Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta)

Karyna Rosario,1 Anisha Dayaram,2 Milen Marinov,2 Jessica Ware,3 Simona Kraberger,2 Daisy Stainton,2 Mya Breitbart1 and Arvind Varsani2,4,5

Correspondence
Arvind Varsani
arvind.varansi@canterbury.ac.nz

Received 12 July 2012
Accepted 16 August 2012
1College of Marine Science, University of South Florida, 140 Seventh Avenue South, St Petersburg, FL 33701, USA
2School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
3Department of Biological Sciences, Rutgers University, 249 University Avenue, Newark, NJ 07102, USA
4Biomolecular Interaction Centre, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
5Electron Microscope Unit, University of Cape Town, Rondebosch, Cape Town 7701, South Africa

Viruses with circular ssDNA genomes that encode a replication initiator protein (Rep) are among the smallest viruses known to infect both eukaryotic and prokaryotic organisms. In the past few years an overwhelming diversity of novel circular Rep-encoding ssDNA (CRESS-DNA) viruses has been unearthed from various hosts and environmental sources. Since there is limited information regarding CRESS-DNA viruses in invertebrates, this study explored the diversity of CRESS-DNA viruses circulating among insect populations by targeting dragonflies (Epiprocta), top insect predators that accumulate viruses from their insect prey over space and time. Using degenerate PCR and rolling circle amplification coupled with restriction digestion, 17 CRESS-DNA viral genomes were recovered from eight different dragonfly species collected in tropical and temperate regions. Nine of the genomes are similar to cycloviruses and represent five species within this genus, suggesting that cycloviruses are commonly associated with insects. Three of the CRESS-DNA viruses share conserved genomic features with recently described viruses similar to the mycovirus Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1, leading to the proposal of the genus Gemycircularvirus. The remaining viruses are divergent species representing four novel CRESS-DNA viral genera, including a gokushovirus-like prokaryotic virus (microphage) and three eukaryotic viruses with Reps similar to circoviruses. The novelty of CRESS-DNA viruses identified in dragonflies using simple molecular techniques indicates that there is an unprecedented diversity of ssDNA viruses among insect populations.

INTRODUCTION

ssDNA viruses with circular genomes were once thought to infect a limited number of hosts and have narrow environmental distributions. Metagenomic studies have challenged this view by revealing the ubiquity of these small viruses. Out of the six viral families composed of viruses with circular ssDNA genomes, four have been consistently detected in a wide variety of environmental samples through recent metagenomic surveys, namely the families Circoviridae, Geminiviridae, Nanoviridae and Microviridae (Rosario & Breitbart, 2011; Rosario et al., 2012). Members of these families have icosahedral virions, with the exception of geminiviruses which have geminate particles, and their genomes encode a well-conserved replication initiator protein (Rep) involved in rolling circle replication (RCR) (Fauquet et al., 2005). These circular Rep-encoding ssDNA (CRESS-DNA) viruses are known to infect a fairly narrow range of hosts within bacteria (family Microviridae), a variety of plants (families Geminiviridae and Nanoviridae), as well as a limited number of vertebrate species (i.e. pigs and birds; family Circoviridae). However, recent studies have associated CRESS-DNA viruses with previously unknown hosts and their widespread detection suggests that this viral type is successful in many environments (Rosario et al., 2012).
The detection of eukaryotic CRESS-DNA viruses in hosts other than plants, pigs and birds indicates that there is a wealth of unexplored taxonomic host groups that may be infected by these viruses or involved in their life cycles (e.g. vectors). A novel virus with similarities to nanoviruses was recently detected in protists, specifically picobiliphytes (Yoon et al., 2011), and the first DNA virus discovered in fungi was a novel viral species (Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1; SsHADV-1) most similar to geminiviruses (Yu et al., 2010). Genomes similar to SsHADV-1 have now been detected in cassava plants (Dayaram et al., 2012), mosquitoes (Ng et al., 2011b) and badger faeces (van den Brand et al., 2012). Animal viruses from the family Circoviridae, specifically the genus Circovirus, have also recently been detected in fish (Lőrincz et al., 2011, 2012). Viruses from a sister group of circoviruses, the proposed genus Cyclovirus, have been discovered in tissues and faecal matter of a wide range of mammals and birds (Ge et al. 2011; Li et al., 2010a, 2011). Furthermore, cycloviruses have been recently detected in dragonflies (Epiprocta), suggesting that members of the family Circoviridae also infect invertebrates (Rosario et al., 2011). Until recently, prokaryotic CRESS-DNA viruses from the family Microviridae (microphages) were thought to only infect enterobacteria (genus Microviridae) and obligate intracellular bacterial parasites (subfamily Gokushovirinae), including Chlamydia, Bdellovibrio and Spiroplasma (Cherwa & Fan, 2011). However, the host range of gokushoviruses was recently extended to a fourth bacterial phylum, Bacteroidetes (Bacteroidetes-associated microviruses, BMVs) (Krupovic & Forterre, 2011). These findings, in conjunction with the detection of novel gokushovirus-like genomes in marine environments (Tucker et al., 2011), suggest that microphage diversity has also been grossly underestimated.

Geminiviruses and nanoviruses are known to be transmitted by insect vectors (Hogenhout et al., 2008); however, there is limited information regarding other CRESS-DNA viral groups in insects. To our knowledge, a single member of the family Microviridae has been isolated from an insect pathogen, specifically Spiroplasma virus 4 (SpV4) which infects the honeybee pathogenic bacterium Spiroplasma melliferum (Renaudin & Bové, 1994). In 2011, a novel cyclovirus (Dragonfly cyclovirus; DfCyV) was discovered in dragonflies from the family Libellulidae (Rosario et al., 2011). DfCyV is the first CRESS-DNA virus identified in dragonflies and represents the first member of the family Circoviridae recovered from insects. Although it is difficult to establish whether DfCyV infects dragonflies based on its molecular detection, the discovery of this virus suggests that there are unknown CRESS-DNA viruses circulating among insect populations.

In an effort to identify CRESS-DNA viruses associated with insects, this study targeted adult dragonflies, which are highly mobile, top predators that have the potential of accumulating viruses from their insect prey over space and time. This study combined the insect-hunting ability of dragonflies with methods that specifically target viruses with circular ssDNA genomes to detect CRESS-DNA viruses circulating in insects. Seventeen complete CRESS-DNA viral genomes were recovered from eight different dragonfly species collected in the Kingdom of Tonga, Bulgaria and the USA (Florida and Puerto Rico), demonstrating that there is a diverse and unexplored community of CRESS-DNA viruses associated with insects. The novel viruses reported here include several distinct cycloviruses, circovirus-like species representing three novel genera, myco-like viruses belonging to the hereby proposed genus Gemycircularvirus and a gokushovirus-like microphage.

**RESULTS AND DISCUSSION**

Although a range of CRESS-DNA viruses has been described in samples from a variety of animals and environments, little effort has been made to identify these viruses in invertebrates. The recent discovery of a cyclovirus in dragonflies provided the first evidence of an animal CRESS-DNA viral group (family Circoviridae) in insects (Rosario et al., 2011). This study has successfully used dragonflies as sampling traps to explore the diversity of CRESS-DNA viruses circulating in winged insect populations, as these highly mobile, top insect predators can accumulate viruses from their insect prey over space and time. Two different strategies were employed to detect CRESS-DNA viruses present in dragonflies, rolling circle amplification (RCA) coupled with restriction enzyme (RE) digestion and degenerate PCR. The first strategy exploited the RCA bias towards circular ssDNA templates followed by RE digestion of enriched viral concatenated genomes into complete, unit-length genomes. Since there is a precedent for detecting the family Circoviridae in dragonflies, the second approach employed a degenerate PCR assay targeting this viral group. Both approaches were used to isolate full genomes, as it has been shown that sequences with similarities to a single ORF may not reflect the genome architecture of known ssDNA viruses (Rosario et al., 2012). The RE digestion method resulted in unit-length genomes that were cloned and sequenced, providing a methodological advantage over the degenerate PCR method. While full genome sequences were also obtained with the PCR method, this required extra effort since the short sequences recovered from the degenerate PCR assay were then used to design primers for inverse PCR to amplify, clone and sequence full genomes. The two approaches proved to be complementary since neither method alone recovered the full suite of CRESS-DNA viruses reported in this study.

Dragonflies were collected from different areas using nets or obtained from archived collections (Table 1). CRESS-DNA viruses were detected in 55% of the dragonfly specimens (n=77), suggesting that these viruses circulate widely in winged insect populations. Notably, CRESS-DNA viruses were recovered from dragonflies regardless of the sample preservation and filtration methods used (Table 1). However, no CRESS-DNA viruses were detected in samples...
Table 1. Dragonfly specimens processed during this study

<table>
<thead>
<tr>
<th>Collection date</th>
<th>Location</th>
<th>Location type</th>
<th>Dragonfly species</th>
<th>Preservation method</th>
<th>Homogenization</th>
<th>0.45 µm Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>07/1990</td>
<td>Poyakonda Biological Station, Finland</td>
<td>Not available</td>
<td>Somatochlora metallica</td>
<td>95% acetone/air-drying</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>07/1991</td>
<td>Novaki-csatorna, Hungary</td>
<td>Not available</td>
<td>Cordulia aenea</td>
<td>95% acetone/air-drying</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>08/1991</td>
<td>Halaszi, Hungary</td>
<td>Not available</td>
<td>Somatochlora flavomaculata</td>
<td>95% acetone/air-drying</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>06/2003</td>
<td>Austria</td>
<td>Not available</td>
<td>Cordulia aenea</td>
<td>95% acetone/air-drying</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>06/2001; 06/2007</td>
<td>Finland</td>
<td>Not available</td>
<td>Somatochlora flavomaculata</td>
<td>95% acetone/air-drying</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>05/2007</td>
<td>Saarburg, Germany</td>
<td>Not available</td>
<td>Cordulia aenea</td>
<td>95% acetone/air-drying</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>06/2004; 06/2007</td>
<td>Bulgaria</td>
<td>Urban areas, lakes</td>
<td>Cordulia aenea</td>
<td>95% acetone/air-drying</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>06/1992; 07/1998; 06/2004; 07/2000</td>
<td>Bulgaria</td>
<td>River, streams, rural areas</td>
<td>Somatochlora meridionalis</td>
<td>95% acetone/air-drying</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>07/2007</td>
<td>Eastern Rhodope Mountains, Bulgaria</td>
<td>River, stream</td>
<td>S. meridionalis*</td>
<td>95% acetone/air-drying</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>04/2010</td>
<td>Tongatapu, Tonga</td>
<td>Agricultural fields, rural residential area</td>
<td>Pantala flavescens*</td>
<td>95% ethanol/air-drying</td>
<td>Disposable pellet pestle</td>
<td>Acrodisc</td>
</tr>
<tr>
<td>04–05/2010</td>
<td>Vava’u, Tonga</td>
<td>Lake</td>
<td>Diplacodes bipunctata*</td>
<td>95% ethanol/air-drying</td>
<td>Disposable pellet pestle</td>
<td>Acrodisc</td>
</tr>
<tr>
<td>09/2010</td>
<td>St. Petersburg, FL (FL1 site)</td>
<td>University campus</td>
<td>Myarthria marcella*, P. flavescens*, Tramea lacerata*, Tramea carolina</td>
<td>Freezing</td>
<td>Disposable pellet pestle</td>
<td>Acrodisc</td>
</tr>
<tr>
<td>09/2010</td>
<td>St. Petersburg, FL (FL2 site)</td>
<td>Lake park</td>
<td>Coryphaeschna ingens*, Erythemis simplicicollis*</td>
<td>Freezing</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>10/2010</td>
<td>Everglades, FL</td>
<td>Creek</td>
<td>Coryphaeschna ingens*, Celithemis eponina, Erythemis simplicicollis</td>
<td>Freezing</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>10/2010</td>
<td>Islamorada, Florida Keys (FL3 site)</td>
<td>Beach</td>
<td>Anax junius*</td>
<td>Freezing</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
<tr>
<td>11/2010</td>
<td>Guanica, PR</td>
<td>Agricultural fields</td>
<td>Erythrodiplax umbra*, Erythrodiplax sp.</td>
<td>Freezing</td>
<td>Glass beads</td>
<td>Sterivex</td>
</tr>
</tbody>
</table>

*Indicates dragonfly samples where CRESS-DNA viruses were identified.
Table 2 CRESS-DNA viral genomes identified in dragonflies

<table>
<thead>
<tr>
<th>Dragonfly species</th>
<th>Detection method*</th>
<th>Genome ID†</th>
<th>GenBank accession no.</th>
<th>Top BLASTX match (Rep pairwise aa identity, %)</th>
<th>Genus‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. flavescens</em> (Fabricius 1798)</td>
<td>RE digestion (<em>BamHI</em>)</td>
<td>DfCyV-1 (TO-DFpB1-2010)</td>
<td>JX185419</td>
<td>Dragonfly cyclovirus (99 %)</td>
<td>Cyclovirus</td>
</tr>
<tr>
<td><em>P. flavescens</em> (Fabricius 1798)</td>
<td>RE digestion (<em>BamHI</em>)</td>
<td>DfCyV-1 (TO-DFpB3-2010)</td>
<td>JX185420</td>
<td>Dragonfly cyclovirus (99 %)</td>
<td>Cyclovirus</td>
</tr>
<tr>
<td><em>P. flavescens</em> (Fabricius 1798)</td>
<td>RE digestion (<em>BamHI</em>)</td>
<td>DfCyV-1 (TO-DFpB5-2010)</td>
<td>JX185421</td>
<td>Dragonfly cyclovirus (99 %)</td>
<td>Cyclovirus</td>
</tr>
<tr>
<td><em>P. flavescens</em> (Fabricius 1798)</td>
<td>PCR</td>
<td>DfCyV-2 (FL1-NZ38-2010)</td>
<td>JX185422</td>
<td>Bat and human faeces cycloviruses (65 %)</td>
<td>Cyclovirus</td>
</tr>
<tr>
<td><em>P. flavescens</em> (Fabricius 1798)</td>
<td>PCR</td>
<td>DfCyV-2 (FL3-8E-2010)</td>
<td>JX185423</td>
<td>Bat and human faeces cycloviruses (65 %)</td>
<td>Cyclovirus</td>
</tr>
<tr>
<td><em>Anax juneii</em> (Drury 1773)</td>
<td>RE digestion (<em>EcoRV</em>)</td>
<td>DfCyV-3 (FL2-5E-2010)</td>
<td>JX185424</td>
<td>Human faeces cyclovirus PK5034 (48 %)</td>
<td>Cyclovirus</td>
</tr>
<tr>
<td><em>Erythemis simplicicollis</em> (Say 1840)</td>
<td>RE digestion (<em>EcoRV</em>)</td>
<td>DfCyV-4 (BG-NZ46-2007)</td>
<td>JX185425</td>
<td>Bat faeces cycloviruses (90 %)</td>
<td>Cyclovirus</td>
</tr>
<tr>
<td><em>Somatochlora meridionalis</em> (Nielsen 1935)</td>
<td>PCR</td>
<td>DfCyV-5 (PR-6E-2010)</td>
<td>JX185426</td>
<td>Bat faeces cycloviruses (68 %)</td>
<td>Cyclovirus</td>
</tr>
<tr>
<td><em>Erythrodiplax umbrata</em> (Linnaeus 1758)</td>
<td>RE digestion (<em>EcoRV</em>)</td>
<td>DfCyV-5 (PR-NZ47-2010)</td>
<td>JX185427</td>
<td>Bat faeces cycloviruses (68 %)</td>
<td>Cyclovirus</td>
</tr>
<tr>
<td><em>Erythrodiplax umbrata</em> (Linnaeus 1758)</td>
<td>PCR</td>
<td>DfCirV (TO-DF3E-2010)</td>
<td>JX185415</td>
<td>Marine circo-like virus CB-A (36.9 %)</td>
<td>Novel</td>
</tr>
<tr>
<td><em>Erythrodiplax umbrata</em> (Linnaeus 1758)</td>
<td>RE digestion (<em>EcoRV</em>)</td>
<td>DfOrV (TO-DF13-2010)</td>
<td>JX185416</td>
<td>Gull circovirus (30 %)/Milk vetch dwarf nanovirus (34.1 %)</td>
<td>Novel</td>
</tr>
<tr>
<td><em>Diplacodes bipunctata</em> (Brauer 1865)</td>
<td>RE digestion (<em>BamHI</em>)</td>
<td>DfOrV (TO-DF14-2010)</td>
<td>JX185417</td>
<td>Gull circovirus (30 %)/Milk vetch dwarf nanovirus (34.1 %)</td>
<td>Novel</td>
</tr>
<tr>
<td><em>D. bipunctata</em> (Brauer 1865)</td>
<td>RE digestion (<em>BamHI</em>)</td>
<td>DfCyV-1 (FL1-NZ37-2010)</td>
<td>JX185418</td>
<td>Bat faeces TM6c circovirus (43.8 %)</td>
<td>Novel</td>
</tr>
<tr>
<td><em>M. marcella</em> (Selys in Sagra 1857)</td>
<td>RE digestion (<em>XmnI</em>)</td>
<td>DfasCV-1 (FL1-2X-2010)</td>
<td>JX185429</td>
<td>SsHADV-1 like virus from badger faeces (51 %)</td>
<td>Proposed Gemycircularvirus</td>
</tr>
<tr>
<td><em>Erythemis simplicicollis</em> (Say 1840)</td>
<td>RE digestion (<em>XmnI</em>)</td>
<td>DfasCV-2 (FL2-5X-2010)</td>
<td>JX185430</td>
<td>SsHADV-1 like virus from cassava (72 %)</td>
<td>Proposed Gemycircularvirus</td>
</tr>
<tr>
<td><em>P. flavescens</em> (Fabricius 1798)</td>
<td>RE digestion (<em>BamHI</em>)</td>
<td>DfasCV-3 (TO-DFS3B2-2010)</td>
<td>JX185428</td>
<td>Mycovirus SsHADV-1 (44 %)</td>
<td>Proposed Gemycircularvirus</td>
</tr>
<tr>
<td><em>Tramea lacerate</em> (Hagen 1861)</td>
<td>RE digestion (<em>XmnI</em>)</td>
<td>DfasM-1 (FL1-NZ54-2010)</td>
<td>JX185431</td>
<td>Chlamydia phage 4 (29.7 %)</td>
<td>Novel</td>
</tr>
</tbody>
</table>

*Refers to the method used to discover a given virus, including: RCA followed by RE digestion (specific enzyme is listed within parentheses) or degenerate circovirus PCR.
† The origin of the sample (i.e. TO, Kingdom of Tonga; FL, Florida, USA, sites 1–3; BG, Bulgaria; PR, Puerto Rico), laboratory identification name and year of collection is given within parentheses. Genomes that share >95 % nt genome-wide identity have the same name and are in bold.
‡ Several of the viruses detected during this study cannot be classified within known CRESS-DNA viral genera.
archived for more than 4 years, raising the possibility that viral DNA was degraded in these samples.

A total of 17 genomes representing a diverse range of CRESS-DNA viruses were identified in this study (Table 2). More than half of the genomes exhibit similarities to members of the family Circoviridae, mainly to cycloviruses. Although the dominance of circoviruses was expected due to the degenerate circovirus PCR assay utilized, results indicate that there is a diverse community of members of the family Circoviridae associated with dragonflies, regardless of the species or collection location. Since the other CRESS-DNA viruses are similar to groups known to infect fungi and bacteria, it is unlikely that these viruses infect dragonflies or their insect prey. However, it is possible that these viruses infect organisms closely associated with insects (e.g. endosymbionts). Of the 17 CRESS-DNA viral genomes sequenced in this study, 12 genomes were unique (<70% genome-wide identity), while the others were closely related variants sharing >95% identity with some of those genomes (Table 2). The diversity and novelty of viral genomes discovered in dragonflies highlights insects.

![Schematic genome organization of cycloviruses (a) and novel viruses encoding a Rep similar to circoviruses (b) identified in dragonflies.](image)
Fig. 2. Maximum-likelihood phylogenetic tree of the Rep found in eukaryotic CRESS-DNA viruses. Viruses identified in dragonflies during this study are highlighted in grey. Sequences identified in animal faeces (f), tissues (t) or meat (meat) products through metagenomics and degenerate PCR assays, as well as bovine (Bo), rodent (Rod), pig and chimpanzee (Chimp) stool-associated circular viruses (SCVs), are specified. Bar, 0.5 amino acid substitutions per site.
as undersampled organisms that harbour a wealth of CRESS-DNA viruses.

Cycloviruses

The CRESS-DNA viral genomes identified in dragonflies were dominated by cycloviruses, which comprised nine of the 17 genomes recovered. Out of the 12 unique genomes recovered, five exhibited similarities and characteristics consistent with cycloviruses (Table 2, Fig. 1a). Similar to other cycloviruses described to date, the cyclovirus genomes identified in this study are smaller than 2000 nt, ranging from 1711 to 1987 nt, contain two major ORFs encoding a putative Rep and capsid protein, have a small or no intergenic region between the 3' ends of major ORFs and exhibit the circovirus nonanucleotide motif TAGTATTAC at the apex of a potential stem–loop structure, encompassing the origin of replication (ori) in the capsid-encoding strand (Fig. 1a). The putative Reps identified in the majority of dragonfly cycloviruses contain all RCR and helicase motifs important for replication activity that have been identified in known eukaryotic CRESS-DNA viral groups (Table S1, available in JGV Online).

The unique cyclovirus genomes described in this study have been identified as DfCyVs 1–5 (GenBank accession nos JX185419–JX185427; Table 2). Phylogenetic analysis of the predicted Rep sequence of the CRESS-DNA viruses indicates that these five dragonfly viruses cluster with the cyclovirus clade (Fig. 2). DfCyV-1 was identified in dragonflies collected in the Kingdom of Tonga and shares 96 % genome-wide identity to the first DfCyV discovered in the same region (Rosario et al., 2011) (Table S2). In contrast, the other cycloviruses identified in dragonflies from Florida (DfCyV-2 and DfCyV-3), Bulgaria (DfCyV-4) and Puerto Rico (DfCyV-5) share less than 70 % genome-wide identity with DfCyV. The highest similarity among dragonfly cycloviruses was observed between DfCyV-2 and DfCyV-4 (67 % identity) and DfCyV-1 and DfCyV-4 (68 % identity). A 147 nt putative intron was identified in the Rep-encoding region of DfCyV-3 with typical donor (GT) and acceptor (AG) splice sites. Introns (~170 nt) have also been identified in the rep gene of cyclovirus genomes from human faeces and meat products (Li et al., 2010a, 2011). Only one of the cycloviruses, DfCyV-2, was identified in more than one dragonfly species (Pantala flavescens and Anax junius) by RE digestion (Table 2). Screening PCR for the DfCyVs resulted in the detection of the same virus, DfCyV-2, in two additional species, Coryphaeschna ingens and Erythemis simplicicollis, whereas the other cycloviruses were not detected in multiple species. The four dragonfly species in which DfCyV-2 was detected were collected from two locations in Florida, separated by ~370 km. This broad DfCyV-2 detection, in terms of dragonfly species and geographical range, is similar to that of DfCyV, which was detected in three different dragonfly species collected from the Tongan islands ~300 km apart (Rosario et al., 2011), indicating that some cycloviruses can be distributed over considerable distances.

Full genome pairwise comparisons against all cycloviruses reported to date revealed that the DfCyVs are most similar (60–88 % identity) to cycloviruses detected in the faeces of insectivorous bats (Ge et al., 2011; Li et al., 2010b) (Table S2). Amino acid identities to the putative capsid and Rep of cycloviruses identified from bat faeces in China (GenBank accession nos JN377566 and JF938080) were as high as 90 and 93 %, respectively, for the DfCyV-4 genome (Table S3). These findings indicate that the cycloviruses found in bat faeces may infect insects consumed by bats rather than the bats themselves. In addition, well-supported cyclovirus clades containing sequences identified from bats, humans and dragonflies highlight the possibility that cycloviruses identified in human faeces may also originate from insect-contaminated food (Fig. 2). Since limited levels of insect contamination in food products do not represent a public health hazard, humans may consume food products containing a considerable number of insects or insect parts even when monitored by quality control agencies (FDA, 2011). Interestingly, no viruses from the genus Circovirus have been identified in insects or insectivorous bats, only cycloviruses and novel CRESS-DNA viruses, suggesting that members of the genus Circovirus may be limited to vertebrate hosts.

Novel CRESS-DNA viruses containing a circovirus-like Rep

Three unique genomes discovered in dragonflies have significant BLASTX (Altschul et al., 1997) similarities to members of the family Circoviridae based on the Rep (Table 2); however, their genome architecture is different from circoviruses and cycloviruses (Fig. 1b). These novel viruses are divergent from known genomes (Fig. 2) and have been named Dragonfly circularisvirus (DfCirV), Dragonfly orbiculatusvirus (DfOrV) and Dragonfly cyclicosvirus (DfCycV) (GenBank accession nos JX185415, JX185417 and JX185418, respectively; Table 2). DfCirV, DfOrV and DfCycV were only detected by screening PCR in the dragonfly species from which they were originally discovered and share less than 45 % amino acid level identity with the Rep of known viruses available in GenBank. The Reps of these novel viruses contain all RCR and helicase motifs characteristic of eukaryotic CRESS-DNA viruses (Table S1). Although the Rep of DfOrV has significant BLASTX similarities to circoviruses, its best pairwise match is to a nanovirus Rep. However, in contrast with nanoviruses, this genome encodes more than one major ORF. A variety of novel CRESS-DNA viruses with similar characteristics have been identified through metagenomic analyses in environmental samples, animal faeces and a protist (reviewed by Rosario et al., 2012).

Since similarities based solely on the Rep do not always reflect known genome architectures, eukaryotic CRESS-DNA viral genomes have been organized into eight classes.
The second major ORF present in the DfOrV, DfCirV and DfCyclV, suggesting that each of these viruses are widespread in nature. Phylogenetic analysis of environments and pig faeces, suggesting that these types of viruses are widespread in nature. The genome classes account for genomes that exhibit similarities to more than one CRESS-DNA viral group and accommodate cases where similarities to a known REP do not reflect the known genome organization for that group. The novel genome architectures identified in DfOrV, DfCirV and DfCyclV do not cluster with any of the known CRESS-DNA viral families, suggesting that each of these viruses belongs to a previously undescribed genus (Fig. 2).

The second major ORF present in the DfOrV, DfCirV and DfCyclV genomes does not have any significant BLASTX matches to known proteins in GenBank (e-values >1). However, this ORF is likely to encode a structural protein due to the limited protein-encoding capacity of known CRESS-DNA viruses. There are only two major ORFs (>200 aa) in eukaryotic CRESS-DNA viral genomes, specifically in circovirus and REP-encoding geminivirus genomes, which encode a REP and a structural protein. Searches in the Pfam database revealed that the unknown ORFs in DfOrV and DfCyclV genomes have weak similarities to known structural proteins, whereas DfCirV does not have any matches (Fig. 1). The unknown ORF of DfOrV encodes a predicted protein rich in arginine residues within the first 40 aa, similar to circovirus capsid proteins (Niagro et al., 1998) and has weak similarities to the Circo_Cap protein family in Pfam (PF02443; e-value 0.0043). Therefore, even though DfOrV exhibits a novel genome organization, both major ORFs exhibit similarities to circovirus proteins. On the other hand, DfCyclV exhibits similarities to two different groups. Although the REP of this genome exhibits significant similarities to circovirus REPs, the unknown ORF has weak similarities to geminivirus (PF00844; e-value 0.0069) and satellite tobacco necrosis virus (PF03898; e-value 0.0005) capsid protein families in Pfam. Genomes exhibiting similarities to the REP of circovirus but a capsid similar to geminivirus have been identified in mosquitoes and environmental samples (reviewed by Rosario et al., 2012). However, the slightly stronger similarities of the DfCyclV unknown ORF to the capsid protein of a ssRNA satellite virus from the family Tombusviridae was unexpected. This situation is reminiscent of BSL_RDHV, a circovirus-like DNA virus encoding a capsid similar to ssRNA viruses, which was recently discovered in a hot spring through metagenomics (Diemer & Stedman, 2012). Combined, these findings suggest that recombination between unrelated ssDNA and ssRNA viruses may lead to novel species. Notably, BSL_RDHV also exhibits a class V genome organization and encodes a circovirus-like REP, as well as a capsid similar to that of tombusviruses. Although BSL_RDHV was reported after our analysis, pairwise comparisons did not reveal any significant sequence similarities between DfCyclV and BSL_RDHV. Nevertheless, the discovery of another genome exhibiting similarities to both DNA and RNA viruses suggests that these novel hybrid viruses may be more widespread than previously recognized.

**Novel myco-like viruses, proposed genus Gemycircularvirus**

Three different dragonfly species (P. flavescens, Myathiria marcella and E. simplicicollis) collected in distant tropical (Kingdom of Tonga) and subtropical (Florida, USA) locations contained novel viruses with similarities to the gemini-like mycovirus SsHADV-1 (Table 2) (Yu et al., 2010). After their discovery using RCA coupled with RE digestion, these myco-like viruses were only detected by screening PCR in the dragonflies in which they were originally discovered and have been named Dragonfly-associated circular virus -1, -2 and -3 (DfasCV-1, -2 and -3) (GenBank accession numbers JX185428–JX185430; Table 2). Similar viruses have been recently reported in cassava leaf material (cassava-associated circular DNA virus, CasCV) (Dayaram et al., 2012), mosquitoes (mosquito VEM SDBVL-G) (Ng et al., 2011b) and badger faeces (Meles meles faecal virus, MmFV) (van den Brand et al., 2012). This study almost doubles the number of available sequences related to SsHADV-1 and highlights the diverse nature of these viruses associated with insects.

Although SsHADV-1 infects a fungal pathogen, it shares significant similarities with the plant-infecting geminiviruses, specifically mastreviruses. In contrast with mastreviruses, SsHADV-1 has a smaller genome, has a putative origin containing a unique nonanucleotide motif TAATATTAT, does not encode a movement protein and lacks an intron in the REP ORF. All other reported viruses similar to SsHADV-1, including the ones found in dragonflies here, exhibit the same genomic characteristics with the exception of the presence of an intron (Fig. 3a). The introns identified in the REP ORFs of these myco-like viruses vary in length from 166 to 226 nt (Table S1). Similar to mastreviruses, the unspliced (i.e. REP A) and spliced REPs of the novel myco-like viruses share RCR motifs while the spliced REPs also contain helicase motifs (Gutierrez, 1999). SsHADV-1 and the novel myco-like viruses also have a geminivirus REP sequence (GRS) motif (Nash et al., 2011) between RCR motifs II and III that is similar to the one observed in mastreviruses. However, the GRS motif found in the myco-like viruses is unique when compared with geminiviruses. Conserved glutamine, alanine and lysine residues in the C terminus of the GRS motifs (Nash et al., 2011) are not present in SsHADV-1 or the novel myco-like viruses. Nevertheless, in silico structural analyses of the predicted spliced REP of CasCV have shown that this REP is likely to be functionally similar to its geminivirus homologues (Dayaram et al., 2012).
Fig. 3. Schematic genome organization (a), Rep maximum-likelihood (ML) phylogenetic tree (b) and pairwise comparisons (c) showing general features and relationships among the mycovirus SsHADV-1 and myco-like viruses recently identified in cassava plants (CasCV), mosquitoes (mosquito VEM SDBVL-G), badger faeces (MmFV) and dragonflies (DfasCV). In genome schematics (a), the Rep B is depicted in two shades of purple indicating predicted full-length ORFs based on alternative start codons (dark) and truncated ORFs using methionine as the start (light). Myco-like viruses identified in this study are highlighted in grey on the phylogenetic tree in (b). Bar, 0.5 amino acid substitutions per site.
Despite the similarities to mastreviruses, phylogenetic analyses of the Rep show that the novel myco-like viruses form a distinct clade with SsHADV-1 (Fig. 3b). Since these viruses have similar genome organizations, have the same unique nonanucleotide motif in the putative ori (Fig. 3a) and share >55% full genome identity in pairwise comparisons (Fig. 3c), we propose to classify the myco-like viruses under a new genus, Gemycircularvirus. Similar to other eukaryotic CRESS-DNA viral groups, the Reps of gemycircularviruses are more conserved than their capsid proteins based on pairwise comparisons (Fig. 3c). The highest similarity among members of this group was observed between DfasCV-2 and CasCV, which share 67% genome-wide pairwise identity. Both of these viruses share the next highest identity with the mycovirus SsHADV-1 (~64%). Since pathogenic and symbiotic associations between fungi and insects have been well documented (e.g. Dowd, 1992; Hajek & St. Leger, 1994), it is possible that the presence of myco-like DfasCVs in dragonflies collected in Florida (USA) and the Kingdom of Tonga may reflect infected fungi associated with the dragonflies themselves or their insect prey. It is difficult to predict whether all gemycircularviruses discovered to date infect fungi, since SsHADV-1 is the only virus discovered from a culture system and experimentally shown to infect its host source. Nevertheless, gemycircularviruses cluster closely with putative integrated CRESS-DNA viral Rep sequences identified within fungal genomes, suggesting a relationship between fungi and these viruses (Fig. 3b). The discovered DfasCVs have expanded the genetic data regarding these novel gemycircularviruses and this study documents their presence in dragonflies for the first time.

Microphage

A novel microphage genome, named Dragonfly-associated microphage 1 (DfasM-1; Genbank accession no. JX185431), was discovered in the dragonfly species Tramea lacerata and later also detected by PCR in M. marcella specimens collected from the same location in Florida. The novel DfasM-1 genome only contains two ORFs with significant BLASTX matches in GenBank (e-value <0.001), which encode putative major capsid proteins (MCPs) and Reps. The MCP and Rep exhibited 38.5 and 29.7% amino acid pairwise identities, respectively, to members of the subfamily Gokushovirinae, specifically to phages infecting Chlamydia. All three RCR motifs characteristic of phage Reps were identified in the putative DfasM-1 Rep (Table S1; Ilyina & Koonin, 1992). A third ORF in the DfasM-1 genome revealed weak similarities in the Conserved Domain Database (e-value 0.007) to the superfamily of minor capsid proteins (PHA00327) found in gokushoviruses. This protein family, identified as VP2 in Chlamydia gokushoviruses, is analogous to the DNA pilot protein found in microviruses (Liu et al., 2000). Although sequences similar to members of the family Microviridae have been detected in a variety of ecosystems (Rosario & Breitbart, 2011), to our knowledge, this is only the second report of a microphage associated with insects. The only member of the genus Spiromicrovirus, SpV4, infects Spiroplasma melliferum, which is a bacterial pathogen of honeybees (Renaudin & Bové, 1994). It is possible that DfasM-1 infects a bacterial commensal or pathogen associated with dragonflies or their insect prey.

All microphages detected in environmental samples to date resemble gokushoviruses, even though they have been detected in distant environments and potentially infect disparate bacterial hosts (Angly et al., 2006; Desnues et al., 2008; Tucker et al., 2011). These findings support the idea that there are no intermediate species between members of the family Microviridae (Brentlinger et al., 2002), as novel environmental sequences clearly fall within the subfamily Gokushovirinae and only phages infecting enterobacteria form the genus Microvirus. BLAST similarities, in conjunction with the genome size of 4472 nt, suggest that DfasM-1 is also similar to members of the subfamily Gokushovirinae. However, the DfasM-1 genome organization is not analogous with other gokushoviruses (Fig. 4a). In addition, phylogenetic analyses of the MCP revealed that DfasM-1 does not cluster with gokushoviruses, microviruses or the recently discovered Bacteroidetes prophages in the proposed subfamily Alpavirinae (Fig. 4b; Krupovic & Forterre, 2011). Therefore, DfasM-1 potentially represents a member of a novel subfamily, indicating that there are additional groups of the family Microviridae that have not been identified.

Concluding remarks

It has been recently noted that CRESS-DNA viruses are more widespread and diverse than previously recognized (Delwart & Li, 2012; Rosario et al., 2012). Furthermore, CRESS-DNA viruses may have an ancient origin, as endogenous viral Rep sequences have been identified in a variety of organisms ranging from unicellular eukaryotes to vertebrates (Liu et al., 2011). Despite these findings, there has not been an effort to investigate CRESS-DNA viruses circulating among insects, which are considered one of the most evolutionarily successful groups on earth. This study captured a diversity of CRESS-DNA viruses present in dragonflies using simple and affordable methods that allowed the detection of both novel viruses and previously described species. Therefore this study demonstrates that CRESS-DNA viruses circulate widely in dragonflies and that these top-end predators can be used as natural sampling tools to explore the untapped CRESS-DNA viral diversity found among winged insect populations. The data gathered from this study using rapid, targeted methods may help elucidate the evolutionary relationships among known CRESS-DNA viruses and those recently discovered in metagenomic surveys.

Since the CRESS-DNA viruses discovered during this study are similar to viruses thought to infect animals, fungi and bacteria, the reported viruses may infect dragonflies, their insect prey or organisms associated with these insects. Although hosts cannot be established based on the
Fig. 4. Schematic genome organization (a) showing analogous ORFs between members of the family Microviridae and maximum-likelihood phylogenetic tree (b) of representative MCPs. The novel microphage DfasM-1 discovered during this study is highlighted in grey in the phylogenetic tree. Since the BMVs were discovered as integrated phages, sequences were collected from bacterial genome positions reported in the literature (Krupovic & Forterre, 2011). Bar, 0.5 amino acid substitutions per site.
molecular assays used here, this study significantly expands the known diversity of CRESS-DNA viruses. The detection of four novel circovirus species in dragonflies collected in tropical, subtropical and temperate regions supports that circoviruses are commonly associated with insects, in contrast with members of the genus *Circovirus* that have only been detected in vertebrates. The discovery of viruses with Reps similar to circoviruses (DFoRV, DCirV, DCycIV) and microphages (DfasM-1) that display novel genome organizations further support the importance of characterizing full genomes rather than basing classification on a single ORF. DFoRV, DCirV and DCycIV expand the number of known viral genomes with class V genome organization, which has been previously described in viruses from environmental samples and faecal matter (Dzier & Stedman, 2012; Rosario et al., 2012). In addition, this study reports the second virus (DCycIV) to exhibit hybrid similarities between CRESS-DNA and RNA viruses, providing supporting evidence for recombination between these disparate genome types. The discovery of three novel mycock-like viruses allowed for the taxonomic description of a novel group, the hereby proposed genus Gemycircularvirus, which exhibits conserved genomic characteristics. This study also uncovered a divergent prokaryotic CRESS-DNA virus, DfasM-1, suggesting that there are novel lineages within the family *Microviridae*. Overall, this study highlights the diversity of CRESS-DNA viruses present in dragonflies, which reflects the unexplored community of ssDNA viruses circulating among insect populations.

**METHODS**

**Sample collection, virus particle purification and DNA extraction.** Various dragonfly species were collected globally from agricultural fields, lakes, beaches and urban areas using nets (Table 1). Upon collection, samples were preserved by either freezing or circulating among insect populations.

**Detection of CRESS-DNA viruses**

RCA followed by RE digestion. RCA with the phi29 polymerase was used to enrich for small circular ssDNA viruses in the dragonfly DNA extracts (Kim et al., 2008). RCA has been used successfully on a wide variety of samples and efficiently recovers a diversity of circular ssDNA viral genome types from environmental samples and small quantities of tissue (<0.5 g), including plant, vertebrate and insect samples (Habib et al., 2006; Ng et al., 2009a, b, 2011a, b; Rosario et al., 2009, 2011). Therefore, DNA extracts were amplified through RCA for 18 h using an Illustra TempliPhi Amplification kit (GE Healthcare). Since RCA of small circular templates results in the formation of concatemers (Fujii et al., 2006), 2–4 μl TempliPhi products were digested with different REs in separate reactions to obtain complete, unit-length linear genomes that could be cloned and sequenced. This method has been successfully used to clone full-length geminivirus, circovirus, caulimovirus and avihepadnavirus genomes (Hadfield et al., 2011; Massaro et al., 2012; Owor et al., 2007; Piasecki et al., 2012; Shepherd et al., 2008; Varsani et al., 2011). In some instances, TempliPhi products from up to four dragonfly specimens belonging to the same species and collected in the same location were pooled for RE digestion reactions. TempliPhi products were digested with EcoRV, BamHI and XmnI REs and resolved on agarose gels. Approximately 1500–4000 nt fragments were excised from agarose gels, cleaned and cloned into either a pET1.2 vector (Fermentas) for EcoRV and XmnI RE digestion products or BamHI-digested pGEM-3Zf (+) vector (Promega) for BamHI RE digestion products. Cloned genomes were commercially sequenced by primer walking. One of the RE digestion products exhibiting similarities to a microphage only yielded a partial sequence. The full genome sequence of this microphage was obtained by designing back-to-back primers from the partial RE digestion sequence and performing inverse PCR, followed by cloning and sequencing. The inverse PCR experiment proceeded as follows, using the Herculase II Fusion Polymerase (Agilent Technologies) and primers listed in Table S4(a): 95 °C for 4 min, followed by 35 cycles of 95 °C for 20 s, 47 °C incrementally decreasing by 0.1 °C in each cycle for 20 s and 72 °C for 3 min 30 s and a final extension at 72 °C for 5 min.

**Degenerate PCR to detect circoviruses.** Degenerate PCR was used to specifically target circlo-like viruses found in dragonflies. For this purpose, a nested PCR assay developed to amplify the conserved Rep protein of circoviruses in faeces and tissues from a diverse range of vertebrates was used (Li et al., 2010a). Briefly, 2 μl DNA extracts were used in 25 μl reactions for the first PCR round, using the CV-F1 (5'-GGIA YiiCICAYTYTICARGG-3') and CV-R1 (5'-AWCAICCRKTA-RAARTCRTC-3') primers. For the nested PCR assay, 1 μl product from the first PCR round was used in 50 μl reactions with CV-F2 (5'-GGIAYICCAYTTGARGG-3') and CV-R2 (5'-TGYTGTYCCR-TAICCRTCCACCA-3') primers. Both PCR rounds were performed following thermocycling conditions described by Li et al. (2010a). Products approximately 400 bp in size were cloned using a TOPO TA Cloning kit (Invitrogen) and commercially sequenced. The resulting sequence information was used to design back-to-back primers for inverse PCR (see Table S4a for primers). Inverse PCR assays were used to obtain full-length genome products that were then cloned and commercially sequenced by primer walking.

Based on sequence data, specific primers were designed for each of the genomes in order to screen all the dragonflies for the presence of these CRESS-DNA viruses through screening PCR (see Table S4b for primers).

**Data analysis**

**Genome annotation and basic analysis.** All genomes were Sanger-sequenced with a minimum of 2 × coverage and assembled using the Sequencher software (Gene Codes Corporation). Final genome sequences were visualized using SeqBuilder from the Lasergene software package (DNASTAR) and, for annotation purposes, ORFs >100 aa were compared against the GenBank non-redundant database on 20 April 2012 using BLASTX (e-value <0.001) (Altschul et al., 1997).
et al., 1997). Predicted amino acid sequences from ORFs without significant matches in GenBank were compared against the Pfam (Punta et al., 2012) and the Conserved Domain (Marchler-Bauer et al., 2011) databases. Pairwise distances between genomes and proteins were calculated using MEGA5, using default settings and pairwise deletion of gaps (Tamura et al., 2011).

**Phylogenetic analyses.** Full CRESS-DNA viral genomes were downloaded from GenBank on 20 April 2012. For eukaryotic CRESS-DNA viral genomes, the Rep and capsid protein-encoding regions were extracted from each genome. Introns within the Rep-encoding region were identified manually when there were multiple ORFs with matches to Reps. Predicted spliced Reps were only considered if they contained all RCR and helicase motifs characteristic of known eukaryotic CRESS-DNA viral Reps (Rosario et al., 2012) and these spliced Reps were used in phylogenetic comparisons. In the case of prokaryotic CRESS-DNA viruses, analyses were based on representative microphage MCP sequences. All alignments were performed in MEGA5 using the MUSCLE algorithm (Edgar, 2004) and manually edited. Maximum-likelihood phylogenetic trees were constructed with the LG model using PHYML (Guindon et al., 2010), with the approximate likelihood ratio test (aLRT) to assess branch support (Anisimova & Gascuel, 2006). Branches with <50% support were collapsed using Mesquite v. 2.75.

**ACKNOWLEDGEMENTS**

This work was funded by the Marsden Fund of New Zealand (UOC0903) and the National Science Foundation (NSF) (DEB-1025915). K.R. was supported by the East Asia and Pacific Summer Institutes program funded by the NSF (USA) and the Royal Society of New Zealand. J.W. was supported in part by the NSF postdoctoral research fellowship 0804424 and by the Rutgers University start-up funds. The authors would like to thank Mr. Giuliani for coordinating sampling efforts in Puerto Rico. Thanks to Natalia von Ellenrieder, Rosser Garrison and Jurg De Marmels for their help with dragonfly taxonomy and to Michael L. May (Rutgers University, New Jersey) for providing archived dragonfly specimens.

**REFERENCES**


Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J Virol* **84**, 1674–1682.


