A peer-reviewed version of this preprint was published in PeerJ on 11 October 2018.

View the peer-reviewed version (peerj.com/articles/5761), which is the preferred citable publication unless you specifically need to cite this preprint.

Virus discovery in all three major lineages of terrestrial arthropods highlights the diversity of single-stranded DNA viruses associated with invertebrates

Karyna Rosario Corresp., 1, Kaitlin A Mettel 1, Bayleigh E Benner 1, Ryan Johnson 1, Catherine Scott 2, So hath Z Yusseff-Vanegas 3, Christopher CM Baker 4,5, Deby L Cassill 6, Caroline Storer 7, Arvind Varsani 8, 9, Mya Breitbart 1

1 College of Marine Science, University of South Florida, Saint Petersburg, Florida, United States
2 Department of Biological Sciences, University of Toronto, Scarborough, Scarborough, Ontario, Canada
3 Department of Biology, University of Vermont, Burlington, Vermont, United States
4 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, United States
5 Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States
6 Department of Biological Sciences, University of South Florida Saint Petersburg, Saint Petersburg, Florida, United States
7 School of Forest Resources and Conservation, University of Florida, Gainesville, Florida, United States
8 The Biodesign Center for Fundamental and Applied Microbiomics, School of Life Sciences, Center for Evolution and Medicine, Arizona State University, Tempe, Arizona, United States
9 Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, South Africa

Corresponding Author: Karyna Rosario
Email address: krosari2@mail.usf.edu

Viruses encoding a replication-associated protein (Rep) within a covalently closed, single-stranded (ss)DNA genome are among the smallest viruses known to infect eukaryotic organisms, including economically valuable agricultural crops and livestock. Although circular Rep-encoding ssDNA (CRESS DNA) viruses are a widespread group for which our knowledge is rapidly expanding, biased sampling towards vertebrates and land plants has limited our understanding of their diversity and evolution. Here we screened terrestrial arthropods for CRESS DNA viruses and report the identification of 44 viral genomes and replicons associated with specimens representing all three major terrestrial arthropod lineages, namely Euchelicerata (spiders), Hexapoda (insects), and Myriapoda (millipedes). We identified virus genomes belonging to three established CRESS DNA viral families (Circoviridae, Genomoviridae, and Smacoviridae); however, over half of the arthropod-associated virus genomes are only distantly related to currently classified CRESS DNA viral sequences. Although members of viral and satellite families known to infect plants (Geminiviridae, Nanoviridae, Alphasatellitidae) were not identified in this study, these plant-infecting CRESS DNA viruses and replicons are transmitted by hemipterans. Therefore, members from six out of the seven established CRESS DNA viral families circulate among arthropods. Furthermore, a phylogenetic analysis of Reps, including endogenous viral sequences, reported to date from a wide array of organisms revealed...
that most of the known CRESS DNA viral diversity circulates among invertebrates. Our results highlight the vast and unexplored diversity of CRESS DNA viruses among invertebrates and parallel findings from RNA viral discovery efforts in undersampled taxa.
Virus discovery in all three major lineages of terrestrial arthropods highlights the diversity of single-stranded DNA viruses associated with invertebrates

Karyna Rosario1*, Kaitlin A. Mettel1, Bayleigh E. Benner1, Ryan Johnson1, Catherine Scott3, Sohath Z. Yusseff-Vanegas3, Christopher C. M. Baker4,5, Deby L. Cassill6, Caroline Storer7, Arvind Varsani8,9 and Mya Breitbart1

1College of Marine Science, University of South Florida, St. Petersburg, FL, United States of America
2Department of Biological Sciences, University of Toronto Scarborough, Scarborough, Ontario, Canada
3Department of Biology, University of Vermont, Burlington, Vermont, United States of America
4Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, United States of America
5Present address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America
6Department of Biological Sciences, University of South Florida Saint Petersburg, St. Petersburg, FL, United States of America
7School of Forest Resources and Conservation, University of Florida, Gainesville, FL, United States of America
8The Biodesign Center for Fundamental and Applied Microbiomics, School of Life Sciences, Center for Evolution and Medicine, Arizona State University, Tempe, Arizona, United States of America
9Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, South Africa

*Corresponding Author:
Karyna Rosario
Email: krosari2@mail.usf.edu
Abstract

Viruses encoding a replication-associated protein (Rep) within a covalently closed, single-stranded (ss)DNA genome are among the smallest viruses known to infect eukaryotic organisms, including economically valuable agricultural crops and livestock. Although circular Rep-encoding ssDNA (CRESS DNA) viruses are a widespread group for which our knowledge is rapidly expanding, biased sampling towards vertebrates and land plants has limited our understanding of their diversity and evolution. Here we screened terrestrial arthropods for CRESS DNA viruses and report the identification of 44 viral genomes and replicons associated with specimens representing all three major terrestrial arthropod lineages, namely Euchelicerata (spiders), Hexapoda (insects), and Myriapoda (millipedes). We identified virus genomes belonging to three established CRESS DNA viral families (Circoviridae, Genomoviridae, and Smacoviridae); however, over half of the arthropod-associated virus genomes are only distantly related to currently classified CRESS DNA viral sequences. Although members of viral and satellite families known to infect plants (Geminiviridae, Nanoviridae, Alphasatellitidae) were not identified in this study, these plant-infecting CRESS DNA viruses and replicons are transmitted by hemipterans. Therefore, members from six out of the seven established CRESS DNA viral families circulate among arthropods. Furthermore, a phylogenetic analysis of Reps, including endogenous viral sequences, reported to date from a wide array of organisms revealed that most of the known CRESS DNA viral diversity circulates among invertebrates. Our results highlight the vast and unexplored diversity of CRESS DNA viruses among invertebrates and parallel findings from RNA viral discovery efforts in undersampled taxa.
Introduction

Virus discovery remains an open-ended endeavor with estimates of more than 99% of viruses within organisms remaining to be sampled (Geoghegan and Holmes 2017). Eukaryotic viruses infecting vertebrates, mainly mammals, and land plants are overrepresented in public databases relative to those infecting invertebrates and unicellular organisms (Mahmoudabadi and Phillips 2018). Therefore, biased sampling has heavily skewed our view of viral diversity and evolution and there is a need to explore ‘non-traditional’ organisms. Efforts investigating single-stranded (ss)RNA viruses in undersampled taxa have identified arthropods, the most diverse and successful group of animals on Earth (Stork 2018), as a rich and untapped reservoir of novel viruses (Shi, Lin, Tian, et al. 2016; Li et al. 2015). Moreover, discovery of divergent viruses in invertebrates has prompted reevaluation of RNA virus evolution concepts and taxonomic frameworks (Shi, Lin, Vasilakis, et al. 2016; Shi, Zhang, and Holmes 2018; Dolja and Koonin 2018). Notably, these studies have identified arthropods as the ultimate ancestral source of some vertebrate- and plant-infecting RNA viruses (Shi, Lin, Tian, et al. 2016). Since arthropods may be central to the evolutionary history of other viral groups, here we survey terrestrial arthropods for the presence of ssDNA viruses with circular genomes, which follow positive-sense ssRNA viruses as the second most abundant group of viruses infecting eukaryotes (Mahmoudabadi and Phillips 2018; NCBI 2018). This study focuses on a subset of eukaryotic ssDNA viruses with covalently-closed circular genomes that encode a replication-associated protein (Rep).

Prior to 2009, eukaryotic circular Rep-encoding ssDNA (CRESS DNA) viruses were thought to be restricted to plants (Geminiviridae and Nanoviridae families) and vertebrates (family Circoviridae), specifically pigs and birds (King 2012). Since then, metagenomic studies have revealed the cosmopolitan and diverse nature of eukaryotic CRESS DNA viruses. CRESS DNA viruses have now been reported from a wide array of organisms, ranging from primates (e.g., Kapusinszky et al. 2017; Ng et al. 2015) to unicellular algae (Yoon et al. 2011), and ecosystems spanning aquatic (e.g., Labonté and Suttle 2013; Lopez-Bueno et al. 2009; Dayaram, Goldstien, et al. 2015; Hewson, Eaglesham, et al. 2013), terrestrial (e.g., Reavy et al. 2015; Kim et al. 2008), airborne (e.g., Whon et al. 2012) and man-made environments (e.g. Kraberger et al. 2015; Rosario, Duffy, and Breitbart 2009; Rosario et al. 2018). The increased detection and expanded diversity of CRESS DNA viruses has resulted in the establishment of four new taxonomic groups by the International Committee for the Taxonomy of Viruses (ICTV), including three new families.
(Genomoviridae, Smacoviridae, Bacilladnaviridae) and the Cyclovirus genus within the family Circoviridae, to accommodate these diverse viruses (Rosario et al. 2017; Krupovic et al. 2016; Kazlauskas et al. 2017; Varsani and Krupovic 2018). Moreover, many CRESS DNA viruses, which are predicted to represent novel families, remain unclassified. In addition, the investigation of endogenous viral sequences has revealed the ancient origin of CRESS DNA viruses by providing evidence indicating that some of these viruses have been infecting diverse animal and plant hosts for millions of years (Belyi, Levine, and Skalka 2010; Dennis, de Souza, et al. 2018; Lefeuvre et al. 2011). Although integration into host chromosomes may be incidental (Krupovic and Forterre 2015), CRESS DNA endogenous viral elements may influence host evolution and biology by contributing to their genetic composition and, perhaps, providing new functional capabilities (Feschotte and Gilbert 2012; Belyi, Levine, and Skalka 2010). Therefore, eukaryotic CRESS DNA viruses are a highly diverse group of viruses that have implications well beyond their recognized agricultural and medical relevance.

All eukaryotic CRESS DNA viruses are minimalists; their small circular genomes (<6 kb) encode less than eight proteins, including a distinctive homologous Rep (Rosario, Duffy, and Breitbart 2012; Kazlauskas et al. 2017). Another salient feature of most CRESS DNA viral genomes is a conserved putative origin of replication (ori) marked by a nonanucleotide motif at the apex of a predicted stem-loop structure where rolling circle replication (RCR) is initiated (Rosario, Duffy, and Breitbart 2012). The presence of a capsid-encoding ORF (ORF) distinguishes CRESS DNA viruses from CRESS DNA satellite molecules or replicons, such as those classified within the family Alphasatellitidae (Briddon et al. 2018). Although these circular molecules do not encode a capsid, a hallmark defining feature of viruses, these replicons have been considered part of the ‘extended viral world’ as these molecules represent successful genetic parasites (Koonin and Dolja 2014).

It has been hypothesized that CRESS DNA viruses may have evolved from interactions between capsid protein genes from RNA viruses and bacterial plasmids on several independent occasions (Koonin, Dolja, and Krupovic 2015; Krupovic 2013). The polyphyletic nature of CRESS DNA viruses, complemented by their high substitution rates (Duffy, Shackelton, and Holmes 2008; Duffy and Holmes 2009; Firth et al. 2009) and predisposition to recombination (Lefeuvre et al. 2009; Martin et al. 2011), even within the rep gene (Kazlauskas, Varsani, and Krupovic 2018; Krupovic et al. 2015), have resulted in the emergence of a highly diverse viral group. This diversity
is also reflected by different genome architectures that, similar to RNA viruses (Shi, Lin, Tian, et al. 2016; Li et al. 2015), suggests plasticity in CRESS DNA virus genomes. CRESS DNA viruses, including viruses classified within the same genus (e.g., Begomovirus), can have monopartite or multipartite genomes. Notably, multipartite genomes have only been observed in plant-infecting CRESS DNA viruses. Based on the arrangement of major ORFs relative to the putative ori, CRESS DNA genomes display eight genome organizations, including those that only encode a Rep and might represent segments of multipartite genomes or satellite molecules (Rosario, Duffy, and Breitbart 2012). However, there does not seem to be a correlation between these genome organizations and phylogenetic relationships amongst various CRESS DNA viruses (Rosario, Schenck, et al. 2015; Quaiser et al. 2016). Despite this limitation, the Rep remains the only tractable phylogenetic marker that can be used to investigate evolutionary relationships among the highly diverse and polyphyletic CRESS DNA viruses.

Most viral discovery studies focus on vertebrate hosts, primarily mammals, which limits our perspective of viral diversity and evolution. The work presented here expands on studies investigating CRESS DNA viruses in invertebrates, which have traditionally been undersampled (Dayaram et al. 2013; Rosario et al. 2012; Kraberger et al. 2018; Rosario, Schenck, et al. 2015; Bistolas et al. 2017; Hewson, Ng, et al. 2013; Wang et al. 2018; Bettarel et al. 2018). We report 44 CRESS DNA genomes recovered from arthropods representing all three major terrestrial arthropod lineages (Giribet and Edgecombe 2012). By performing a phylogenetic analysis of Reps from CRESS DNA genomes reported from a wide array of organisms and those identified as endogenous viral elements, we demonstrate that most of the previously described CRESS DNA viral phylogenetic diversity circulates among invertebrates. In addition, database searches using newly detected Reps led to the detection of an unreported endogenous cyclovirus-like element within a genome scaffold from a rodent-infecting nematode. Although cycloviruses have been mainly detected in feces from various mammals and homogenized tissues from insects (Rosario et al. 2017), endogenous cyclovirus elements indicate that these viruses are able to infect both arthropod (mites) (Liu et al. 2011) and non-arthropod parasitic invertebrates.
Materials and Methods

Sample collection and processing

A variety of terrestrial arthropods that were opportunistically sampled were screened for CRESS DNA viral sequences (Table 1). Samples included members from all three major groups of terrestrial arthropods including Hexapoda (Class Insecta; Orders: Hymenoptera, Coleoptera, Odonata, Dermaptera, Diptera, Orthoptera, Lepidoptera, Ephemeroptera, Blattodea), Euchelicerata (Class Arachida; Order: Araneae), and Myriapoda (Classes: Diplopoda and Chilopoda). All specimens were identified to the most specific taxonomic rank possible through identification by experts or using DNA barcoding (see below) when taxonomic identifications were not available. Samples were processed following methods used to detect CRESS DNA viruses in marine invertebrates (Rosario, Schenck, et al. 2015) and insects (Rosario et al. 2012; Dayaram et al. 2013). Briefly, specimens were serially rinsed three times using sterile suspension medium (SM) buffer [0.1 M NaCl, 50 mM Tris-HCl (pH 7.5), 10 mM MgSO$_4$] to remove debris. A small piece of tissue was dissected from representative specimens and stored at -20°C for DNA barcoding. Each specimen or pooled sample composed of up to ten specimens from the same species was homogenized by bead-beating using 1.0 mm sterile glass beads in a bead beater (Biospec Products, USA) for 60 - 90 seconds and homogenates were centrifuged at 6000 x g for 6 minutes. Viral particles were then partially purified from supernatants by filtering through a 0.45 µm Sterivex filter (Millipore, USA) and nucleic acids were extracted from 200 µl of filtrate using the QIAmp MinElute Virus Spin Kit (Qiagen, USA).

DNA barcoding was performed to identify any arthropods positive for CRESS DNA viruses that were not taxonomically identified by experts. For this purpose, DNA was extracted from tissue samples using the Quick-DNA Tissue/Insect Kit (Zymo Research, USA) following the manufacturer’s instructions. The mitochondrial cytochrome oxidase I (COI) gene was then amplified through PCR using the universal COI primers LCO1490 (5’GGTCAACAAATCATAAAGATATTGG3’) and HCO2198 (5’TAAACTTCAGGGTGACCAAAAAATCA3’) (Folmer et al. 1994). Fifty microliter PCR reactions contained the following: 1.5 mM MgCl$_2$, 1X Apex NH$_4$ Buffer, 0.5 µM primer LCO1490, 0.5 µM primer HCO2198, 5% DMSO, 1 µg/µl BSA, 1 U Apex Red Taq DNA Polymerase (Genesee Scientific, USA), and 3 µl of template DNA. Thermocycling conditions consisted of an initial denaturation at 95°C for 2 min, followed by 35 cycles of 94°C for 1 min,
48°C for 1 min incrementally decreasing the temperature by 0.1°C on each cycle, and 72°C for 1 min, with a final extension at 72°C for 7 min. Mitochondrial COI gene PCR products were commercially sequenced using LCO1490 and HCO2198 primers. Sequences were compared against GenBank through BLASTn searches. Sequences sharing >95% identity with sequences in the database were classified to species, whereas sequences with nucleotide identities below this threshold were classified at the genus level.

Detection of CRESS DNA viral genomes and genome completion

Small circular templates, such as CRESS DNA genomes, were enriched by amplifying DNA extracts through rolling circle amplification (RCA) using the Illustra TempliPhi Amplification Kit (GE Healthcare, USA) (Haible, Kober, and Jeske 2006; Kim and Bae 2011). RCA products were digested using a suite of six-cutter FastDigest restriction enzymes (Thermo Fisher Scientific, USA), including BamHI, EcoRV, Pdml, HindIII, KpnI, PstI, XhoI, SmaI, BglII, EcoRI, XbaI, and NcoI. Three microliters of RCA product from each sample were digested with each enzyme in separate reactions following the manufacturer’s instructions to obtain complete, unit-length genomes. Products of the restriction digests were resolved on an agarose gel and fragments ranging in size from 1 – 4 kb were excised and purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research, USA). In most cases, products resulting from blunt-cutting enzyme digestions were cloned into the pJET1.2 vector using the CloneJET PCR Cloning kit (Thermo Fisher Scientific, USA), whereas products containing sticky ends were cloned using pGEM-3Zf(+) vectors (Promega, USA) pre-digested with the appropriate enzyme. However, if there were difficulties cloning into pre-digested pGEM-3Zf(+) vectors, sticky-end digestion products were cloned into the pJET1.2 vector following the manufacturer’s sticky-end cloning protocol. Cloned digest products were then Sanger sequenced using vector primers. If these preliminary sequences showed significant similarities to CRESS DNA viral sequences, complete genome sequences were obtained through primer walking of cloned unit-length genomes or through inverse PCR using back-to-back primers designed from preliminary sequences. For the latter, PCR products were cloned using the CloneJET PCR Cloning kit and Sanger sequenced using vector primers and primer walking.
CRESS DNA genome sequence analyses

CRESS DNA genome sequences were assembled in Geneious version R7 (Biomatters, New Zealand) with default parameters for de novo assemblies. Major, non-overlapping ORFs > 100 amino acids were identified and annotated using SeqBuilder version 11.2.1 (Lasergene, USA). Partial genes or genes that seemed interrupted were screened for potential introns using GENSCAN (Burge and Karlin 1997). Genomes that did not contain a putative capsid-encoding ORF, based on BLASTx (Altschul et al. 1990) or remote protein homology searches using HHpred (Söding, Biegert, and Lupas 2005), were further investigated by looking at intrinsically disorder protein (IDP) profiles of non-Rep encoding ORFs using the neural network based VL3 disorder predictor on DisProt (Sickmeier et al. 2007). If non-Rep encoding ORFs contained a high proportion of disordered residues within the first 100 amino acids, they were considered putative capsid proteins (Rosario, Schenck, et al. 2015). The potential ori for each genome was identified by locating the canonical nonanucleotide motif ‘NANTATTAC’ observed in most CRESS DNA genomes (Rosario, Duffy, and Breitbart 2012), or similar sequences (Varsani and Krupovic 2018; Krupovic et al. 2016), and evaluating if the identified nonamer was found at the apex of a predicted stem loop structure using the Mfold web server (Zuker 2003). Genome-wide and Rep amino acid sequence pairwise identities (PI) were calculated using the Sequence Demarcation Tool (SDT) version 1.2 (Muhire, Varsani, and Martin 2014) to evaluate taxonomic relationships among CRESS DNA genomes identified in this study and those found in GenBank.

Phylogenetic analyses

To evaluate how the novel CRESS DNA sequences identified in this study compared to previously reported CRESS DNA viral genomes, we constructed a phylogenetic tree from Rep amino acid sequences recovered from a wide array of organisms. For this purpose, Rep sequences were downloaded from GenBank on April 2018. These sequences included Reps from members of six established CRESS DNA viral families, including Geminiviridae (9 genera), Nanoviridae (2 genera), Circoviridae (2 genera), Genomoviridae (9 genera), Bacilladnaviridae (4 genera), and Smacoviridae (six genera), as well as satellite molecules from the Alphasatellitidae (11 genera) and other CRESS DNA viral genomes that remain unclassified. To reduce the number of sequences in the analysis while still being able to assess diversity, Rep sequences representing established taxonomic groups were clustered based on a 70% amino acid identity cut off using CD HIT (Fu et
However, if more than 20 sequences remained after clustering for a given group, sequences were clustered using a 50% identity cut off. All sequences outside of the established CRESS DNA families were grouped by organism (e.g., rodent associated sequences) and sequences within each group were clustered using a 70% amino acid identity cut off. In addition to sequences representing exogenous viruses and replicons, CRESS DNA-like endogenous viral sequences (CEVs) reported from various organisms were included (Dennis, de Souza, et al. 2018; Liu et al. 2011; Dennis, Flynn, et al. 2018). Selected CEVs did not contain any early stop codons or frameshifts. The final dataset contained 489 Rep sequences.

An alignment was performed using MUSCLE (Edgar 2004) as implemented in MEGA7 (Kumar, Stecher, and Tamura 2016) and manually edited by inspecting and aligning sequences based on the presence of conserved RCR and superfamily 3 (SF3) helicase motifs (Rosario et al. 2017; Varsani and Krupovic 2018; Varsani and Krupovic 2017; Kazlauskas et al. 2017). The alignment was trimmed close to the RCRI and helicase arginine finger motifs and final aligned sequences, including CEVs, were at least 200 amino acids in length (Data S1). The alignment was used to construct an unrooted approximately-maximum-likelihood phylogenetic tree using FastTree 2 (Price, Dehal, and Arkin 2010) with default settings. The phylogenetic tree was edited using TreeGraph 2 (Stöver and Müller 2010) to collapse branches with support below a given threshold of Shimodaira–Hasegawa (SH)-like support and FigTree (http://tree.bio.ed.ac.uk/software/figtree/) was used for tree visualization and editing. The same alignment and tree editing strategies were used for all phylogenetic trees presented in this study.

CRESS DNA genome sequences representing species from established CRESS DNA viral groups were further investigated for genera and/or species level classification assignment. All species level assignments were based on current species demarcation criteria for each viral family (Rosario et al. 2017; Varsani and Krupovic 2017; Varsani and Krupovic 2018). However, some genomes with similarities to members of the Genomoviridae and Circoviridae families seemed to fall outside existing genera; thus, further phylogenetic analyses were undertaken to determine the assignment of these CRESS DNA viruses as putative members of these families. Rep sequences representing members from each of the nine Genomoviridae genera and closely related viruses, including geminiviruses, were aligned and a maximum likelihood (ML) phylogenetic tree was constructed using PhyML (Guindon et al. 2010) with the LG+G+I substitution model. The ML tree was then rooted with viral sequences from members of the Geminiviridae (Varsani and
Krupovic 2017). For the family Circoviridae, Rep sequences representing members from the Circovirus and Cyclovirus genera as well as closely related sequences and CEVs falling within this family were aligned. A midpoint rooted ML phylogenetic tree was then constructed using PhyML with automatic selection of substitution model through the Smart Model Selection (SMS) using the Akaike Information Criterion (AIC) (Lefort, Longueville, and Gascuel 2017).

**Results**

*CRESS DNA viruses identified in all three major lineages of terrestrial arthropods*

More than 500 specimens representing a diversity of terrestrial arthropods were analyzed for the presence of CRESS DNA viruses through RCA followed by restriction enzyme digestion and cloning. Our efforts resulted in the detection of 44 unique (<80% genome-wide pairwise identity) CRESS DNA genomes (Table 2). Consistent with known CRESS DNA genomes, the genomes are small in size (< 3.5 kb) and contain a putative ori marked by a conserved nonanucleotide motif at the apex of predicted stem loop structure, with six out of the eight described CRESS DNA genome organizations detected based on the position of the ori relative to the major ORFs (Rosario, Duffy, and Breitbart 2012). In addition to an identifiable Rep-encoding ORF, most (n=34) of these CRESS DNA genomes encode a putative capsid protein based on similarities to capsid proteins found in public databases. Ten of the detected genomes do not encode an ORF with significant matches to known capsid proteins; however, the non-Rep encoding ORFs in eight of these genomes have similar IDP profiles to those seen in other CRESS DNA viruses, suggesting they encode a putative capsid protein (Data S2) (Rosario, Schenck, et al. 2015). The three genomes for which a capsid-encoding ORF could not be identified were named as ‘circular molecules’ to distinguish these replicons from bona fide CRESS DNA viral genomes.

Half of the CRESS DNA genomes described here were identified in spiders (Class Arachnida; Order Araneae), despite the fact that >70% of the samples processed in this study were insects (Class Insecta; data not shown). The high prevalence of CRESS DNA genomes in spiders is even more striking considering that most spiders were processed individually (Table 1), as opposed to many of the insect species for which multiple individuals were pooled. To our knowledge, these genomes represent the first exogenous CRESS DNA viruses reported from spiders. CRESS DNA viruses were also widely detected in insects with viral genomes retrieved from specimens representing four orders, including ants (Hymenoptera), beetles (Coleoptera), flies
(Diptera), grasshoppers and crickets (Orthoptera), and termites (Blattodea). In addition, we detected a genome from a millipede (Class Diplopoda) representing the first CRESS DNA virus associated with a member of the Subphylum Myriapoda. The low number of CRESS DNA viruses identified from members of the Myriapoda may be a consequence of uneven sampling since only seven specimens from this group were processed. Similarly, groups for which no CRESS DNA viruses were identified, including Odonata, Dermaptera, Lepidoptera, Ephemeroptera, and Chilopoda, had low sample numbers (<10) (data not shown).

Terrestrial arthropods harbor diverse novel CRESS DNA viruses

Over half of the genomes (55%) identified in this study shared <70% genome-wide PI with previously reported sequences (Table 1) and could not be assigned to an existing CRESS DNA group. Phylogenetic analysis of Rep amino acid sequences retrieved from a wide array of organisms illustrated the wide phylogenetic distribution of the arthropod CRESS DNA viruses and replicons identified here (Fig. 1). Some of the arthropod-associated CRESS DNA genomes falling outside of established taxonomic groups were most closely related to isolates that have not been assigned to either genera or families. We identified four genomes from spiders that were most closely related to a circularisvirus reported from dragonflies (Rosario et al. 2012). Phylogenetic analysis revealed other circularisvirus-like genomes retrieved from dragonflies (accession KM598396, Dayaram, Potter, et al. 2015) and bat feces (accession KT732823, Male et al. 2016) (Fig. 1). All of these circularisvirus-like genomes contained similar genomic features, including unisense organization, similar size (~1.9 – 2 kb), and a putative ori on the Rep-encoding strand (Table 2). In addition, circularisvirus-like genomes shared > 57% genome-wide PIs among each other, which is similar to genome-wide PIs reported for established CRESS DNA viral families (Varsani and Krupovic 2017; Varsani and Krupovic 2018; Rosario et al. 2017). Conserved circularisvirus genomic characteristics and genome-wide PIs may grant the formation of a new group of CRESS DNA viruses.

Fire ant associated circular virus 1 (FaACV-1), a viral genome retrieved from fire ants has features characteristic of the crucivirus group, members of which have been mainly reported from environmental samples (Diemer and Stedman 2012; Quaiser et al. 2016; Roux et al. 2013; Steel et al. 2016; McDaniel et al. 2014; Krupovic et al. 2015). Namely, the FaCV-1 genome contains a Rep-encoding ORF most similar to that of CRESS DNA viruses and a putative capsid protein with
significant similarities to capsid proteins of ssRNA viruses from the family *Tombusviridae*. FaCV-1 is most closely related to a crucivirus isolated from wastewater (Table 2). Both FaCV-1 and the wastewater associated crucivirus have two major ORFs arranged in ambisense orientation and share 55% genome-wide PI, suggesting that these genomes may belong to the same viral family. However, currently there is no classification framework for cruciviruses. In addition, we identified a molecule, cybaeus spider associated circular molecule 1 (CySACM-1), containing a single ORF most similar to a Rep-encoding ORF from a crucivirus identified from peatland (Quaiser et al. 2016). The findings presented here suggest that cruciviruses circulate in insects and may be associated with both terrestrial and aquatic (Hewson, Ng, et al. 2013; Bistolas et al. 2017) arthropods.

Although the aim of this study was to identify CRESS DNA genomes, four novel circular molecules that did not encode a Rep were detected (Data S2). These included two molecules, leaf-footed bug associated circular molecule 1 (LfBACM-1) and Spider associated circular molecule 2 (SACM-2), that only contained a single major ORF encoding a putative capsid. The small genome size (<1.2 kb) of these molecules is reminiscent of capsid-encoding genomic segments from multipartite CRESS DNA viruses from the family *Nanoviridae* (Gronenborn 2004). Indeed, LfBACM-1 is most similar to a genomic segment from a novel multicomponent CRESS DNA virus discovered in the feces of fruit-eating bats (Male et al. 2016). However, the SACM-2 putative capsid protein sequence is most similar to the capsid encoded by a presumably monopartite CRESS DNA virus discovered from a sewage oxidation pond (Krabarger et al. 2015). Surprisingly, the remaining two molecules, longjawed orbweaver circular molecule 1 (LjOrbCM-1) and giant house spider associated circular molecule 1 (GhSACM-1), encoded a protein most similar to the large T antigen (LT) encoded by polyomaviruses. The LjOrbCM-1 genome only contained the LT-encoding ORF, whereas GhSACM-1 encoded an additional major, non-overlapping ORF. However, the non-LT encoding ORF of GhSACM-1 was not predicted to encode a structural protein based on homology searches or IDP profiles. These four non-CRESS DNA molecules will not be discussed further, but these findings are noteworthy since they support studies describing capsid-encoding molecules potentially representing novel multipartite viruses associated with unsuspected organisms (Male et al. 2016) and the presence of episomal polyoma-like replicons in spiders (Buck et al. 2016).
Terrestrial arthropods harbor a diversity of species representing new members of established CRESS DNA viral groups

The CRESS DNA genomes that could be assigned to previously reported taxa were dominated by members of the family Genomoviridae, which included genomes retrieved from spiders (n = 7), flies (n = 2), grasshoppers (n = 1) and termites (n = 1) (Table 2). Phylogenetic analysis based on the Rep indicated that the newly identified viruses belong to three genera (Gemycircularvirus, Gemykibivirus, and Gemykolovirus) within the family Genomoviridae (Fig. 2). The majority of arthropod-associated genomoviruses identified here belong to the genus Gemycircularvirus, which is the genus containing the highest number species within the family (Varsani and Krupovic 2017). Based on the species demarcation criteria of 78% genome-wide PI (Varsani and Krupovic 2017), two of the seven identified gemycircularviruses represent new isolates from the classified species dragonfly associated gemycircularvirus 1 and sewage derived gemycircularvirus 4. The remaining five gemycircularviruses represent new species. Spider associated circular viruses (SACVs) 1 and 2 represent a new gemycircularvirus species that was identified in four species of spiders, with SACV-1 and -2 isolates sharing 79-98% genome-wide PI. In addition to gemycircularviruses, we identified two isolates, fly associated circular virus 2 (FlyACV-2) and cybaeus spider associated circular virus 2 (CySACV-2), representing members of the genus Gemykibivirus. Although both of these viruses represent novel gemykibivirus species, a similar isolate to FlyACV-2 (92% genome-wide PI) has been reported from pig feces (Nádia et al. 2017). Lastly, grasshopper associated circular virus 1 (GhACV-1) and tubeweb spider associated circular virus 1 (TuwSACV-1), represent two new species of the genus Gemykolovirus.

In addition to viral genomes that clearly fall within the well-established family Genomoviridae, we identified three genomes from fungus-farming termites that belong to a group of unclassified viruses that appear to be intermediate between genomoviruses and geminiviruses (Fig. 2). In a brief report noting the prevalence of these termite associated circular viruses (TACVs) in African Odontotermes sp. mounds we indicated that these genomes were most similar to members of the Genomoviridae (Kerr et al. 2018). However, Rep phylogenetic analysis indicate that only TACV-2 belongs to the Genomoviridae (genus Gemycircularovirus), while TACV-1, -3, and -4 represent a new group of viruses. Furthermore, TACV-3 and -4 have top BLAST matches to geminiviruses (Table 2), but these genomes cluster closer to genomoviruses than geminiviruses (Fig. 2). There are a number of unclassified sequences retrieved from various environmental
sources that fall in a similar phylogenetic position with TACV-1, -3, and -4, which may grant the
formation of new taxonomic groups.

Members of the family *Circoviridae*, genus *Cyclovirus*, were detected in spiders (n = 2),
flies (n = 1) and ants (n = 1) (Table 2). Based on the species demarcation criteria of 80% genome-
wide PI (Rosario et al. 2017), two of these genomes, arboreal ant associated circular virus 1
(AaACV-1) and soft spider associated circular virus 1 (SoSACV-1), represent novel cyclovirus
species. Fly associated circular virus 1 (FlyACV-1) and spinybacked orbweaver circular virus 2
(SpOrbCV-2) are new isolates of the classified species cockroach associated cyclovirus 1
(CroACV-1) and dragonfly associated cyclovirus 3 (DfACyV-3), respectively. While FlyACV-1
seems to be a new variant of the CroACV-1 species sharing 85% genome-wide PI with this
cyclovirus, SpOrbCV-2 is the same variant as DfACyV-3 sharing 99% PI with this isolate.
Interestingly, DfACyV-3 was discovered from a dragonfly collected in the same region in Florida,
USA (Rosario et al. 2012) as SpOrbCV-2 indicating that this cyclovirus species has been
circulating in the region for at least seven years.

In addition to the four cycloviruses, we identified a genome, fly associated circular virus
5 (FlyACV-5), which is most closely related to cycloviruses and shares the genomic features
characteristic of members in this genus (Table 2). Genome-wide PIs between FlyACV-5 and
known cyclovirus species are within the accepted range for members of the *Circoviridae* (> 55%
genome-wide PIs) (Rosario et al. 2017). However, phylogenetic analysis of Rep sequences from
members of the family *Circoviridae* did not support the placement of FlyACV-5 in either of the
established genera for this family (Fig. 3). FlyACV-5 was most closely related to a CRESS DNA
virus retrieved from bat feces, the Pacific flying fox faeces associated circular DNA virus-8
(PfffACV-8) (Male et al. 2016). Since both FlyACV-5 and PfffACV-8 have genomic features
characteristic of the genus *Cyclovirus* and share genome-wide PIs >55% with members of this
genus, these genomes may represent a novel group within the family *Circoviridae*. The
phylogenetic analysis also revealed two cycloviruses, namely Soft spider associated circular virus
1 (SoSACV-1, accession MH545516) and Pacific flying fox faeces associated circular DNA virus-
2 (PfffACV-2, accession KT732786) that seem to be intermediate between circoviruses and
cycloviruses (Fig. 3). However, at present, these genomes have been classified as cycloviruses
based on their genome organization, which is a mirror image of that observed in circoviruses
(Rosario et al. 2017).
Two new members of the family Smacoviridae were identified in flies (Table 2). Both isolates, fly associated circular viruses (FlyACV) -3 and -4, represent new species belonging to the genus Porprismacovirus based on the species demarcation criteria of 77% genome-wide PI and genus demarcation threshold of 40% Rep amino acid sequence PI (Varsani and Krupovic 2018). Porprismacovirus is by far the genus with the highest number of species in the family (Varsani and Krupovic 2018). Although both FlyACV-3 and -4 represent new species, FlyACV-3 is closely related to an unclassified smacovirus isolated from macaque feces (Kapuinszky et al. 2017). Therefore, FlyACV-3 and the unclassified macaque associated smacovirus represent variants of the same Porprismacovirus species.

In addition to viruses most closely related to members of established CRESS DNA taxonomic groups, we identified an isolate representing a cricket-infecting virus that has not been classified. Cricket associated circular virus 1 (CrACV-1), identified in a store-bought cricket, represents an isolate of Achaeta domesticus volvovirus (AdVVV) (Pham, Bergoin, and Tijssen 2013; Pham et al. 2013). The four volvovirus genomes that have been reported to date, including CrACV-1, have been recovered from commercial crickets and share >99% genome-wide PI, thus representing a single viral species.

Detection of a cyclovirus endogenous element in a non-arthropod invertebrate

Analysis of BLASTn matches for fly associated circular viruses (FlyACV) -6 and -7 in the GenBank non-redundant database revealed weak matches to nematodes. Although these initial BLAST matches were non-significant, this prompted a search in the WormBase Parasite database (Howe et al. 2017) using the FlyACVs putative Rep sequences as queries. This search led to the detection of a previously unreported CEV from Hymenolepis microstoma, commonly known as rodent tapeworm. The H. microstoma CEV is embedded within a 251 kb genome scaffold (accession LN902886) (Tsai et al. 2013). The GenBank record for this genome scaffold noted two Rep-associated coding sequences (CDS) that were positioned next to each other. One of the CDS contains a near full-length Rep (accession CDS32196), whereas the second one (accession CDS32195) is interrupted at the SF3 Walker-A motif, at which point a partial capsid is encoded in the same reading frame. Inclusion of the near full-length Rep from H. microstoma in the phylogenetic analysis showed that this sequence falls within the Cyclovirus genus (Fig. 3). Although CEVs have been previously noted from parasitic helminths (Liu et al. 2011), this is the
first cyclovirus CEV reported from nematodes. Notably, the *H. microstoma* CEV Rep sequence is most closely related to a cyclovirus sequence reported from rodents (accession KY370028). The putative endogenous capsid sequence is most similar to a cyclovirus reported from cat feces (Zhang et al. 2014), which is also closely related to the *H. microstoma* CEV and the rodent cyclovirus based on the Rep (Fig. 3). Therefore, both the CEV Rep and capsid sequences indicate that *H. microstoma* has been infected at some point by a cyclovirus. To our knowledge, this is the first evidence indicating that non-arthropod invertebrates serve as hosts for cycloviruses.

**Discussion**

**Most CRESS DNA viral diversity circulates among arthropods and other invertebrates**

From the relatively small-scale survey presented here, it is clear that terrestrial arthropods harbor an extensive diversity of CRESS DNA viruses. Combining our results with previous reports, exogenous CRESS DNA viruses have now been reported in organisms from four out of the five major branches of the Arthropod Tree of Life (Giribet and Edgecombe 2012), including Euchelicerata (Class Arachnida) (Kraberger et al. 2018; Wang et al. 2018), Hexapoda (Class Insecta) (Rosario et al. 2012; Dayaram et al. 2013; Dayaram, Potter, et al. 2015; Padilla-Rodriguez, Rosario, and Breitbart 2013; Kraberger et al. 2017; Tikhe and Husseneder 2017), Myriapoda (Class Diplopoda), and Crustacea (Classes Malacostraca, Maxillopoda, Copepoda, Branchiopoda) (Rosario, Schenck, et al. 2015; Hewson, Ng, et al. 2013; Bistolas et al. 2017; Dunlap et al. 2013). To the best of our knowledge, no studies have specifically looked for CRESS DNA viruses within the remaining major arthropod branch, the Pycnogonida (sea spiders). Spiders were identified as an unsuspected rich reservoir of CRESS DNA viral diversity, harboring most of the genomes from distinct viral groups identified in this survey. However, it should be noted that spiders are insectivores; thus, it is possible that the wide array of CRESS DNA viral diversity they contain is partially the result of accumulating CRESS DNA viruses from their insect prey. Similarly, a high diversity of CRESS DNA viruses has been reported from dragonflies, which are also top insect predators (Rosario et al. 2012; Dayaram et al. 2013). Since our methods might have recovered viruses from dietary content, it is possible that generalist arthropod predators may contain a broader range of CRESS DNA viruses than dietary specialists. Additionally, the discovery of some CRESS DNA viruses in multiple arthropod species may be due to overlapping diets. Phylogenetic analysis of the conserved Rep indicates that many of the diverse CRESS DNA viruses found within
the terrestrial arthropods and other invertebrates fall outside established CRESS DNA viral families and do not form cohesive phylogenetic groups. This observation suggests that the CRESS DNA viral diversity associated with arthropods and other invertebrates has been grossly underestimated and that additional sampling of these groups would continue to expand the CRESS DNA virosphere.

The diversity of CRESS DNA genomes identified in this study spans the entire CRESS DNA phylogenetic tree, representing Reps recovered from a wide array of eukaryotic organisms (Fig. 1). Moreover, the CRESS DNA diversity falling outside established taxonomic groups that has been detected within arthropods and other invertebrates overwhelms the diversity reported from vertebrate organisms and plants, despite the fact that the latter groups have been heavily sampled. Members of the \textit{Alphasatellitidae}, \textit{Geminiviridae} and \textit{Nanoviridae} families were not detected in our survey, which did not include plant virus insect vectors. However, these plant-infecting CRESS DNA viruses and satellite molecules have long been known to circulate among hemipteran vectors (Hogenhout et al. 2008), which have been exploited to discover viral species found in a given area (Rosario et al. 2016; Rosario, Seah, et al. 2015; Ng et al. 2011). Therefore, arthropod associated viruses include members from five out of the six CRESS DNA viral families that have been identified as monophyletic (Kazlauskas et al. 2017). The remaining established CRESS DNA family that was not identified in our survey, \textit{Bacilladnaviridae}, includes viruses infecting unicellular algae and has only been reported from aquatic environments (Kazlauskas et al. 2017).

We identified novel viruses representing new members from the \textit{Genomoviridae}, \textit{Smacoviridae}, and \textit{Circoviridae} families. Genomoviruses were the most diverse group, with 12 genomes recovered from spiders and insects from the orders Blattodea, Coleoptera, Diptera, and Orthoptera. The arthropod associated genomoviruses represent at least three genera and highlight the ever growing diversity and wide distribution of this group of viruses (Krupovic et al. 2016). Two CRESS DNA viral sequences representing members of the \textit{Smacoviridae}, which have been mainly recovered from feces from various mammals (Ng et al. 2015; Steel et al. 2016; Varsani and Krupovic 2018), were recovered from blow flies (Diptera: Calliphoridae) collected in the Caribbean. The novel fly associated smacoviruses, FlyACV -3 and -4, were recovered from blow flies collected using a chicken carcass as bait (Yusseff-Vanegas and Agnarsson 2017). Since blow flies are known to feed on feces and tissues from various vertebrates, including mammals, it is
likely that FlyACV -3 and -4 represent dietary content. Nevertheless, the detection of fly associated
smacoviruses and a smacovirus in dragonflies (Dayaram, Potter, et al. 2015) indicates that this
group of viruses circulates within arthropods.

Members of the family *Circoviridae* present a unique distribution relative to all of the other
established CRESS DNA viral groups infecting multicellular organisms. All of the *Circoviridae*
members identified in this study represent the genus *Cyclovirus*. Although cycloviruses have been
identified in both vertebrates and arthropods, to date, members of the genus *Circovirus* have
mainly been reported from the former. Moreover, this observation is consistent with CEV searches
(Liu et al. 2011; Dennis, de Souza, et al. 2018; Belyi, Levine, and Skalka 2010) including a survey
of more than 680 animal genomes, ~50% of which were invertebrates (Dennis, Flynn, et al. 2018).
Viruses reported from ticks are the only arthropod associated CRESS DNA viruses belonging to
the genus *Circovirus* (Wang et al. 2018; Tokarz et al. 2018). Since ticks are hematophagous
parasites that feed exclusively on the blood of birds and mammals (Basu and Charles 2017), it is
possible that tick associated circoviruses represent vertebrate-infecting viruses, in particular avian
circoviruses (Fig. 3). Interestingly, bona fide circoviruses have not been reported from mosquitoes,
a major group of blood-feeding arthropods of public health relevance. However, this might reflect
the scarcity of mosquito DNA viromes reported to date. Despite these caveats, the available data
suggests that cycloviruses circulate in a wide array of invertebrates and mammals, whereas
circoviruses are mainly restricted to vertebrates and, perhaps, blood-feeding arthropod vectors.

In the Rep-based phylogeny, cycloviruses appear basal with respect to circoviruses (Fig. 3). Based on the higher diversity of cycloviruses described to date and the wider distribution of
these viruses in both vertebrates and invertebrates, it is conceivable that cycloviruses are ancestral
to circoviruses. We also note that there is a group of cycloviruses recovered from spiders and an
insectivorous bat that seem intermediate between other cycloviruses and the circovirus clade (Fig. 3).
Moreover, we detected viral sequences with cyclovirus genome organization from blow flies
and an insectivorous bat that do not fall within the cyclovirus Rep clade that may represent a novel
group within the *Circoviridae*. Further sampling of CRESS DNA viruses found in arthropods and
other invertebrates may help resolve phylogenetic relationships among members of the
*Circoviridae*. Nevertheless, our phylogenetic analysis supports the idea that there are distinct
groups of cycloviruses (Dennis, Flynn, et al. 2018). In addition, ancestors to members of the genus
*Circovirus* may be found within arthropods.
It appears that the complete phylogenetic breadth of CRESS DNA viral diversity that has been reported to date circulates within arthropods and other invertebrates, which is analogous to what has been noted for RNA viruses (Shi, Lin, Tian, et al. 2016; Li et al. 2015; Shi et al. 2018). Few CRESS DNA Rep phylogenetic clusters are represented by viral sequences recovered from either vertebrates, plants, or fungi alone (Fig. 1). This observation includes established CRESS DNA viral groups as well as novel sequences that have not been assigned to taxonomic groups. In addition, the vast majority of CRESS DNA sequences recovered from plants and fungi, including CEVs, fall near or within the closely related Geminiviridae and Genomoviridae clades. Vertebrate-associated CRESS DNA sequences that fall outside established groups have been mainly reported from fecal samples and most are intermixed with sequences that have been reported from invertebrates (Fig. 1). However, we identified one divergent clade that only included sequences from CRESS DNA viral isolates recovered from mammal feces and may represent a vertebrate-infecting lineage. More sampling, including blood or tissue samples as opposed to fecal samples, is needed to confirm this possibility. Despite the presence of plant-specific (Geminiviridae and Nanoviridae) and potentially vertebrate-specific viral lineages (genus Circovirus), most of the CRESS DNA viral diversity identified in vertebrates, plants and fungi is nested within the much broader genetic diversity of invertebrate associated viruses.

Related CRESS DNA viruses identified in disparate organisms

The phylogenetic analysis revealed many Rep sequences from disparate sources grouping together in the same clade (Fig. 1). Even when looking at broad source classifications, such as vertebrates, plants, arthropods and other invertebrates, few of the clades that fall outside of the established CRESS DNA viral groups represent isolates retrieved from similar sources. Moreover, the same ‘source intermixing’ can be observed within established CRESS DNA groups (Figs. 2-3). CRESS DNA viral isolates representing members of the Genomoviridae are a notable example. Genomovirus genomes have been recovered from plants, fungi, vertebrates and arthropods; however, there is no clear separation of genomovirus groups based on the source (Fig. 2). With such phylogenetic distribution, it is tempting to speculate about potential cross-species CRESS DNA virus transmission. However, since many CRESS DNA viruses have been identified through molecular assays alone, it is difficult to predict the host for most of these viruses, including those that fall within established CRESS DNA groups. Therefore, we cannot make inferences
regarding horizontal CRESS DNA virus transmission. Furthermore, cross species transmission
between arthropod and vertebrate infecting viruses has been deemed unlikely (Dennis, Flynn, et
al. 2018). Nevertheless, available data suggests that closely related CRESS DNA viruses circulate
among disparate organisms, providing opportunities for cross-viral species interactions that may
lead to recombination and the emergence of new viral species (Roux et al. 2013; Krupovic 2013).

The detection of related CRESS DNA viruses in disparate eukaryotic organisms might be
partly explained by vectored viruses and/or viruses infecting hosts that interact closely with other
species in a similar niche. The two CRESS DNA viral families whose members infect plants and
were not identified in our survey, namely Geminiviridae and Nanoviridae, are transmitted by
hemipteran vectors where these viruses may be found in high titers (Czosnek et al. 2017; Watanabe
and Bressan 2013). In contrast to arthropod-borne animal-infecting RNA viruses and some
vectored plant RNA viruses, it is thought that CRESS DNA plant viruses do not replicate or
express genes in their vector (i.e., non-propagative transmission) (Dietzgen, Mann, and Johnson
2016). However, there is evidence showing genetic changes of a nanovirus within its aphid vector
(Sicard et al. 2015). Moreover, some begomoviruses (family Geminiviridae) can replicate within
their whitefly vector (Czosnek et al. 2017) and alter the whitefly feeding behavior to result in
enhanced virus transmission (Liu et al. 2013). Therefore, there is evidence for complex cross-
kingdom interactions between plant-infecting CRESS DNA viruses and their insect vectors.

In addition to recognized interactions between plant-infecting CRESS DNA viruses and
their hemipteran vectors, CRESS DNA viruses may be present in systems where cross-kingdom
species interactions are intertwined. We discovered CRESS DNA viruses in organisms that have
been previously investigated for their role in model symbiosis systems, including fungus- and
plant-insect systems. Interestingly, CRESS DNA viruses discovered in both fungus-farming
insects investigated here, including Bark beetle associated circular virus 1 (BbACV-1) and
TACVs, were most closely related to members of the Genomoviridae (Fig. 2). Infection assays
with Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1), the only
genomovirus with a confirmed host (Yu et al. 2010), suggest that some genomoviruses are able to
infect both fungi and insects. The primary SsHADV-1 host is a plant fungal pathogen; however,
the virus is able to replicate in a mycophagous insect that it potentially uses as a transmission
mechanism (Liu et al. 2016). Although we are not able to discern if BbACV-1 and TACVs infect
either the farmer (insect) or the crop (fungus) in these ancient agricultural systems (Mueller and
Gerardo 2002), their discovery provides two additional examples where genomovirus-insect-fungi interactions are tightly connected.

Another notable example is the novel AaACV-1 cyclovirus, which was identified multiple times in all three of ant species tested from the whistling thorn (*Vachellia (Acacia) drepanolobium*) ant-plant system in East Africa and genomes recovered from each ant species shared > 98% PI. The three ant species tested (*Tetraponera penzigi*, *Crematogaster nigriceps* and *Crematogaster mimosae*) live in an obligate mutualism with the acacia tree throughout its range (Young, Stubblefield, and Isbell 1997), protecting their host plants from mammalian herbivores, but also shaping aspects of the host environment such as plant-associated fungal communities, in a species specific manner (Baker et al. 2017). Although the ants inhabit domatia (arthropod-occupied chambers) of the same acacia tree species, each tree typically hosts a single colony at any point in time (Palmer et al. 2000). Opportunities for direct viral transmission between colonies of the same or different species are therefore likely to be restricted to infrequent antagonistic interactions between colonies occupying neighboring trees, so we did not expect to discover the same AaACV-1 virus in all three ant species. On the other hand, each tree typically hosts multiple colonies sequentially over its lifetime, providing the ants with shelter in domatia and food in the form of extrafloral nectar. The common host plant thus represents a good candidate mechanism for the circulation of AaACV-1 in these three different ant species. It remains to be determined what role, if any, AaACV-1 infection plays in this system. Interestingly, an endogenous cyclovirus Rep has been identified in another arboreal ant (*Pseudomyrmex gracilis*) (Fig. 3) (Dennis, Flynn, et al. 2018) suggesting that cycloviruses may be common in complex ant-plant symbiotic systems (Clement et al. 2008).

Genomic fossil record supports widespread distribution of CRESS DNA viruses among invertebrates

Evidence from CEVs supports that CRESS DNA viruses infect or have infected a diversity of organisms, including hosts from four of the five supergroups of eukaryotes (Liu et al. 2011; Dennis, de Souza, et al. 2018; Dennis, Flynn, et al. 2018; Belyi, Levine, and Skalka 2010). Here we took advantage of previously reported CEVs and put them in a phylogenetic context with extant exogenous CRESS DNA viruses recovered from a wide array of organisms. Analyzed CEVs only included Rep sequences that were over 200 amino acids in length and did not contain any early
stop codons or frameshifts. Therefore, analyzed CEVs potentially represent relatively recent CRESS DNA viral infections or functional elements co-opted by the host (Dennis, Flynn, et al. 2018; Liu et al. 2011). Notably, a CEV identified in the germline of the brown recluse spider *Loxosceles reclusa* (sequence ID: CVE.Loxosceles_reclusa.7, Dennis, Flynn, et al. 2018) is most closely related to the longjawed orbweaver circular virus 2 (LjOrbCV-2, accession MH545529) identified here, suggesting that viruses similar to LjOrbCV-2 infect spiders. Interestingly, all of the CEVs reported from vertebrates clustered within the genus *Circovirus*, whereas those reported from invertebrates were distributed across the CRESS DNA Rep phylogenetic tree (Fig. 1). Similarly, CEVs found in fungal genomes clustered near the *Genomoviridae* and *Geminiviridae* clades. Therefore, the available CEV data support that CRESS DNA viruses infecting vertebrates, plants, and fungi show a limited phylogenetic distribution compared to viruses found in invertebrates.

CEVs have revealed multiple insertions from related as well as distinct CRESS DNA viruses in various vertebrate host germlines (Dennis, de Souza, et al. 2018). Multiple CEVs have also been reported from the same invertebrate species, with up to 19 sequences identified in a given host (Dennis, Flynn, et al. 2018) and our phylogenetic analysis supports that divergent CRESS DNA viruses can infect the same invertebrate host species (Fig. 1). For example, we were able to include in our analysis seven distinct Rep sequences previously identified as CEVs from brine flies (genus *Ephydra*) (Dennis, Flynn, et al. 2018), representing two species, *E. hidrans* (n = 5) and *E. gracilis* (n = 2). Surprisingly, none of the endogenized Rep sequences from brine flies clustered close to each other. The closest phylogenetic neighbors to *Ephydra* Rep CEVs perhaps reflect the ecology of these unique dipterans (Class Insecta) that live at the interface between aquatic and terrestrial habitats. Brine flies live in hypersaline alkaline lakes feeding on benthic algae and, whereas the larval stages are underwater, adults are considered terrestrial (Herbst 1980). The various *Ephydra* CEVs were most closely related to Reps recovered mainly from aquatic invertebrates. However, one of the CEVs clustered with a Rep representing a CRESS DNA virus recovered from another dipteran (mosquitoes).

CEVs identified in parasitic invertebrates, including protozoans and nematodes, also demonstrate infection by distinct CRESS DNA elements. The presence of Rep-like sequences in the genomes of enteric protozoan parasites, including *Giardia* and *Entamoeba*, were noted more than 10 years ago (Gibbs et al. 2006). Further investigation of CEVs from these parasites showed
that various Rep elements are transcribed, at least in *E. histolytica*, suggesting that the Rep has
been co-opted for the benefit of the parasite (Liu et al. 2011). We were able to include seven Rep
CEVs previously detected in *Giardia intestinalis* (n = 2) (Gibbs et al. 2006) and three species of
*Entamoeba* (n = 5) (Fig. 1) (Liu et al. 2011). The two analyzed Rep sequences from *Giardia*
clustered together, whereas sequences from *E. histolytica* clustered in two groups. Therefore, the
unicellular parasite *E. histolytica* has been infected by at least two distinct CRESS DNA elements.
Although we were not able to include multiple CEVs from the same nematode species, three out
of four CEVs from parasitic nematodes clustered together, suggesting that there is a nematode-
infecting lineage of CRESS DNA viruses or at least an ancestral one. Notably, extrachromosomal
virus-like elements encoding a Rep have been reported from a free-living freshwater nematode
(Rebrikov et al. 2002) and, thus, CRESS DNA-like elements may be more common in nematodes
than previously recognized.

CRESS DNA viral infection of parasitic organisms transmitted through the fecal-oral route
further exemplify the difficulties associated with predicting potential hosts for viruses identified
in feces. There is a possibility that CRESS DNA viral sequences identified in vertebrate feces
actually represent parasite-infecting CRESS DNA viruses since some parasites are ubiquitous and
can be found in high numbers in fecal matter (e.g., Oates et al. 2012). For example, three of the
CEVs from enteric protozoan parasites clustered close to CRESS DNA viral sequences retrieved
from mammal feces (Fig. 1). Two of these were *Giardia* CEVs that clustered with FlyACV-6 and
-7 identified here and rodent associated viruses (Phan et al. 2011). Since blow flies feed on fecal
matter, FlyACV-6 and -7 may represent ingested viruses from feces that may infect either rodents
or a parasitic protozoan host. Similarly, we detected a CEV in a rodent-infecting parasitic
tapeworm, *H. microstoma*, whose Rep was most similar to cycloviruses recovered from rodents
and cat feces (Fig. 3). It is difficult to establish with the available information if the cat and rodent
associated cycloviruses infect these mammals or co-occurring parasitic organisms. Alternatively,
it is an intriguing possibility that parasitic eukaryotes can act as CRESS DNA viral vectors, which
would result in tightly connected CRESS DNA virus-eukaryotic parasite-host interactions.

The detection of CEVs in a wide diversity of parasitic eukaryotes, including nematodes,
protozoans, and arthropods indicates that the role of parasitic organisms in CRESS DNA viral
ecology and evolution should be explored. Moreover, there are extant exogenous CRESS DNA
viruses circulating in ectoparasitic arachnids (mites and ticks) (Kraberger et al. 2018; Tokarz et al.
The detection of divergent RNA viruses in parasitic nematodes has also highlighted the need to further investigate the role of parasites in virus evolution (Shi, Lin, Tian, et al. 2016). The scarcity of data regarding parasite-associated microbes, including viruses, has been recognized and efforts are underway to try to address this knowledge gap (Dheilly et al. 2017). Exploration of undersampled invertebrate taxa and non-fecal samples from vertebrates will certainly provide more insights into the evolution of CRESS DNA viruses. In turn, the discovery of divergent CRESS DNA viruses will inform efforts investigating the genomic fossil record to better understand viral evolution and host biology (Feschotte and Gilbert 2012; Krupovic and Forterre 2015).

Conclusion

Here we described CRESS DNA viruses from terrestrial arthropods with a widespread phylogenetic distribution, including members of yet unclassified viral groups. The cosmopolitan distribution of CRESS DNA viruses, combined with the dynamic nature of these viruses, which seem to commonly exchange genetic information (Kazlauskas, Varsani, and Krupovic 2018; Quaiser et al. 2016), may help explain the unprecedented diversity recognized within recent years. It is important to note that our analysis underestimates the diversity within CRESS DNA viral genomes since we only included genome sequences associated with specific organisms, as opposed to environments (e.g., sewage, seawater), which would have added more than 200 Rep sequences to our phylogeny. In spite of this, it is clear that arthropods and other invertebrates harbor an extensive diversity of CRESS DNA viruses that dwarfs the genomic diversity observed in vertebrates, plants, and fungi. CRESS DNA viruses are emerging as a dominant and diverse group in the eukaryotic DNA viral world, with each report of novel genomes expanding the boundaries of this group. In contrast to other eukaryotic DNA viruses, CRESS DNA viruses are associated with a wide range of organisms across the tree of life, from unicellular eukaryotes to chordates, reflecting their ancient evolutionary history. Viral discovery efforts in undersampled taxa promise to reveal a more complete view of CRESS DNA virus diversity that will elucidate evolutionary linkages among these successful ‘genetic parasites’ (Koonin and Dolja 2014).
Acknowledgements

We would like to thank students (Class of 2019) from the Laura Mercado Specialized School of Agro-ecology in San Germán, Puerto Rico for enthusiastically collecting specimens for this project during an outreach activity and Dr. José L. Agosto from the University of Puerto Rico, Río Piedras Campus for his assistance in conducting such activity. We also acknowledge Rachel C. Harbeitner, Dawn B. Goldsmith, Allison Cohen, Parker E. Jernigan, Sidney Fulford, and Carina Graham for their help with sample processing. Thanks to Naomi E. Pierce (Harvard University) and Jon G. Sanders (University of California, San Diego) for facilitating access to ant samples from Kenya. Thanks to the CarBio team (http://www.islandbiogeography.org/participants.html) for their help collecting blow fly samples in the Caribbean.
Tables

Table 1. Sample information and identified CRESS DNA genomes.

Table 2. CRESS DNA genome information, accession numbers, and taxonomic groups identified in this study.

Figures

Figure 1. Approximately maximum likelihood phylogenetic tree of replication-associated protein (Rep) amino acid sequences representing CRESS DNA viruses, replicons, and CRESS DNA-like endogenous viral (CEV) elements recovered from various organisms. Branches representing sequences associated with various types of organisms are distinguished by color. Clades containing Rep sequences falling within established CRESS DNA viral groups including the Genomoviridae (Genomo), Geminiviridae (Gemini), Bacilladnaviridae (Bac), Circoviridae (Circo), Smacoviridae (Smaco), Nanoviridae (N) and Alphasatellitidae (Alpha) were merged and are highlighted in grey. Branches representing CEVs identified in Ephydra spp. (dipteran), including E. gracilis (EpG) and E. hians (EpH), and nematodes (N) are specified. CEVs identified in protists are also specified, including Entamoeba (Ent), Giardia intestinalis (G) and Blastocystis hominis (B). Entamoeba species are further distinguished, namely E. invadens (EntI), E. histolytica (EntH), and E. dispar (EntD). Reps identified in this study are highlighted with schematics of terrestrial arthropods showing their source and broad phylogenetic distribution. The new circularisvirus clade is highlighted in light blue. Branches with <80% Shimodaira–Hasegawa (SH)-like support were collapsed.

Figure 2. Maximum likelihood phylogenetic tree of replication-associated protein (Rep) amino acid sequences representing members of the family Genomoviridae and related CRESS DNA viruses. Branches representing sequences associated with various types of organisms and environmental sources are distinguished by color. Bars on the right indicate clades representing genomovirus genera and unclassified sequences; geminiviruses were used as an outgroup. Genomovirus Reps identified in this study are highlighted with schematics of terrestrial arthropods and include viruses identified from sierra dome spiders (SdSACV), pimoid spiders (PiSACV), tubeweb spiders (TuwSACV), grasshoppers (GhACV) and termites (TACV). Viruses identified in multiple species of spiders are identified as spider associated circular viruses (SACV). Branches with <70% Shimodaira–Hasegawa (SH)-like support were collapsed.
Figure 3. Midpoint rooted maximum likelihood phylogenetic tree of selected replication-associated protein (Rep) amino acid sequences representing members of the family *Circoviridae* and related CRESS DNA viruses. Branches representing viruses associated with various invertebrate and vertebrate organisms are distinguished by color. Bars on the right indicate clades representing the *Cyclovirus* and *Circovirus* genera. Rep sequences representing CRESS DNA-like endogenous viral (CEV) elements are highlighted with stars. Cyclovirus Reps identified in this study are highlighted with schematics of terrestrial arthropods and include viruses identified from flies (FlyACV), ants (AaACV), soft spiders (SoSACV) and spinyback orbweavers (SpOrbCV). Reps representing unclassified genome sequences forming non-*Circoviridae* clades were merged and are highlighted in grey (accessions: KX246259, KR528563, KM598407, KR528546, KM874290, KM874319, KM874343, KT945164). Branches with <70% Shimodaira–Hasegawa (SH)-like support were collapsed.

Supplemental Material

**Supplemental Data S1** – Alignment used to create approximately maximum likelihood tree shown in Figure 1 in fasta format. Details regarding sequence acronyms, descriptions, and accession numbers can be found in Supplemental Data S2.

**Supplemental Data S2** – Excel workbook containing three worksheets (W#) with the following information: details regarding intrinsically disordered protein (IDP) profiles (W1), detailed information for sequences found in the alignment used to create approximately maximum likelihood tree shown in Figure 1 (Data S1) (W2), and a table listing accession numbers and information regarding non-CRESS DNA molecules detected in terrestrial arthropods (W3).
References


Dolja, V. V., and E. V. Koonin. 2018. 'Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer', Virus Research, 244: 36-52.


Gronenborn, B. 2004. 'Nanoviruses: genome organisation and protein function', Veterinary Microbiology: 103-09.


Hewson, I., G. Ng, W. Li, B. A. LaBarre, I. Aguirre, J. G. Barbosa, M. Breitbart, A. W. Greco, C. M. Kearns, A. Looi, L. R. Schaffner, P. D. Thompson, and N. G. Hairston. 2013. 'Metagenomic identification, seasonal dynamics, and potential transmission mechanisms of a Daphnia-associated single-stranded DNA virus in two temperate lakes', Limnology and Oceanography, 58: 1605-20.


Kerr, M., K. Rosario, C. C. M. Baker, and M. Breitbart. 2018. 'Discovery of four novel circular single-stranded DNA viruses in fungus-farming termites', Genome Announcements, 6: e00318.


Koonin, E. V., and V. V. Dolja. 2014. 'Virus world as an evolutionary network of viruses and capsidless selfish elements', Microbiology and Molecular Biology Reviews, 78: 278-303.

Kraberger, S., G. R. Argüello-Astorga, L. G. Greenfield, C. Galilee, D. Law, D. P. Martin, and
A. Varsani. 2015. 'Characterisation of a diverse range of circular replication-associated
protein encoding DNA viruses recovered from a sewage treatment oxidation pond',
*Infection, Genetics and Evolution*, 31: 73-86.

Kraberger, S., J. E. Polston, H. M. Capobianco, R. I. Alcalá-Briseño, R. S. Fontenele, and A.
Varsani. 2017. 'Genomovirus genomes recovered from *Echinothrips americanus* sampled
in Florida, USA', *Genome Announcements*, 5: e00445-17.

Kraberger, S., G. A. Visnovsky, R. F. van Toor, M. F. Male, K. Waits, R. S. Fontenele, and A.
Varsani. 2018. 'Genome sequences of two single-stranded DNA viruses identified in
*Varroa destructor*, *Genome Announcements*, 6: e00107-18.

Krupovic, M. 2013. 'Networks of evolutionary interactions underlying the polyphyletic origin of
ssDNA viruses', *Current Opinion in Virology*, 3: 578-86.

Krupovic, M., and P. Forterre. 2015. 'Single-stranded DNA viruses employ a variety of
mechanisms for integration into host genomes', *Annals of the New York Academy of
Sciences*, 1341: 41-53.

of widespread single-stranded DNA viruses', *Archives of Virology*, 161: 2633-43.

Krupovic, M., N. Zhi, J. Li, G. Hu, E. V. Koonin, S. Wong, S. Shevchenko, K. Zhao, and N. S.
Young. 2015. 'Multiple layers of chimerism in a single-stranded DNA virus discovered
by deep sequencing', *Genome Biology and Evolution*, 7: 993-1001.

Kumar, S., G. Stecher, and K. Tamura. 2016. 'MEGA7: Molecular Evolutionary Genetics

Labonté, J. M., and C. A. Suttle. 2013. 'Previously unknown and highly divergent ssDNA viruses
populate the oceans', *The ISME Journal*, 7: 2169-77.


Holmes, and Y. Z. Zhang. 2015. 'Unprecedented genomic diversity of RNA viruses in
arthropods reveals the ancestry of negative-sense RNA viruses', *Elife*, 4: e05378.

'Multiple forms of vector manipulation by a plant-infecting virus: *Bemisia tabaci* and

Liu, H. Q., Y. P. Fu, B. Li, X. Yu, J. T. Xie, J. S. Cheng, S. A. Ghabrial, G. Q. Li, X. H. Yi, and
D. H. Jiang. 2011. 'Widespread horizontal gene transfer from circular single-stranded
DNA viruses to eukaryotic genomes', *BMC Evolutionary Biology*, 11: 276.

2016. 'Fungal DNA virus infects a mycophagous insect and utilizes it as a transmission
vector', *Proceedings of the National Academy of Sciences*, 113: 12803–08.

diversity of the viral community from an Antarctic lake', *Science*, 326: 858-61.


Pham, H. T., M. Bergoin, and P. Tijsen. 2013. 'Acheta domesticus volvovirus, a novel single-stranded circular DNA virus of the house cricket', *Genome Announcements*, 1: e00079.

Pham, H. T., H. Iwao, M. Bergoin, and P. Tijsen. 2013. 'New volvovirus isolates from *Acheta domesticus* (Japan) and *Gryllus assimilis* (United States)', *Genome Announcements*, 1: e00328-13.


Stöver, B. C., and K. F. Müller. 2010. 'TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses', BMC Bioinformatics, 11: 7.

Tikhe, C. V., and C. Husseneder. 2017. 'Metavirome sequencing of the termite gut reveals the presence of an unexplored bacteriophage community', Frontiers in Microbiology, 8: 2548.


Figure 1 (on next page)

Approximately maximum likelihood phylogenetic tree of replication-associated protein (Rep) amino acid sequences from CRESS DNA viruses, replicons, and CRESS DNA-like endogenous viral (CEV) elements.

Approximately maximum likelihood phylogenetic tree of replication-associated protein (Rep) amino acid sequences representing CRESS DNA viruses, replicons, and CRESS DNA-like endogenous viral (CEV) elements recovered from various organisms. Branches representing sequences associated with various types of organisms are distinguished by color. Clades containing Rep sequences falling within established CRESS DNA viral groups including the Genomoviridae (Genomo), Geminiviridae (Gemini), Bacilladnaviridae (Bac), Circoviridae (Circo), Smacoviridae (Smaco), Nanoviridae (N) and Alphasatellitidae (Alpha) were merged and are highlighted in grey. Branches representing CEVs identified in Ephydra spp. (dipteran), including *E. gracilis* (EpG) and *E. hians* (EpH), and nematodes (N) are specified. CEVs identified in protists are also specified, including *Entamoeba* (Ent), *Giardia intestinalis* (G) and *Blastocystis hominis* (B). Entamoeba species are further distinguished, namely *E. invadens* (EntI), *E. histolytica* (EntH), and *E. dispar* (EntD). Reps identified in this study are highlighted with schematics of terrestrial arthropods showing their source and broad phylogenetic distribution. The new circularisvirus clade is highlighted in light blue. Branches with <80% Shimodaira–Hasegawa (SH)-like support were collapsed.
0.7 amino acid substitutions per site
Maximum likelihood phylogenetic tree of replication-associated protein (Rep) amino acid sequences representing members of the family *Genomoviridae* and related CRESS DNA viruses.

Maximum likelihood phylogenetic tree of replication-associated protein (Rep) amino acid sequences representing members of the family *Genomoviridae* and related CRESS DNA viruses. Branches representing sequences associated with various types of organisms and environmental sources are distinguished by color. Bars on the right indicate clades representing genomovirus genera and unclassified sequences; geminiviruses were used as an outgroup. Genomovirus Reps identified in this study are highlighted with schematics of terrestrial arthropods and include viruses identified from sierra dome spiders (SdSACV), pimoid spiders (PiSACV), tubeweb spiders (TuwSACV), grasshoppers (GhACV) and termites (TACV). Viruses identified in multiple species of spiders are identified as spider associated circular viruses (SACV). Branches with <70% Shimodaira–Hasegawa (SH)-like support were collapsed.
SH-like branch support
- >0.95
- >0.9-0.95
- >0.7-0.9

0.3 amino acid substitutions per site
Figure 3 (on next page)

Midpoint rooted maximum likelihood phylogenetic tree of selected replication-associated protein (Rep) amino acid sequences representing members of the family *Circoviridae* and related CRESS DNA viruses.

Midpoint rooted maximum likelihood phylogenetic tree of selected replication-associated protein (Rep) amino acid sequences representing members of the family *Circoviridae* and related CRESS DNA viruses. Branches representing viruses associated with various invertebrate and vertebrate organisms are distinguished by color. Bars on the right indicate clades representing the *Cyclovirus* and *Circovirus* genera. Rep sequences representing CRESS DNA-like endogenous viral (CEV) elements are highlighted with stars. Cyclovirus Reps identified in this study are highlighted with schematics of terrestrial arthropods and include viruses identified from flies (FlyACV), ants (AaACV), soft spiders (SoSACV) and spinyback orbweavers (SpOrbCV). Reps representing unclassified genome sequences forming non-*Circoviridae* clades were merged and are highlighted in grey (accessions: KX246259, KR528563, KM598407, KR528546, KM874290, KM874319, KM874343, KT945164). Branches with <70% Shimodaira–Hasegawa (SH)-like support were collapsed.
<table>
<thead>
<tr>
<th>Table 1 (on next page)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample information and identified CRESS DNA genomes.</td>
</tr>
</tbody>
</table>

Sample information and identified CRESS DNA genomes.
<table>
<thead>
<tr>
<th>Year</th>
<th>Location</th>
<th>Species name (common name)</th>
<th>Samples</th>
<th>Identified genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>Kenya</td>
<td>Crematogaster nigriceps (Arboreal ant)</td>
<td>Pool (2)</td>
<td>Arboreal ant associated circular virus 1</td>
</tr>
<tr>
<td>2011</td>
<td>Kenya</td>
<td>Tetraponera penzigi (Arboreal ant)</td>
<td>Pool (2)</td>
<td>Arboreal ant associated circular virus 1</td>
</tr>
<tr>
<td>2011</td>
<td>Kenya</td>
<td>Crematogaster mimosae (Arboreal ant)</td>
<td>Pool</td>
<td>Arboreal ant associated circular virus 1</td>
</tr>
<tr>
<td>2013</td>
<td>FL USA</td>
<td>Solenopsis invicta (Fire ant)</td>
<td>Pool</td>
<td>Fire ant associated circular virus 1</td>
</tr>
<tr>
<td>2013</td>
<td>FL USA</td>
<td>Xylosandrus amputates (Bark beetle)</td>
<td>Pool</td>
<td>Bark beetle associated circular virus 1</td>
</tr>
<tr>
<td>2014</td>
<td>Puerto Rico</td>
<td>Dineuteus sp. (Water beetle)*</td>
<td>Single (4)</td>
<td>Water beetle associated circular virus 1</td>
</tr>
<tr>
<td>2013</td>
<td>Store</td>
<td>Gryllus assimilis (Field cricket)</td>
<td>Pool</td>
<td>Cricket associated circular virus 1</td>
</tr>
<tr>
<td>2011</td>
<td>FL USA</td>
<td>Romulea microptera (Lubber grasshopper)</td>
<td>Single</td>
<td>Grasshopper associated circular virus 1</td>
</tr>
<tr>
<td>2013</td>
<td>Nevis</td>
<td>Lucilia rica (Blow fly)*</td>
<td>Pool</td>
<td>Fly associated circular virus 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fly associated circular virus 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fly associated circular virus 5</td>
</tr>
<tr>
<td>2013</td>
<td>St. Barts</td>
<td>Fannia sp. (Dung fly)*</td>
<td>Pool</td>
<td>Fly associated circular virus 2</td>
</tr>
<tr>
<td>2013</td>
<td>Dom. Republic</td>
<td>Lucilia retroversa (Blow fly)</td>
<td>Pool</td>
<td>Fly associated circular virus 4</td>
</tr>
<tr>
<td>2013</td>
<td>Guadeloupe</td>
<td>Lucilia rica (Blow fly)*</td>
<td>Pool</td>
<td>Fly associated circular virus 5</td>
</tr>
<tr>
<td>2013</td>
<td>St. Barts</td>
<td>Lucilia rica. (Blow fly)*</td>
<td>Pool</td>
<td>Fly associated circular virus 7</td>
</tr>
<tr>
<td>2015</td>
<td>NH USA</td>
<td>Oxidus sp. (Greenhouse millipede)*</td>
<td>Single</td>
<td>Millipede associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Parasteatoda tepidariorum (Common house spider)</td>
<td>Single</td>
<td>Common house spider circular molecule 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Cybaeus signifer</td>
<td>Single</td>
<td>Cybaeus spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Cybaeus signifer</td>
<td>Single</td>
<td>Cybaeus spider associated circular virus 2</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Cybaeus signifer</td>
<td>Single</td>
<td>Cybaeus spider associated circular virus 3</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Steatoda grossa (False black widow spider)</td>
<td>Single</td>
<td>False black widow spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Eratigena duellina (Giant house spider)</td>
<td>Single</td>
<td>Giant house spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Eratigena duellina (Giant house spider)</td>
<td>Single</td>
<td>Giant house spider associated circular virus 2</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Eratigena duellina (Giant house spider)</td>
<td>Single</td>
<td>Giant house spider associated circular virus 3</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Eratigena duellina (Giant house spider)</td>
<td>Single</td>
<td>Giant house spider associated circular virus 4</td>
</tr>
<tr>
<td>2014</td>
<td>Puerto Rico</td>
<td>Nephila sp. (Golden silk orbweaver)*</td>
<td>Single</td>
<td>Golden silk orbweaver associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>FL USA</td>
<td>Leucauge argyra (Longjawed orbweaver)*</td>
<td>Single</td>
<td>Longjawed orbweaver circular virus 1</td>
</tr>
<tr>
<td>2014</td>
<td>Puerto Rico</td>
<td>Leucauge argyra (Longjawed orbweaver)*</td>
<td>Single</td>
<td>Longjawed orbweaver circular virus 2</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Pimoa altioculata (Pimoid spider)*</td>
<td>Single</td>
<td>Pimoid spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Pimoa altioculata (Pimoid spider)*</td>
<td>Single</td>
<td>Pimoid spider associated circular virus 2</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Neriere litigiosa (Sierra dome spider)*</td>
<td>Single</td>
<td>Sierra dome spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Neriere litigiosa (Sierra dome spider)*</td>
<td>Single</td>
<td>Sierra dome spider associated circular virus 2</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Cybaeidae (Soft spider)</td>
<td>Single</td>
<td>Soft spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Cybaeus signifer</td>
<td>Single</td>
<td>Spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Segestria pacifica (Tubeweb spider)</td>
<td>Single</td>
<td>Spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Eratigena atrica (Giant house spider)</td>
<td>Single</td>
<td>Spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Parasteatoda tepidariorum (Common house spider)</td>
<td>Single</td>
<td>Spider associated circular virus 2</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Segestria pacifica (Tubeweb spider)*</td>
<td>Single</td>
<td>Spider associated circular virus 3</td>
</tr>
<tr>
<td>2014</td>
<td>FL USA</td>
<td>Gasteracantha cancriformis (Spinybacked orbweaver)*</td>
<td>Single</td>
<td>Spinybacked orbweaver circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>FL USA</td>
<td>Gasteracantha cancriformis (Spinybacked orbweaver)*</td>
<td>Single</td>
<td>Spinybacked orbweaver circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>FL USA</td>
<td>Gasteracantha cancriformis (Spinybacked orbweaver)*</td>
<td>Single</td>
<td>Spinybacked orbweaver circular virus 2</td>
</tr>
<tr>
<td>Year</td>
<td>Location</td>
<td>Species</td>
<td>Sample Type</td>
<td>Virus Name</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>---------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>2017</td>
<td>FL USA</td>
<td>Cyrtophora sp. (Tentweb spider)</td>
<td>Single</td>
<td>Tentweb spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Segestria pacifica (Tubeweb spider)*</td>
<td>Single</td>
<td>Tubeweb spider associated circular virus 1</td>
</tr>
<tr>
<td>2017</td>
<td>Victoria BC</td>
<td>Dysdera crocata (Woodlouse hunter spider)</td>
<td>Single</td>
<td>Woodlouse hunter spider associated circular virus 1</td>
</tr>
<tr>
<td>2015</td>
<td>Kenya</td>
<td>Odontotermes sp. (Fungus-farming termite)</td>
<td>Pool</td>
<td>Termite associated circular virus 1</td>
</tr>
<tr>
<td>2015</td>
<td>Kenya</td>
<td>Odontotermes sp. (Fungus-farming termite)</td>
<td>Pool (2)</td>
<td>Termite associated circular virus 1, Termite associated circular virus 3, Termite associated circular virus 4</td>
</tr>
</tbody>
</table>

*Location abbreviations: FL = Florida, NH = New Hampshire, St. Barts = Saint Barthelemy, Dom. Republic = Dominican Republic, BC = British Columbia, Store = Carolina Biological Supply

Many specimens were taxonomically identified by sample providers. However, some specimens were identified through DNA barcoding and are indicated with an asterisk (*); nd = not determined.

Samples processed as individuals (Single) or pools (Pool) composed of up to 10 individuals from the same species are distinguished. Some CRESS DNA genomes were identified in multiple individuals or pools (number of samples is specified within parenthesis).
Table 2 (on next page)

CRESS DNA genome information, accession numbers, and taxonomic groups identified in this study.

CRESS DNA genome information, accession numbers, and taxonomic groups identified in this study.
<table>
<thead>
<tr>
<th>Accession Number(s)</th>
<th>Genotype¹</th>
<th>Taxonomic Affiliation²</th>
<th>Genome size</th>
<th>Nonanucleotide Motif³</th>
<th>BLAST match source (accession number)⁴</th>
<th>Identity (%)⁵</th>
<th>Rep</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>MH54551- MH545513</td>
<td>Arbitral ant associated circular virus 1</td>
<td>Circoviridae</td>
<td>1769</td>
<td>TAGTATTAC (II)</td>
<td>Bat feces (KM382269)</td>
<td>73*</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>MH545514</td>
<td>Fly associated circular virus 1</td>
<td>Circoviridae</td>
<td>1722</td>
<td>TAGTATTAC (II)</td>
<td>Cockroach (JX569794)</td>
<td>89</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>MH545516</td>
<td>Soft spider associated circular virus 1</td>
<td>Circoviridae</td>
<td>1937</td>
<td>TAGTATTAC (II)</td>
<td>Shrew feces (AB379879)</td>
<td>63*</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>MH545515</td>
<td>Spinybacked orbweaver circular virus 2</td>
<td>Circoviridae</td>
<td>1707</td>
<td>TAGTATTAC (II)</td>
<td>Dragonfly (JX185424)</td>
<td>99*</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>MH545522</td>
<td>Cybaeus spider associated circular virus 1</td>
<td>Circoviridae</td>
<td>1991</td>
<td>TAACTACTAC (V)</td>
<td>Dragonfly (JX185415)</td>
<td>61</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>MH545520</td>
<td>Golden silk orbweaver associated circular virus 1</td>
<td>Circularisvirus</td>
<td>2054</td>
<td>CAGTATTAC (V)</td>
<td>Dragonfly (JX185415)</td>
<td>63</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>MH545521</td>
<td>Longjawed orbweaver circular virus 1</td>
<td>Circularisvirus</td>
<td>1905</td>
<td>CATTATTAC (V)</td>
<td>Dragonfly (JX185415)</td>
<td>60</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>MH545518- MH545519</td>
<td>Spinybacked orbweaver circular virus 1</td>
<td>Circularisvirus</td>
<td>1955</td>
<td>CAGTATTAC (V)</td>
<td>Dragonfly (JX185415)</td>
<td>63</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>MH545523</td>
<td>Fire ant associated circular virus 1</td>
<td>Crucivirus</td>
<td>3226</td>
<td>TATGTGTA (A)</td>
<td>Wastewater (KY487859)</td>
<td>61</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>MH545497</td>
<td>Bark beetle associated circular virus 1</td>
<td>Genomoviridae</td>
<td>2237</td>
<td>TAAATATTAT (II)</td>
<td>Dragonfly (JX185429)</td>
<td>96*</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>MH545498</td>
<td>Cybaeus spider associated circular virus 2</td>
<td>Genomoviridae</td>
<td>2344</td>
<td>TAAATATTAT (II)</td>
<td>Whitefly (KY230625)</td>
<td>67*</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>MH545499</td>
<td>Giant house spider associated circular virus 1</td>
<td>Genomoviridae</td>
<td>2093</td>
<td>TAAATATTAT (II)</td>
<td>Llama feces (KT862245)</td>
<td>73*</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>MH545500</td>
<td>Grasshopper associated circular virus 1</td>
<td>Genomoviridae</td>
<td>2309</td>
<td>TAACTACTAT (II)</td>
<td>Bat feces (KT732803)</td>
<td>62*</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>MH545501</td>
<td>Pimoid spider associated circular molecule 1⁶</td>
<td>Genomoviridae</td>
<td>1662</td>
<td>TAAATTTAT (II)</td>
<td>Llama feces (KT862245)</td>
<td>69*</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>MH545502</td>
<td>Pimoid spider associated circular virus 1</td>
<td>Genomoviridae</td>
<td>2240</td>
<td>TAAATTTAT (II)</td>
<td>Sewage (KJ547640)</td>
<td>100*</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>MH545503- MH545505</td>
<td>Spider associated circular virus 1</td>
<td>Genomoviridae</td>
<td>2214-2216</td>
<td>TAAATTTAT (II)</td>
<td>Bird feces (KF371636)</td>
<td>67*</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>MH545506</td>
<td>Spider associated circular virus 2</td>
<td>Genomoviridae</td>
<td>2204</td>
<td>TAAATTTAT (II)</td>
<td>Cow feces (KT862253)</td>
<td>84*</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>MH545502- MG917675</td>
<td>Termite associated circular virus 1</td>
<td>Genomoviridae</td>
<td>2222-2226</td>
<td>TAAATTTAT (II)</td>
<td>Thrips (KY308268)</td>
<td>74*</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>MH545501</td>
<td>Tubeweb spider associated circular virus 1</td>
<td>Genomoviridae</td>
<td>2174</td>
<td>TAAATTTAT (II)</td>
<td>Thrips (KY308270)</td>
<td>63*</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>MH545524</td>
<td>Fly associated circular virus 3</td>
<td>Smacoviridae</td>
<td>2537</td>
<td>TAGTGTTTAC (IV)</td>
<td>Macaques feces (KU043428)</td>
<td>83</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>MH545525</td>
<td>Fly associated circular virus 4</td>
<td>Smacoviridae</td>
<td>2546</td>
<td>TAGTGTTTAC (IV)</td>
<td>Chimpanzee feces (GQ351275)</td>
<td>57</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>MH545526</td>
<td>Cricket associated circular virus 1</td>
<td>Smacoviridae</td>
<td>2515</td>
<td>TAGTATTAC (II)</td>
<td>Cricket (KC794539)</td>
<td>100</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>MH545538</td>
<td>Common house spider circular molecule 1⁷</td>
<td>Unclassified</td>
<td>1833</td>
<td>TATTATTAC⁸ (V)</td>
<td>Giant panda feces (MF327573)</td>
<td>62</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>MH545543</td>
<td>Cybaeus spider associated circular molecule 1⁹</td>
<td>Unclassified</td>
<td>1989</td>
<td>TAGCCTAA (VIII)</td>
<td>Peatland (KX388505)⁹</td>
<td>58*</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>MH545542</td>
<td>False black widow spider associated circular virus 1</td>
<td>Unclassified</td>
<td>2199</td>
<td>TATTATTAC (I)</td>
<td>Reclaimed water (NC_013023)</td>
<td>61</td>
<td>62</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. CRESS DNA genome information, accession numbers, and taxonomic groups identified in this study.
<table>
<thead>
<tr>
<th>Accession</th>
<th>Species</th>
<th>Taxonomy</th>
<th>Year</th>
<th>Nonamer</th>
<th>Host</th>
<th>Database ID</th>
<th>Blast E-value</th>
<th>Pairwise Identity (%)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MH545517</td>
<td>Fly associated circular virus 5</td>
<td>Unclassified</td>
<td>1997</td>
<td>TAGTATTAC (II)</td>
<td>Bat feces (KT732825)</td>
<td></td>
<td>97</td>
<td>93</td>
<td>A few genomes were not considered viral and were labelled as ‘molecules’ for the following reasons: (A) capsid-encoding open reading frame (ORF) seemed truncated, (B) non-Rep encoding ORF displayed ordered IDP profiles within N-terminus (B), or (C) genome only contained a single major ORF.</td>
</tr>
<tr>
<td>MH545530</td>
<td>Fly associated circular virus 6</td>
<td>Unclassified</td>
<td>2103</td>
<td>TAGTATTAC (IV)</td>
<td>Wastewater (KY487991)*</td>
<td></td>
<td>61</td>
<td>59</td>
<td>Groups that do not represent established taxonomic groups by ICTV are non-italicized, including Circularisvirus, Crucivirus, and Volvovirus.</td>
</tr>
<tr>
<td>MH545531</td>
<td>Fly associated circular virus 7</td>
<td>Unclassified</td>
<td>2010</td>
<td>TAGTATTAC (IV)</td>
<td>Wastewater (KY487991)*</td>
<td></td>
<td>65</td>
<td>66</td>
<td>Most nonamers were located at the apex of a predicted hairpin structure, with the exception of a circular molecule identified with the symbol (^). Genome organizations, using the specified nonamer as a reference, are indicated within parenthesis according to genotypes discussed by Rosario et al. 2012.</td>
</tr>
<tr>
<td>MH545536</td>
<td>Giant house spider associated circular virus 2</td>
<td>Unclassified</td>
<td>2040</td>
<td>TAGTATTAC (V)</td>
<td>Sphaerid clam (KP153475)</td>
<td></td>
<td>68</td>
<td>64</td>
<td>Best BLAST matches for identified CRESS DNA genomes. Some of the most closely related viruses to CRESS DNA viruses and replicons identified here, based on BLAST searches, contain a different genome organization and are indicated with the symbol (#).</td>
</tr>
<tr>
<td>MH545537</td>
<td>Giant house spider associated circular virus 3</td>
<td>Unclassified</td>
<td>2290</td>
<td>TATTATTAC (I)</td>
<td>Amphipod (KC248416)</td>
<td></td>
<td>61</td>
<td>59</td>
<td>Pairwise identities (PIs) between identified CRESS DNA genomes and their best BLAST match. Nucleotide PIs between replication-associated proteins (Rep) were calculated based on predicted spliced coding regions. Genomes containing Rep-encoding ORFs interrupted by an intron are marked with the symbol (*).</td>
</tr>
<tr>
<td>MH545541</td>
<td>Giant house spider associated circular virus 4</td>
<td>Unclassified</td>
<td>2494</td>
<td>TAAATATTAC (IV)</td>
<td>Wastewater (KY487963)</td>
<td></td>
<td>62</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>MH545529</td>
<td>Longjawed orbweaver circular virus 2</td>
<td>Unclassified</td>
<td>2321</td>
<td>CAGTATTAC (VI)</td>
<td>Damselfly (KM598408)</td>
<td></td>
<td>58</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>MH545532</td>
<td>Millipede associated circular virus 1</td>
<td>Unclassified</td>
<td>1987</td>
<td>TAGTATTAC (II)</td>
<td>Estuarine snail (NC_026646)</td>
<td></td>
<td>59</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>MH545534</td>
<td>Pimoid spider associated circular virus 2</td>
<td>Unclassified</td>
<td>2125</td>
<td>TAGTATTAC (I)</td>
<td>Bat (KJ641721)*</td>
<td></td>
<td>62</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>MH545535</td>
<td>Sierra dome spider associated circular virus 2</td>
<td>Unclassified</td>
<td>1860</td>
<td>TAGTATTAC (V)</td>
<td>Giant panda feces (NC_035196)</td>
<td></td>
<td>57</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>MH545539</td>
<td>Spider associated circular virus 3</td>
<td>Unclassified</td>
<td>1889</td>
<td>CAACCACCTC (I)</td>
<td>Ice shelf pond (NC_024478)</td>
<td></td>
<td>57</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>MH545533</td>
<td>Tentweb spider associated circular virus 1</td>
<td>Unclassified</td>
<td>2127</td>
<td>TAGTATTAC (II)</td>
<td>Dragonfly larvae (KF738884)</td>
<td></td>
<td>62</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>MG917674</td>
<td>Termite associated circular virus 1</td>
<td>Unclassified</td>
<td>2155</td>
<td>TAAATTAC (II)</td>
<td>Chicken feces (K056250)</td>
<td></td>
<td>61*</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>MG917676</td>
<td>Termite associated circular virus 3</td>
<td>Unclassified</td>
<td>2220</td>
<td>TAATGTTAC (II)</td>
<td>Shrub (KT214387)</td>
<td></td>
<td>57*</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>MG917677</td>
<td>Termite associated circular virus 4</td>
<td>Unclassified</td>
<td>2152</td>
<td>TAATGTTAC (II)</td>
<td>Tomato (NC_036591)</td>
<td></td>
<td>58*</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>MH545527</td>
<td>Water beetle associated circular virus 1</td>
<td>Unclassified</td>
<td>2244</td>
<td>CAGTATTAC (II)</td>
<td>Ice shelf pond (NC_024478)</td>
<td></td>
<td>56</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>MH545528</td>
<td>Water beetle associated circular virus 1</td>
<td>Unclassified</td>
<td>2176</td>
<td>TAAATAGTAG (II)</td>
<td>Amphipod (KC248416)</td>
<td></td>
<td>57*</td>
<td>58</td>
<td></td>
</tr>
</tbody>
</table>

1 A few genomes were not considered viral and were labelled as ‘molecules’ for the following reasons: (A) capsid-encoding open reading frame (ORF) seemed truncated, (B) non-Rep encoding ORF displayed ordered IDP profiles within N-terminus (B), or (C) genome only contained a single major ORF.  
2 Groups that do not represent established taxonomic groups by ICTV are non-italicized, including Circularisvirus, Crucivirus, and Volvovirus.  
3 Most nonamers were located at the apex of a predicted hairpin structure, with the exception of a circular molecule identified with the symbol (^). Genome organizations, using the specified nonamer as a reference, are indicated within parenthesis according to genotypes discussed by Rosario et al. 2012.  
4 Best BLAST matches for identified CRESS DNA genomes. Some of the most closely related viruses to CRESS DNA viruses and replicons identified here, based on BLAST searches, contain a different genome organization and are indicated with the symbol (#).  
5 Pairwise identities (PIs) between identified CRESS DNA genomes and their best BLAST match. Nucleotide PIs between replication-associated proteins (Rep) were calculated based on predicted spliced coding regions. Genomes containing Rep-encoding ORFs interrupted by an intron are marked with the symbol (*).