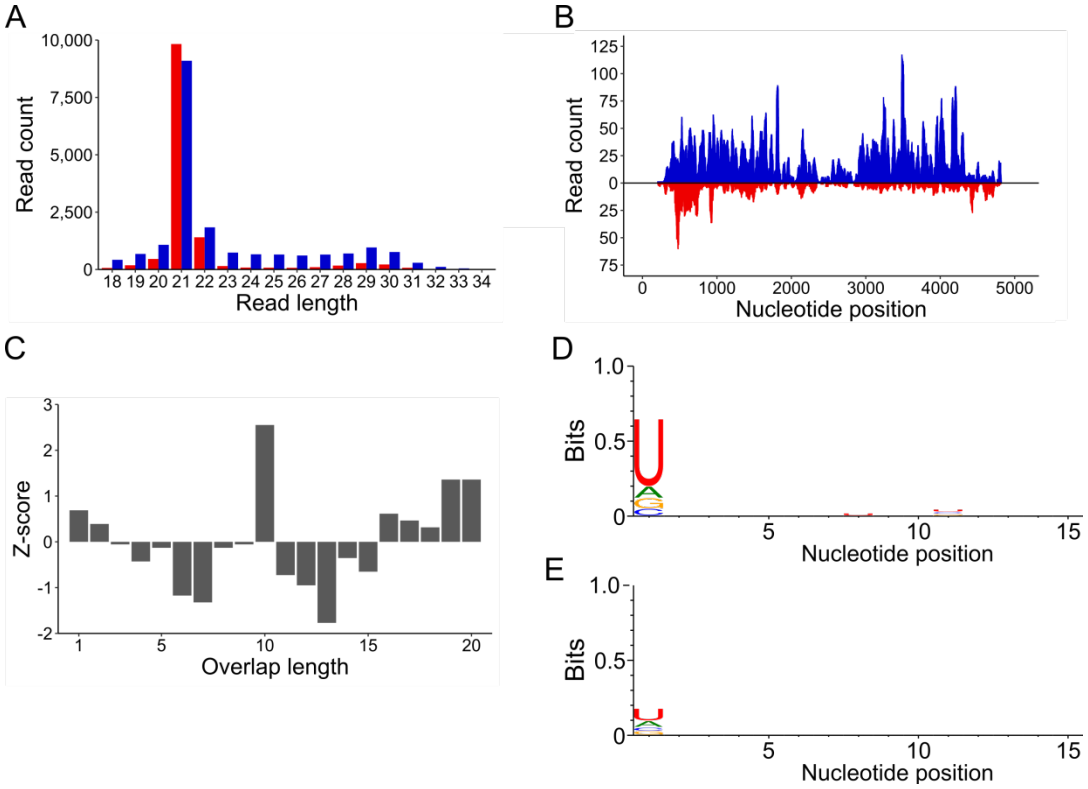
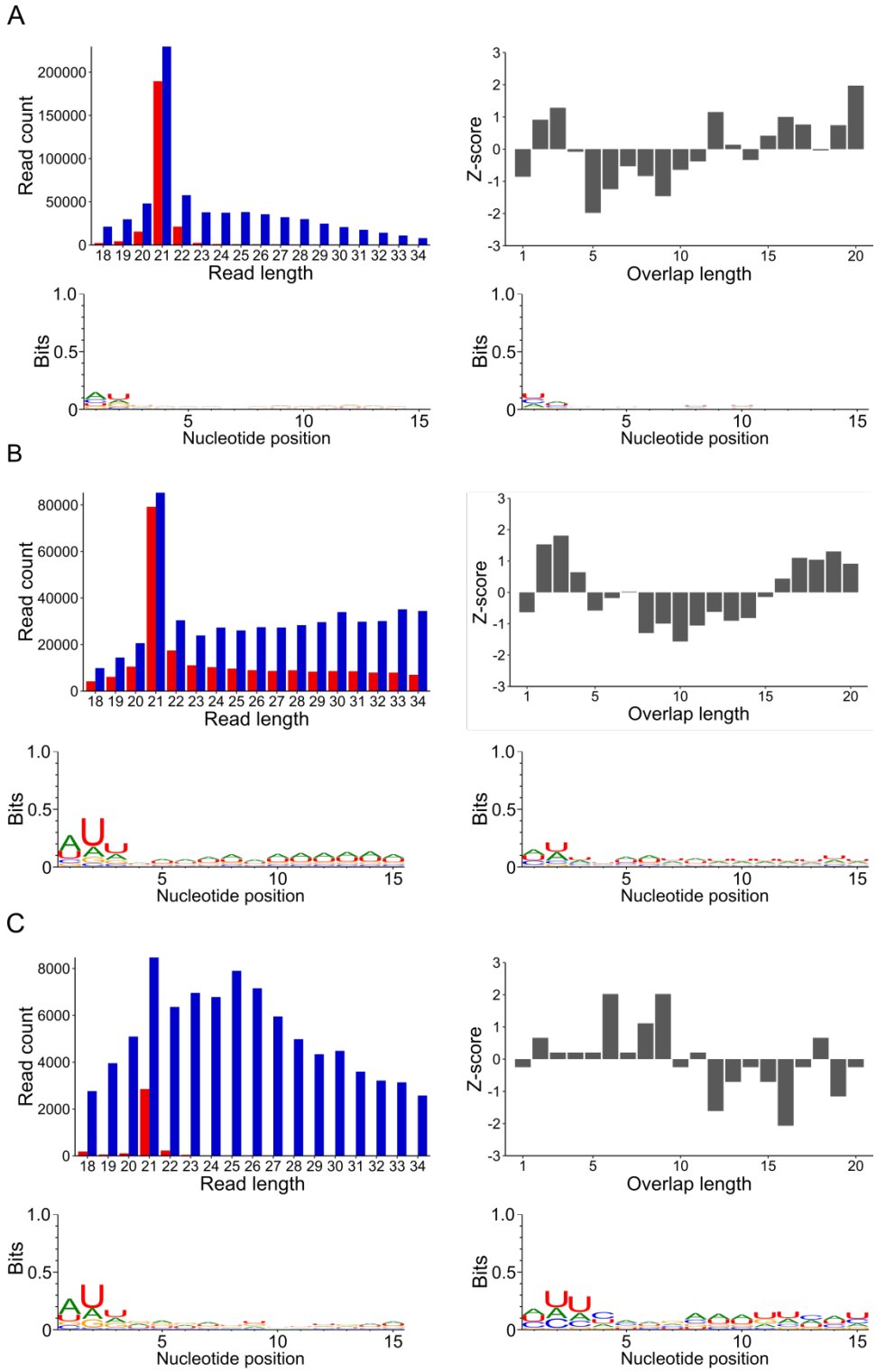


Fig. S6. DcDV is targeted by ping-pong-dependent vpiRNAs in DcDV-infected *D. citri* insects from CRF-Uru. sRNAs were purified from the progeny of CRF-Uru *D. citri* that had been infected with DcDV by intrathoracic injection (see Fig. S5a & c). (A) Length distribution of sRNAs mapping to the DcDV genome in DcDV-infected *D. citri* insects from CRF-Uru. To account for the bidirectional transcription strategy of DcDV, sRNA mapping polarity was assigned from mapping location based on the start and stop positions of the canonical DcDV transcripts. Red = antisense sRNAs, blue = sense sRNAs. (B) Positions of 27-32 nt sRNAs shown in (A). Red = antisense sRNAs, blue = sense sRNAs. (C) Z-scores for the indicated overlap distances between the 5' ends of complementary 27-32 nt sRNAs shown in (A). (D & E) Sequence logos for the 27-32 nt sRNAs shown in (A). Sequence logos for antisense sRNAs (D) or sense sRNAs (E) are shown.

1367 **Fig. S7**

1368



1369
1370
1371
1372

Fig. S7 continued

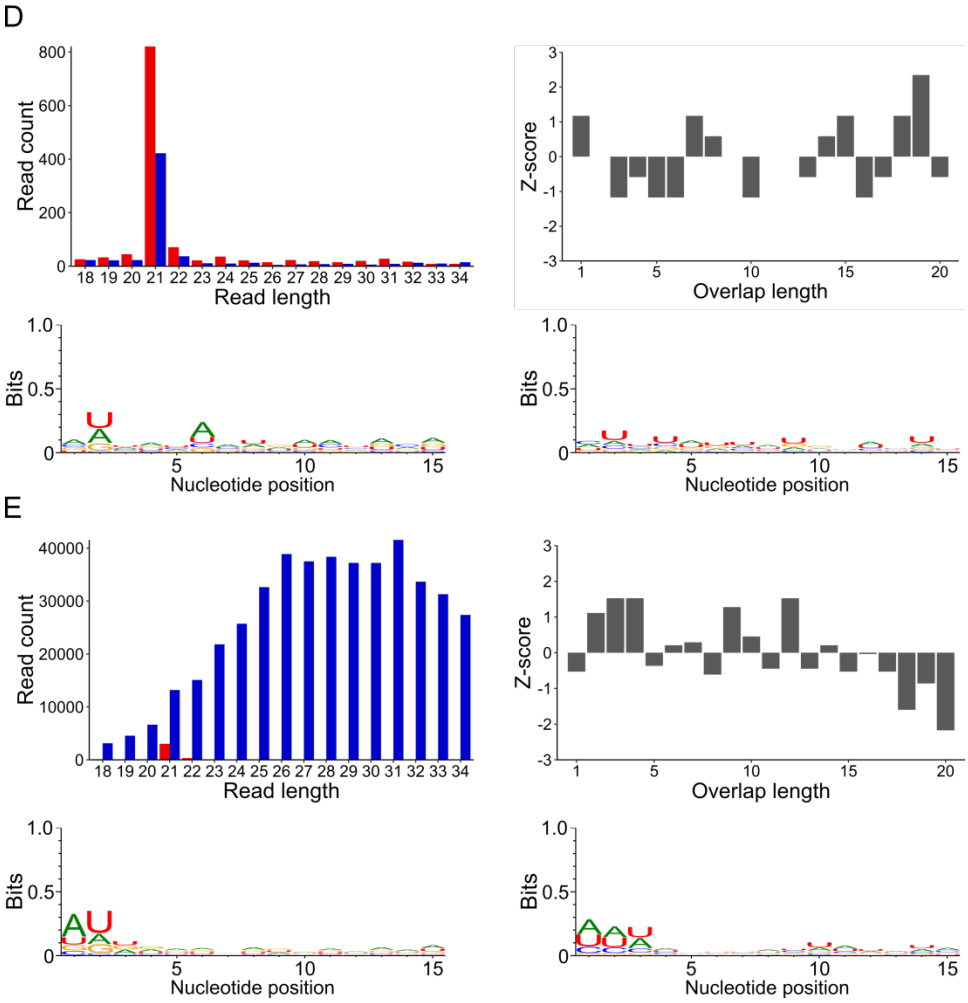


Fig. S7. (A-E) Analysis of sRNAs from various field-collected *D. citri* populations infected with one or more *D. citri*-specific viruses. (A) *D. citri* from the US state of Florida infected with Diaphorina citri flavi-like virus (sRNA GenBank accession no. SRX1164135, viral genome GenBank accession no. NC_030453.1). (B) *D. citri* from China infected with Diaphorina citri reovirus (sRNA GenBank accession no. SRX1164134, viral genome GenBank accession no.'s KT698830.1, KT698831.1, KT698832.1, KT698833.1, KT698834.1, KT698836.1). (C) *D. citri* from China infected with Diaphorina citri picorna-like virus (sRNA GenBank accession no. SRX1164134, viral genome GenBank accession no. KT698837.1). (D) *D. citri* from China infected with Diaphorina citri bunyavirus (sRNA GenBank accession no. SRX1164134, viral genome GenBank accession no.'s KT698825.1, KT698824.1, KT698823.1). (E) *D. citri* from Brazil infected with Diaphorina citri picorna-like virus (sRNA GenBank accession no. SRX1164127, viral genome GenBank accession no. KT698837.1). Upper left panels: length distribution of sRNAs mapped to the various viral genomes. Red = antisense, blue = sense. Upper right panels: Z-scores for the indicated overlap distances between the 5' ends of complementary 27-32 nt sRNAs mapping to opposite strands of the various viral genomes.

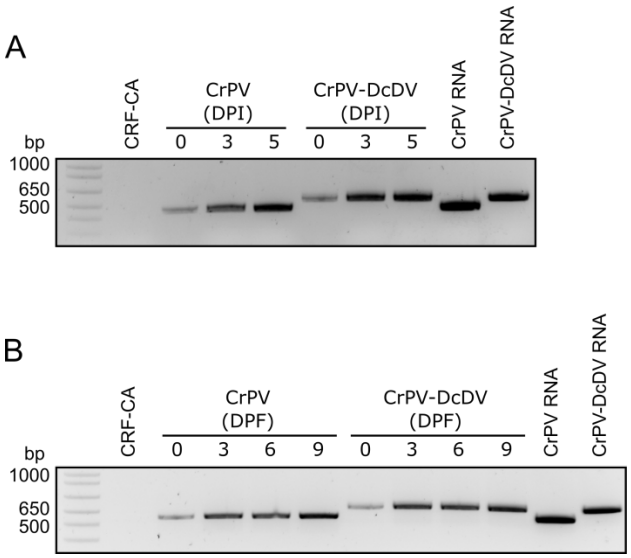
Sequence data are described in (25,26). Lower panels: Sequence logos for 27-32 nt sRNAs mapping to the various viral genomes. Lower left = antisense, Lower right = sense.

Fig. S8

```
CrPV-DcDNV (1352-1473)   CTTTTGATGA GTGGTgatgt cgaatctaat cctggtcctg CAGACCGTTC ACCTTCTCCA   [ 60]
DcDNV (803-859)          ---------- ---------- ---------- ---------g CAGACCGTTC ACCTTCTCCA   [ 60]
NW_007380266 (8062-8118) ---------- ---------- ---------- ---------g CAGACCGTTC ACCTTCTCCA   [ 60]

CrPV (1171-1224)         CTTTTGATGA GTGGTgatgt cgaatctaat cctggtcctg ---------- ----------   [ 60]

CrPV-DcDNV (1352-1473)   GGACCTTCTA CTGCATATCG CTATTGTAGC GAGGAAgatg tcgaatctaa tcctggtcct  [ 120]
DcDNV (803-859)          GGACCTTCTA CTGCATATCG CTATTGTAGC GAGGAA---- ---------- ----------  [ 120]
NW_007380266 (8062-8118) GGACCTTCTA CTGCATATCG CTATCGTGGC GAGGAA---- ---------- ----------  [ 120]

CrPV (1171-1224)         ---------- ---------- ---------- ---------- ---------- ----------  [ 120]

CrPV-DcDNV (1171-1473)   gTGCAATCGC GCCCC [135]
DcDNV (803-859)          ---------- ----- [135]
NW_007380266 (8062-8118) ---------- ----- [135]
CrPV (1171-1224)         -TGCAATCGC GCCCC [135]
```

Fig. 8. Alignment of a portion of the CrPV-DcDV genome, the recombinant DcDV sequence present within CrPV-DcDV, the region of ENS corresponding to the recombinant DcDV sequence present within CrPV-DcDV (represented by the GenBank accession no. of *D. citri* genomic scaffold 2850, NW_007380266), and a portion of the wild-type CrPV sequence. Numbers in parentheses indicate the nucleotide positions of the sequence shown. The CrPV 1A cleavage site is shown in lowercase letters. The recombinant DcDV sequence present within CrPV-DcDV and the corresponding region of ENS are shown in italics.

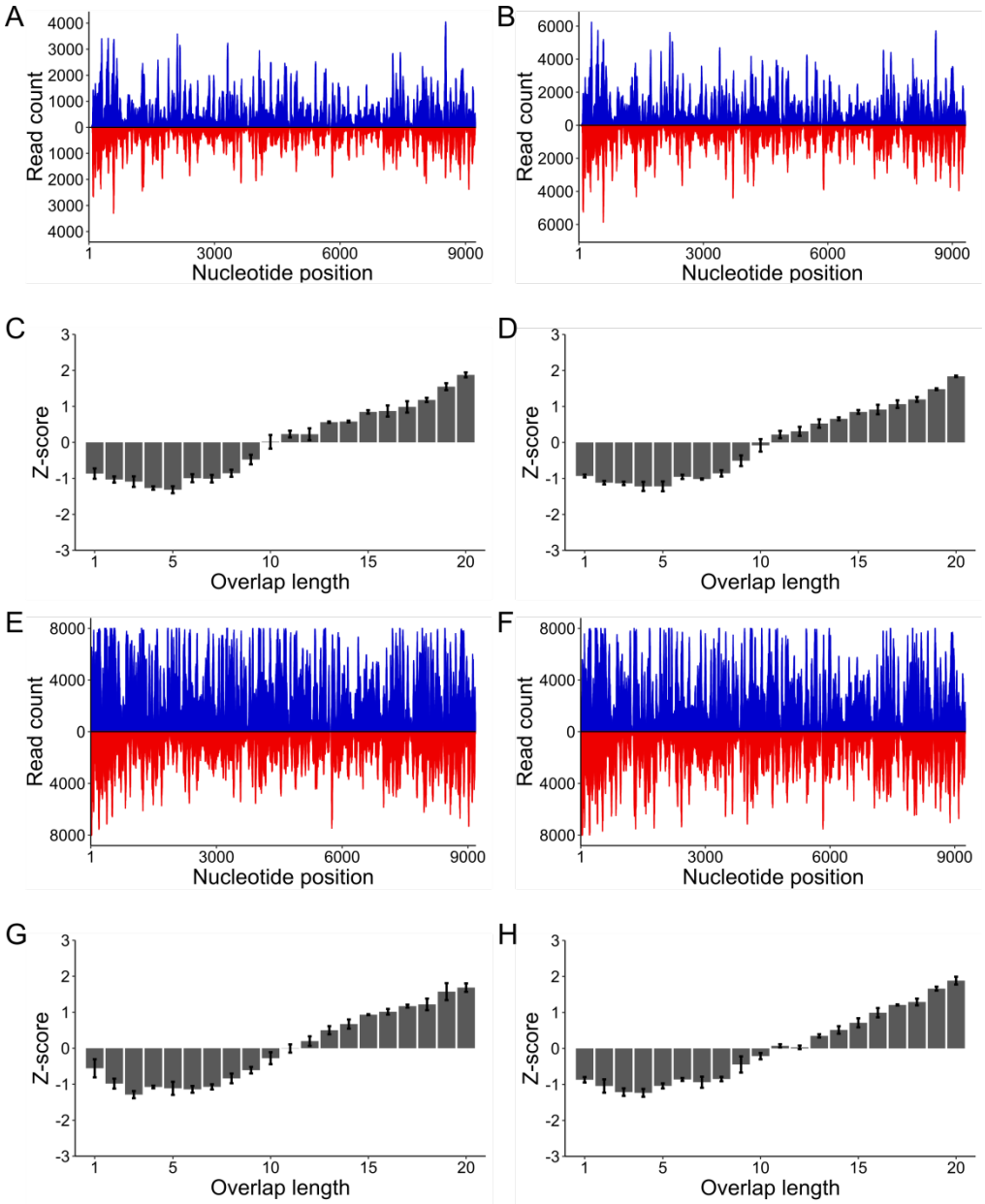
1477 **Fig. S9**
1478



1479
1480
1481 **Fig. S9.** RT-PCR products produced using primers flanking the site into which recombinant
1482 DcDV sequence was inserted in CrPV-DcDV (primers 15 and 16). RNA from all biological
1483 replicates from each day shown in Fig. 4.6a (A) or Fig. 4.7a (B) was pooled and used as
1484 templates for RT-PCR. RNA from CRF-CA *D. citri* was used as a negative control. *In vitro*
1485 transcribed wild-type CrPV or CrPV-DcDV RNA was used as a positive control. DPI = days
1486 post injection. DPF = days post feeding.
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509

1510 **Fig. S10**

1511



1512

1513 **Fig. S10.** (A & B) Positions of 21 nt sRNAs mapping to the wild-type CrPV genome (A) or the
1514 CrPV-DcDV genome (B) during infection of CRF-CA *D. citri* initiated by intrathoracic injection
1515 of virions. Red = antisense sRNAs, blue = sense sRNAs. Reads counts are an average of three
1516 independent libraries. (C & D) Z-scores for the indicated overlap distances between the 5' ends
1517 of complementary 27-32 nt sRNAs mapping to opposite strands of the wild-type CrPV genome
1518 (C) or the non-recombinant portion of the CrPV-DcDV genome (D) during infection of CRF-CA
1519 *D. citri* initiated by intrathoracic injection of virions. Z-scores represent the average of three
1520 independent libraries. Error bars indicate standard deviation. (E & F) Positions of 21 nt sRNAs
1521 mapping to the wild-type CrPV genome (E) or the CrPV-DcDV genome (F) during infection of
1522 CRF-CA *D. citri* initiated by oral acquisition of virions. Red = antisense sRNAs, blue = sense
1523 sRNAs. Reads counts are an average of three independent libraries. (G & H) Z-scores for the

1524 indicated overlap distances between the 5' ends of complementary 27-32 nt sRNAs mapping to
1525 opposite strands of the wild-type CrPV genome (G) or the non-recombinant portion of the CrPV-
1526 DcDV genome (H) during infection of CRF-CA *D. citri* initiated by oral acquisition of virions.
1527 Z-scores represent the average of three independent libraries. Error bars indicate standard
1528 deviation.
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569

1570 **Table S1**

Primer #	Sequence (5’ → 3’)
1	GGGAAAATCGGTCTCATGCTGCTGTT
2	AAATTGGAGTTAGAAGCTAAAGTTACC
3	TGTAACTTTAGTATGGACTGTTCAGC
4	TTCTTTCACTCCAAACTCTTTCTAGAATGC
5	CCTCTGAGTTCTGCTCCAGC
6	AACAGCAGCATGAGACCGATTTTCCC
7	TGACACCGCTAAGCCTTCC
8	ACACTTTTTTAAAAGCGATAAGTTGTC
9	CCCTGGACTTTGAACAGGAA
10	CATTTGCGGTGAACGATTCC
11	CCAGAACATCGATATGGATCAAC
12	TTGAATTTGGGCTGTTAGTGTC
13	GTGCAATCGCGCCCCGTATATG
14	AGGACCAGGATTAGATTCGACATCAC
15	CCAGAACATCGATATGGATCAAC
16	/phos/TTGAATTTGGGCTGTTAGTGTC†
17	/phos/AGATTCGACATCTTCCTCGCTACAATAGCG†
18	AATCCTGGTCCTGTGCAATCGCGCC
19	GGAAGAATGGCGAATTTTGA
20	CTCGGGAAAGATCCATTTCA
21	TCGTGACATCAAGGAGAAGCTGTGC
22	TTGACCGTCGGGAAGTTCGTAGGAT
23	TATAAATAGAGGCGTTTCTGTTGTG
24	TAATACGACTCACTATAGGGCGATCGCCAAGCCTTCC
25	TAATACGACTCACTATAGGGTATAAATAGAGGCGTTTCTGTTGTG
26	CGATCGCCAAGCCTTCC
27	TAATACGACTCACTATAGGGTGACACCGCTAAGCCTTCC

1571 † /phos/ indicates 5’ phosphorylation
1572
1573
1574
1575