High global diversity of cycloviruses amongst dragonflies

Anisha Dayaram,1 Kristen A. Potter,2 Angela B. Moline,3 Dana Drake Rosenstein,4 Milen Marinov,3 John E. Thomas,5,6 Mya Breitbart,7 Karyna Rosario,7 Gerardo R. Argüello-Astorga8 and Arvind Varsani1,9,10

Members of the family Circoviridae, specifically the genus Circovirus, were thought to infect only vertebrates; however, members of a sister group under the same family, the proposed genus Cyclovirus, have been detected recently in insects. In an effort to explore the diversity of cycloviruses and better understand the evolution of these novel ssDNA viruses, here we present five cycloviruses isolated from three dragonfly species (Orthetrum sabina, Xanthocnemis zealandica and Rhionaeschna multicolor) collected in Australia, New Zealand and the USA, respectively. The genomes of these five viruses share similar genome structure to other cycloviruses, with a circular ~1.7 kb genome and two major bidirectionally transcribed ORFs. The genomic sequence data gathered during this study were combined with all cyclovirus genomes available in public databases to identify conserved motifs and regulatory elements in the intergenic regions, as well as determine diversity and recombinant regions within their genomes. The genomes reported here represent four different cyclovirus species, three of which are novel. Our results confirm that cycloviruses circulate widely in winged-insect populations; in eight different cyclovirus species identified in dragonflies to date, some of these exhibit a broad geographical distribution. Recombination analysis revealed both intra- and inter-species recombination events amongst cycloviruses, including genomes recovered from disparate sources (e.g. goat meat and human faeces). Similar to other well-characterized circular ssDNA viruses, recombination may play an important role in cyclovirus evolution.

INTRODUCTION

A wealth of ssDNA viruses is being discovered due to emerging technologies such as metagenomics and next generation sequencing. Although most of these ssDNA viruses are similar in encoding only two ORFs, they have
high diversity in genome organization and in the replicationassociated protein (Rep) (Blinkova et al., 2010; Ge et al., 2011; Kapoor et al., 2012; Kim et al., 2012; Li et al., 2010, 2011; Phan et al., 2011; Rosario et al., 2011, 2012a; Sikorski et al., 2013; van den Brand et al., 2012), reviewed by Rosario et al. (2012b). Within this diverse group of novel ssDNA viruses there is a group of viruses that share significant similarities to the well-characterized circovirus isolates of the family Circoviridae. This, in turn, has led to the proposal of a new genus, Cyclovirus, within the family Circoviridae. Similar to circoviruses, the cycloviruses encode two ORFs that are bidirectionally transcribed, the Rep and the capsid protein (CP). Cycloviruses were first discovered in the stool samples from humans in Pakistan, Tunisia and Nigeria, chimpanzees in Central Africa and meat products of farm animals from Nigeria and Pakistan (Li et al., 2010). Subsequently, cycloviruses have been isolated from insectvorous bats and insects, including various dragonfly species and a Florida woods cockroach (Delwart & Li, 2012; Ge et al., 2011; Li et al., 2011; Padilla-Rodriguez et al., 2013; Rosario et al., 2011, 2012a).

Dragonflies (Odonata: Epiprocta) were the first insects from which circular ssDNA viruses, similar to vertebrate-infecting viruses, were characterized (Rosario et al., 2011). This first report analysed 21 ssDNA viral genomes (~1.7 kb) from three different dragonfly species collected in the Kingdom of Tonga revealing that they represented a single cyclovirus species, and presenting evidence of recombination amongst the reported viral genomes. Subsequently, a study identified diverse cycloviruses in dragonflies from Bulgaria, the USA (Florida) and Puerto Rico indicating the widespread distribution of these viruses (Rosario et al., 2012a). Recently, a novel cyclovirus was described from another insect, Euryctis floridana (Florida woods cockroach-associated cyclovirus isolate GS140; FW/CasCyV-GS140; GenBank accession no. JX569794), suggesting that cycloviruses are indeed widespread amongst insect species (Padilla-Rodriguez et al., 2013).

In addition to cycloviruses, diverse novel ssDNA viruses have been recovered from dragonflies, and thus we have proposed that dragonflies, being top-end insect predators, could be used as ssDNA viral sampling tools in ecosystems by combining their insect-hunting ability with methods that enrich for circular ssDNA viruses (Rosario et al., 2012a). By sampling dragonflies in an agricultural field in Puerto Rico, we identified a novel plant-infecting geminivirus and an associated satellite DNA, including the first mastrevirus and alphasatellite-like molecule, usually associated with begomoviruses, from the Caribbean (Rosario et al., 2013). Insects transmit most plant viruses, and the above example clearly demonstrates the bioaccumulation of insect-transmitted and insect-infecting viruses in dragonflies.

As a continuing effort to determine the diversity and evolution of cycloviruses amongst insects, in this paper we characterize at the genome and protein level five new cycloviruses, three of which represent new species. The reported cycloviruses were characterized from single specimens belonging to three different dragonfly species (Orthetrum sabina, Xanthocnemis zealandica and Rhionaeschna multicolor) collected at Wappa Falls dam, Brisbane (Australia), Lake Pearson, Canterbury (New Zealand) and the Kachina Wetlands of Arizona (USA), respectively. These novel dragonfly cycloviruses, in conjunction with cycloviruses reported in GenBank, were used to perform phylogenetic and recombination analyses amongst members of the proposed genus Cyclovirus. A classification scheme for cycloviruses is proposed based on the analyses presented here.

### RESULTS AND DISCUSSION

#### Viral genome analysis

Five cyclovirus genomes were recovered from three dragonfly species (Fig. 1; Table S1, available in JGV Online). Three genomes were recovered from *R. multicolor* (collected in the USA; 1742–1761 nt), one from *X. zealandica* (New Zealand; 1757 nt) and one from *O. sabina* (Australia; 1764 nt). Other than these five viral genomes, we did not identify any other circular ssDNA viruses within these samples using methods described is this report. All of the genomic features identified in the genomes reported here are characteristic of cycloviruses rather than circoviruses. Circoviruses are slightly larger (1759–2063 nt) than most cycloviruses (1723–1867 nt) (Li et al., 2010). The Rep and CP are transcribed in opposite directions in both virus genera; however, in cycloviruses the putative origin of replication (ori), which exhibits a conserved nonanucleotide motif at the apex of a stem–loop structure, is not found on the Rep-encoding strand, suggesting that the virion strand of cycloviruses and circoviruses is different (Rosario et al., 2012b). Circoviruses and cycloviruses both have a long intergenic region (LIR) that contains the ori; however, cycloviruses lack a short intergenic region that occurs at the 3’ ends of the Rep- and CP-encoding ORFs in circoviruses. The sizes and genome organizations of the viruses detected in dragonflies are similar to that of other cycloviruses. Two major ORFs are arranged in an ambisense organization separated by an LIR (270–339 nt) that contains the putative ori with the conserved nonanucleotide motif (TAGTATTAC), and the CP-encoding ORF is present on the same strand as the putative ori (Fig. 1a). A *BLASTN* (Altschul et al., 1990) analysis of the five full genomes showed that they are most closely related to cycloviruses isolated from chickens, human faecal matter or dragonflies (Table S2). The maximum-likelihood (ML) phylogenetic trees revealed some level of clustering between closely related isolates and indicate that a dragonfly cyclovirus (DCyV-3) is one of the most divergent cycloviruses reported to date (Fig. 1b).

#### Classification of cycloviruses

In an effort to establish an identity threshold for the classification of cyclovirus species, we investigated the distribution of genome-wide pairwise identities amongst
Fig. 1. (a) Genome organization of the cycloviruses recovered from *R. multicolor*, *O. sabina* and *X. Zealandica*. (b) ML phylogenetic tree showing the relationships of the full genomes of all cyclovirus sequences available in GenBank plus those determined in this study (in bold). GenBank accession numbers are shown in square brackets. aLRT, Approximate likelihood-ratio test. Bar, 0.5 nucleotide substitutions per site.
distribution of pairwise identities calculated using SDTv1.0 level cut-off at 76 % genome-wide identity based on the global distribution of highly similar cycloviruses. Please note that the current circovirus species demarcation is based on pairwise identities including gaps as a fifth character state and if this was analysed using our methods with pairwise deletion of gaps, this cut-off would come down to ~78 %. For the purpose of this study, we have demarcated a species level cut-off at 76 % genome-wide identity based on the distribution of pairwise identities calculated using SDTv1.0 (Muhire et al., 2013) (Fig. 2). These criteria, which are similar to the 75 % genome-wide identity demarcation criteria for members of the genus Circovirus (Biagini et al., 2012), could be adopted by other studies.

Based on the circovirus classification scheme proposed here, we expanded the known geographical range of a previously reported circovirus species and identified three novel species. There are two circovirus genomes, which share 99.7 % pairwise identity, isolated from dragonflies collected in Arizona, USA (Fig. 2) that share 86 % pairwise identity with a species isolated from a dragonfly from Bulgaria (DFCyV-4; GenBank accession no. JX183425) and, thus, are tentatively named DFCyV-4 (US-DFKWGX-2012; GenBank accession no. KC512916) and DFCyV-4 (US-DFKWGB-2012; GenBank accession no. KC512917). Interestingly, DFCyV-4 seems to be the same species reported from faecal samples collected from insectivorous bats in China (Bat YN-BtCV3; GenBank accession no. JF938080 and Bat BtCV-01238; JN377566) since they share >85 % genome-wide identity. The remaining three genomes reported in this study share <75 % pairwise identity with all other circoviruses, hence they have been tentatively named DFCyV-6 (US-DFKWGX-2012; GenBank accession no. KC512918), DFCyV-7 (NZ-DFNZ3-2011; GenBank accession no. KCS12919) and DFCyV-8 (AU-DFB007B-2011; GenBank accession no. KC512920). DFCyV-6 shares 72 % pairwise identity with DFCyV-4 and ~60–72 % pairwise identity to all other DFCyVs and cycloviruses. DFCyV-7 and DFCyV-8 share 59–68 % to all other cycloviruses, including all the DFCyVs (Fig. 2). The criteria we have used to classify the cycloviruses from dragonflies could easily be adopted as a basis for classifying all circovirus isolates.

A recent proposal submitted to the International Committee on Taxonomy of Viruses (assigned code: 2011.011a-bbbV) proposes the creation of two subfamilies (cycloviruses and circoviruses) within the family Circoviridae. Each subfamily being further divided into genera, species and isolates. The proposal also suggests the use of CP amino acid sequences as a possibility for the classification of members of the family Circoviridae. It proposes that viral isolates that share between 40 and 75 % identity could be classified as species and those that share 20–40 % identity be members of different genera. We see some limitations with the use of CP sequences on their own, the primary one being the high sequence divergence found amongst CPs (making it difficult to generate robust credible alignments) and a second limitation being recombination. Therefore sequence analysis for taxonomic purposes of segments of genomes is not ideal for viral classification, especially amongst highly recombinant viruses such as most ssDNA viruses. Nonetheless, CP (amino acid identity)-based classification of all the currently sequenced cycloviruses would yield 11 genera and 29 species in contrast to the 25 species we identify using full genome analysis. Without a doubt, as more sequence information becomes available the resolution of the analysis will improve and we should be able to establish a robust classification based on full genomes coupled with CP and Rep sequences.

Recombination analysis

Circular ssDNA viruses are known to have mechanistic predispositions for recombination and are notorious for high recombination rates (Cai et al., 2012; Cheung, 2009; Julian et al., 2012, 2013; Lefeuvre et al., 2009; Martin et al., 2011; Massaro et al., 2012; Mu et al., 2012; Varsani et al., 2011). We found evidence of three intra-species recombination events within cycloviruses, all within the highly sampled DFCyV-1 isolates from the Kingdom of Tonga, similar to those described by Rosario et al. (2011). We also found evidence of inter-species recombination in the remaining diverse cycloviruses (Fig. 3). In most cases we were able to identify ancestral sequences from which the recombinant region was derived, and these results corroborate with our ML phylogenetic analysis of the Rep and CP (Fig. 4). We also identified a highly recombinant viral genome, Human PK5006 (GenBank accession no. GQ404844), in which almost 65 % of its genome is recombinant with recombinant regions derived from Goat Pkgoat11 cyclovirus (HQ738636) and Human PK5222 (GQ404846) ancestral viral sequences. Therefore recombination seems to be a prevalent mechanism in cycloviruses.

Rep and CP analysis

Pairwise amino acid identities were calculated using MUSCLE (Edgar, 2004), and a summary of these are represented in a two-dimensional colour plot in Fig. S1. It is clear from the ML phylogenetic tree (Fig. 4) and pairwise distance analysis of the Rep and CP amino acid sequences that there is greater diversity within the CP compared with the Rep. For example, with the exception of DFCyV-3 all the cyclovirus Reps share >48 % pairwise identity whereas in the case of the CP it is >29 %. The Reps of DFCyV-4 and the two related bat cycloviruses share >93 % pairwise identity whereas the CP share >84 % pairwise identity. The Reps of DFCyV-4, -5, -6 and -7 share >52, 72, 50 and 54 % pairwise identity, respectively, with all other cycloviruses other than with DFCyV-3 (they share ~39, 47, 45 and 44 % pairwise identity). The CPs of DFCyV-4 and DFCyV-5 share ~54 % pairwise identity, whereas the rest of the DFCyV strains share
Fig. 2. (a) Two-dimensional pairwise plot of genome-wide pairwise identities (%) of all the cycloviruses. (b) Distribution of pairwise identities (1485 comparisons) amongst cycloviruses.
pairwise identity of ~34–40%. Interestingly, the CP of DfCyV-7 shares 53% pairwise identity with cycloviruses isolated from human faecal samples from Pakistan and Nigeria (PK5006, GenBank accession no. GQ404844; NG12, GQ404854). BlastP analysis revealed similar results, and these are summarized in Table S2.

Several conserved motifs present in the majority of eukaryotic circular ssDNA virus Reps (reviewed by Rosario et al. (2012b) were identified within the putative Reps of all five cycloviruses reported here, including rolling circle replication (RCR) motifs I, II and III, as well as superfamily 3 (SF3) helicase motifs (Table 1). The function of RCR motif I (FTxNN) is still unknown, however it may be involved in sequence specific recognition of iterons in the LIR (Argüello-Astorga & Ruiz-Medrano, 2001). Details of the iteron sequences found in the LIR are provided in Table 1. RCR motif II (HLQGxxNL for cycloviruses) is thought to coordinate the binding of metal ions (Ilyina & Koonin, 1992). RCR motif III (YCSKxGx for cycloviruses) is believed to be a catalytic site where DNA cleavage takes place during replication initiation (Heyraud-Nitschke et al., 1995; Laufs et al., 1995). SF3 helicase motifs are characterized by a highly conserved domain that contains three motifs within the ~100 aa region, reviewed by Rosario et al. (2012b). The SF3 helicase Walker-A motif (GxxGTGKS for cycloviruses) may act as a deoxyribonucleotide triphosphate (dNTP) binding domain, although it is also thought that this motif may be involved in helicase-like activity in the Rep during RCR. Walker-B and motif C are believed to be involved in controlling helicase activity through the dNTP binding and P-loop NTPase (nucleotide-triphosphatase) domains.

**Sequence analysis of the LIR**

Analysis of the intergenic region of the five cycloviruses sequenced in this study revealed the presence of short iterative sequences (iterons) located close to the stem–loop element containing the conserved nonanucleotide TAGTATTAC (see Table 1 for details of iteron sequences). The iterons are actual or putative Rep-binding sites identified in several groups of circular ssDNA viruses (Argüello-Astorga et al., 1994; Argüello-Astorga & Ruiz-Medrano, 2001; Gutiérrez, 1999; Hanley-Bowdoin et al., 2000; Steinfeldt et al., 2012; Timchenko et al., 2000). The DfCyV-7 iterons exhibited a nucleotide core sequence (CGTCCCAc) different from those of the homologous elements in DfCyV-4, DfCyV-6 and DfCyV-8, all of

---

**Fig. 3.** Recombination analysis results. (a) Schematic of putative recombinant regions. (b) Details of recombination events detected in cyclovirus genomes using detection methods RDP (R), GENECONV (G), BOOTSCAN (B), MAXCHI (M), CHIMAERA (C), SISCAN (S) and 3SEQ (T) implemented in the recombination detection software, RDP 4. Only detection methods with associated P values <0.05 are shown and the given P value is for the detection method highlighted in bold. GenBank accession numbers are shown in square brackets.
which showed iterons with a CGTARC core sequence. In addition, the DfCyV-7 iterative elements exhibited a distinctive arrangement, with one inverted repeat at the 5′ border of the stem–loop element and two direct repeats partially overlapping the right stem arm. In contrast, DfCyV-4, DfCyV-6 and DfCyV-8 displayed three direct repeats, although with a different arrangement in all cases (Fig. 5). Subsequent examination of the intergenic region of all other cycloviruses showed that iterons displaying CGTAAC and/or CGTAGC core sequences are predominant in the proposed genus...
<table>
<thead>
<tr>
<th>Isolate*</th>
<th>Country</th>
<th>Insect species</th>
<th>RCR motifs</th>
<th>SF3 helicase motifs</th>
<th>Iterons†</th>
</tr>
</thead>
<tbody>
<tr>
<td>DfCyV-1</td>
<td>Kingdom of Tonga</td>
<td>Pantala flavescens</td>
<td>FTWNN</td>
<td>GPPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIQGFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKSGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-1</td>
<td>Kingdom of Tonga</td>
<td>P. flavescens</td>
<td>FTWNN</td>
<td>GPPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIQGFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKSGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-1</td>
<td>Kingdom of Tonga</td>
<td>P. flavescens</td>
<td>FTWNN</td>
<td>GPPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIQGFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKSGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-1</td>
<td>Kingdom of Tonga</td>
<td>D. bipunctata</td>
<td>FTWNN</td>
<td>GPPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIQGFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKSGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-1</td>
<td>Kingdom of Tonga</td>
<td>T. tillarga</td>
<td>FTWNN</td>
<td>GPPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIQGFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKSGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-1</td>
<td>Kingdom of Tonga</td>
<td>P. flavescens</td>
<td>FTWNN</td>
<td>GPPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIQGFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKSGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-1</td>
<td>Kingdom of Tonga</td>
<td>D. bipunctata</td>
<td>FTWNN</td>
<td>GPPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIQGFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKSGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-1</td>
<td>Kingdom of Tonga</td>
<td>T. tillarga</td>
<td>FTWNN</td>
<td>GPPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIQGFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKSGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-1</td>
<td>Kingdom of Tonga</td>
<td>T. tillarga</td>
<td>FTWNN</td>
<td>GPPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIQGFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKSGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-2</td>
<td>USA</td>
<td>P. flavescens</td>
<td>FTVNN</td>
<td>GPPGSKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQFANL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-2</td>
<td>USA</td>
<td>Anax junius</td>
<td>FTVNN</td>
<td>GPPGSKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQFANL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-3</td>
<td>USA</td>
<td>Erythematissimplicicollis</td>
<td>FTLN</td>
<td>GDGPAKS VIIDDFYGW</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQFNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCTKGDD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-4</td>
<td>USA</td>
<td>Rhionaeschnamulticolor</td>
<td>FTWNN</td>
<td>GLPGTGS VIIDDFYGW</td>
<td>ITTN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQYANL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKAGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-4</td>
<td>USA</td>
<td>R. multicolor</td>
<td>FTWNN</td>
<td>GLPGTGS VIIDDFYGW</td>
<td>ITTN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQYANL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKAGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-4</td>
<td>Bulgaria</td>
<td>Somatochla meridionalis</td>
<td>FTWNN</td>
<td>GLPGTGS VIIDDFYGW</td>
<td>ITTN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQYANL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKAGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-5</td>
<td>Puerto Rico</td>
<td>Erythrodiplax umbrata</td>
<td>FTLN</td>
<td>GPTSGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCRKSGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-5</td>
<td>Puerto Rico</td>
<td>E. umbrata</td>
<td>FTLN</td>
<td>GPTSGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCRKSGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-6</td>
<td>USA</td>
<td>R. multicolor</td>
<td>FTWNN</td>
<td>GEPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQFANL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKAGK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-7</td>
<td>New Zealand</td>
<td>Xanthochromniszealandica</td>
<td>FTWNN</td>
<td>GAPGTGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKSGI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DfCyV-8</td>
<td>Australia</td>
<td>Orthetrum sabina</td>
<td>FTWNN</td>
<td>GPPGSKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQYANL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YCSKAGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FWCasCyV-8</td>
<td>Australia</td>
<td>Euryticos floridana</td>
<td>FTWNN</td>
<td>GPTSGKS VIIDDFYGW</td>
<td>ITSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLGQFCNL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isolate*</td>
<td>Country</td>
<td>Insect species</td>
<td>RCR motifs</td>
<td>SF3 helicase motifs</td>
<td>Iterons†</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>----------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td>Bat CyCV-TB [HQ738637]</td>
<td>USA</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCKKGNK</td>
</tr>
<tr>
<td>Bat CyV-GF4c [HM228874]</td>
<td>USA</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCSKAGD</td>
</tr>
<tr>
<td>Bat YN-BtCV3 [IF938080]</td>
<td>China</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGYANL</td>
<td>YCSKAGE</td>
</tr>
<tr>
<td>Bat YN-BtCV2 [IF938079]</td>
<td>China</td>
<td>–</td>
<td>FTLNN</td>
<td>HLQGFCNL</td>
<td>YCSKAGN</td>
</tr>
<tr>
<td>Bat YN-BtCV4 [IF938081]</td>
<td>China</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCSKSGD</td>
</tr>
<tr>
<td>Bat YN-BtCV5 [IF938082]</td>
<td>China</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCSKSGI</td>
</tr>
<tr>
<td>Beef PKbeef23 [HQ738634]</td>
<td>Pakistan</td>
<td>–</td>
<td>WTLNN</td>
<td>HLQGFCNL</td>
<td>YCSKSGE</td>
</tr>
<tr>
<td>Chicken NGchicken15 [HQ738644]</td>
<td>Nigeria</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCSKSGI</td>
</tr>
<tr>
<td>Chicken NGchicken8 [HQ738643]</td>
<td>Nigeria</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCSKSGI</td>
</tr>
<tr>
<td>Chimpanzee chimp12 [GQ404850]</td>
<td>Central Africa †</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGYVNL</td>
<td>YCRKSGI</td>
</tr>
<tr>
<td>Goat PKgoat11 [HQ738636]</td>
<td>Pakistan</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCSKSGI</td>
</tr>
<tr>
<td>Human NG14 [GQ404855]</td>
<td>Nigeria</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCSKSGI</td>
</tr>
<tr>
<td>Human NG12 [GQ404854]</td>
<td>Nigeria</td>
<td>–</td>
<td>FTLNN</td>
<td>HLQGFCNL</td>
<td>YCSKAGN</td>
</tr>
<tr>
<td>Human PK5006 [GQ404844]</td>
<td>Pakistan</td>
<td>–</td>
<td>FTWND</td>
<td>HLQGFCNL</td>
<td>YCSKSGI</td>
</tr>
<tr>
<td>Human PK5034 [GQ404845]</td>
<td>Pakistan</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCKKAGH</td>
</tr>
<tr>
<td>Human PK5222 [GQ404846]</td>
<td>Pakistan</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCSKTI</td>
</tr>
<tr>
<td>Human PK5510 [GQ404847]</td>
<td>Pakistan</td>
<td>–</td>
<td>FTWNN</td>
<td>HLQGFCNL</td>
<td>YCSKSGE</td>
</tr>
<tr>
<td>Human TN18 [GQ404858]</td>
<td>Tunisia</td>
<td>–</td>
<td>WTLNN</td>
<td>HLQGFCNL</td>
<td>YCSKSGE</td>
</tr>
</tbody>
</table>

*GenBank accession numbers are shown in square brackets.
†Number of iterons with similar sequences are listed in parentheses.
‡Specimens were collected from Tanzania, Cameroon, Uganda, Rwanda, Central African Republic, Republic of the Congo and the Democratic Republic of the Congo.
Cyclovirus (see Table 1). Indeed, only two cycloviruses besides DfCyV-7 exhibited iterons with a distinctive sequence, namely, DfCyV-3 and FWCasCyV-GS140, displaying CGRcCCC and CGGTACA core sequences, respectively (Fig. 5). Given that a recent theoretical analysis of Rep proteins encoded by diverse groups of circular ssDNA viruses mapped the high-affinity DNA binding specificity determinants (SPDs) to two discrete amino acid clusters surrounding the RCR motifs I and II (London˜o et al., 2010), we decided to analyse the cyclovirus Rep proteins to determine whether there is also a correlation between the iteron sequence and the putative SPDs previously identified in other ssDNA viruses. This new analysis revealed that cycloviruses with iterons exhibiting a CGTARC core sequence encode Rep proteins displaying an SPD-region 1 (SPD-r1) with a TxR sequence and an SPD region 2 (SPD-r2) with a PxR motif (Fig. 6). The only exception to this rule was the DfCyV-5 Rep, exhibiting SPDs with VxR and GxR sequences, respectively (data not shown). The concurrent changes in both the SPD-r1 and SPD-r2 elements of DfCyV-5 Rep suggest the existence of different SPD combinations determining similar high affinity for CGTARC iterons. Interestingly, the Rep proteins encoded by the three cycloviruses (i.e. DfCyV-3, DfCyV-7 and FWCasCyV) possessing distinctive iterons have SPDs that are also unique (Fig. 6), hence supporting the predicted correlation between Rep SPDs and cognate iterons in ssDNA viruses that replicate by an RCR mechanism.

Concluding remarks

Dragonflies have been shown to harbour a diversity of novel circular ssDNA viruses, in particular cycloviruses (Rosario et al., 2012a). Including the genomes described in this study, eight different cyclovirus species have been discovered in dragonflies. Some of these cycloviruses have a widespread geographical distribution, with the broadest distribution observed for DfCyV-4, which has been identified in Bulgaria and the USA. Therefore cycloviruses clearly circulate widely in winged-insect populations. Overall, cycloviruses have been identified in a variety of sample types and organisms, including faeces (bats, humans and chimpanzees), muscle tissues (bats), meat products (cows, chicken, camel, sheep

---

**Fig. 5.** Nucleotide sequence and organization of origin of replication-associated iterative sequences of cycloviruses isolated from insects. Red arrows show the orientation of the iterons with respect to the stem–loop element. Numbers in green denote the nucleotides spanned between a specific iteron and the start codon of the ORF at one end. Lower-case letters in an iterated element indicate a nucleotide that does not match in all the iterons of a virus.
Global diversity of cycloviruses amongst dragonflies diversity of cycloviruses amongst dragonflies

and goats) and insect abdomens (dragonflies and Florida woods cockroach). This contrasts with members of their sister group under the family Circoviridae, genus Circovirus, which have only been associated with vertebrates. Although cycloviruses may have a broader host range than circoviruses, which have only been associated with vertebrates. Although the similarities in their Rep, cycloviruses have distinct genomic characteristics from circoviruses. Since currently there are no species demarcation criteria for the proposed genus Cyclovirus, we took advantage of the diversity of cycloviruses that have been recently reported from dragonflies and other sources to propose a classification scheme. Based on our analyses, a species demarcation cut-off of 76% genome-wide pairwise identity should be used. The characterization of novel insect cycloviruses, whether by sampling viruses bioaccumulated in dragonflies or by directly sampling a variety of insects, will help us better understand evolutionary links within the family Circoviridae as well as improve taxonomic classification criteria.

**METHODS**

**Viral particle purification and DNA extraction.** Adult dragonfly specimens, including O. sabina (n = 1), X. zealandica (n = 1) and R. multicolor (n = 1), were caught using insect nets in Australia (Wappa Falls dam, Queensland), New Zealand (Lake Pearson, Canterbury) and the USA (Kachina Wetlands of Arizona), respectively. The dragonfly samples were preserved either in 95% ethanol or by freezing upon collection. The abdomen from each specimen was used. The characterization of novel insect cycloviruses, whether by sampling viruses bioaccumulated in dragonflies or by directly sampling a variety of insects, will help us better understand evolutionary links within the family Circoviridae as well as improve taxonomic classification criteria.

**Global diversity of cycloviruses amongst dragonflies**

**Fig. 6.** Summary of potential DNA-binding SPDs of cyclovirus Rep proteins. Amino acid identities identified as putative SPDs are shaded. These residues cluster into two discrete regions labelled as SPD-r1 and SPD-r2. The conserved RCR motifs I and II are indicated at the top of the alignments.
resulting supernatant was then filtered in a step-wise fashion through syringe filters (Sartorius Stedim Biotech), first through a 0.45 µm pore size filter and then through a 0.2 µm filter to partially purify virus particles. Viral DNA was then extracted from the filtrate using a High Pure Viral Nucleic Acid kit (Roche).

**Enrichment of circular ssDNA and identification of novel ssDNA viruses.** Circular DNA molecules in the viral DNA extracts were enriched by rolling circle amplification using TempliPhi (GE Healthcare) as described previously (Dayaram et al., 2012; Rosario et al., 2011, 2012a, 2013; Sikorski et al., 2013). The resulting concatenated DNA was digested with BamHI, EcoRI and XmnI in separate reactions yielding between ~1 and 1.7 kb DNA fragments. These fragments were gel purified and cloned into pGEM5ZF (+) (Promega) plasmid restricted with EcoRI or into pUC-19 plasmid vector restricted with BamHI or Smal (Fermentas). The resulting clones were sequenced at Macrogen (South Korea) by primer walking. The genomes of the putative viruses were verified as complete genomes either by designing back-to-back primers followed by PCR amplification of the genome using Kapa HiFi DNA polymerase (Kapa Biosystems), cloning the amplicon and sequencing the recombinant plasmid or by restriction mapping (see Table S1 for details).

**Viral genome and phylogenetic analysis.** The genomic sequences were assembled using DNAMAN (version 5.2.9; Lynnon Biosoft), and preliminary analysis using BLASTX and tBLASTX (Altschul et al., 1990) with GTR+I+G4 as the best substitution model and MAXCHI (Smith, 1992); CHIMAERA (Posada & Crandall, 2001); RDP (Martin & Rybicki, 2005); GENECONV (Padidam et al., 1999); BOOTSCAN (Martin et al., 2005); MAXCHI (Smith, 1992); CHIMAERA (Posada & Crandall, 2001); SISCAN (Gibbs et al., 2000); and 3SEQ (Boni et al., 2007). Only recombination events detected by a minimum of three methods and coupled with clear phylogenetic evidence were considered as putative recombination events.

**ACKNOWLEDGEMENTS**

This work was supported by a Brian Mason trust of New Zealand grant and a Marsden fund of New Zealand grant awarded to A.V. (UOC0903) coupled with personal financial contributions from the various authors of this manuscript. K.R. and M.B. were funded through the US National Science Foundation grant DEB-1239976.

**REFERENCES**


Global diversity of cycloviruses amongst dragonflies diversity of cycloviruses amongst dragonflies


