



# Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae

Arvind Varsani, Mart Krupovic

## ► To cite this version:

Arvind Varsani, Mart Krupovic. Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae. *Virus Evolution*, 2017, 3 (1), pp.vew037. 10.1093/ve/vew037 . pasteur-01977359

**HAL Id: pasteur-01977359**

**<https://pasteur.hal.science/pasteur-01977359>**

Submitted on 10 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family *Genomoviridae*

Arvind Varsani,<sup>1,2,†</sup> and Mart Krupovic<sup>3,\*‡</sup>

<sup>1</sup>The Biodesign Center for Fundamental and Applied Microbiomics, School of Life sciences, Center for Evolution and Medicine, Arizona State University, Tempe, AZ 85287, USA, <sup>2</sup>Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory 7700, South Africa and <sup>3</sup>Unité Biologie moléculaire du Gène chez les Extrêmophiles, Department of Microbiology, Institut Pasteur, 25 rue du Docteur Roux, Paris 75015, France

\*Corresponding authors: E-mails: [krupovic@pasteur.fr](mailto:krupovic@pasteur.fr); [arvind.varsani@asu.edu](mailto:arvind.varsani@asu.edu)

<sup>†</sup><http://orcid.org/0000-0003-4111-2415>

<sup>‡</sup><http://orcid.org/0000-0001-5486-0098>

## Abstract

With the advent of metagenomics approaches, a large diversity of known and unknown viruses has been identified in various types of environmental, plant, and animal samples. One such widespread virus group is the recently established family *Genomoviridae* which includes viruses with small (~2–2.4 kb), circular ssDNA genomes encoding rolling-circle replication initiation proteins (Rep) and unique capsid proteins. Here, we propose a sequence-based taxonomic framework for classification of 121 new virus genomes within this family. Genomoviruses display ~47% sequence diversity, which is very similar to that within the well-established and extensively studied family *Geminiviridae* (46% diversity). Based on our analysis, we establish a 78% genome-wide pairwise identity as a species demarcation threshold. Furthermore, using a Rep sequence phylogeny-based analysis coupled with the current knowledge on the classification of geminiviruses, we establish nine genera within the *Genomoviridae* family. These are *Gemycircularvirus* ( $n = 73$ ), *Gemyduguivirus* ( $n = 1$ ), *Gemygorvirus* ( $n = 9$ ), *Gemykibivirus* ( $n = 29$ ), *Gemykolovirus* ( $n = 3$ ), *Gemykrognavirus* ( $n = 3$ ), *Gemykroznavirus* ( $n = 1$ ), *Gemytondovirus* ( $n = 1$ ), *Gemyvongvirus* ( $n = 1$ ). The presented taxonomic framework offers rational classification of genomoviruses based on the sequence information alone and sets an example for future classification of other groups of uncultured viruses discovered using metagenomics approaches.

**Key words:** *Genomoviridae*; CRESS DNA viruses; replication-associated protein; ssDNA viruses.

## 1. Introduction

Viral metagenomics, fostered by powerful high-throughput sequencing methods, has recently revolutionized our perception of virus diversity in the environment. Many novel groups of uncultivated viruses have been discovered during the past decade, including viruses with small, moderately-sized, and even large genomes (Yau et al. 2011; Roux et al. 2012; Labonte and Suttle,

2013; Dutilh et al. 2014; Yutin et al. 2015; Zhou et al. 2015; Dayaram et al. 2016; Steel et al. 2016). Many of these virus groups remain unclassified. To embrace the constantly growing output from viral metagenomics studies, virus taxonomy is increasingly switching from the traditional classification guided by biological features, such as serology, virion morphology or host range, to predominantly sequence-guided practices (Simmonds et al. 2017). Sequence-guided virus classification is relatively straightforward

when the new viruses fall into existing taxa, with well-defined demarcation criteria. However, in the absence of isolated representatives and established taxonomic framework, rational definition of appropriate taxonomic ranks, such as families, genera, and species, for novel groups of uncultured viruses might be considerably more complex. Solutions to this problem are perhaps most urgently needed in the case of single-stranded (ss) DNA viruses, which are extremely widespread in nature. Due to their small genomes sizes, high mutation and recombination rates (Duffy and Holmes 2008; Duffy and Holmes 2009; Firth et al. 2009; Harkins et al. 2009, 2014; Grigoras et al. 2010; Martin et al. 2011; Streck et al. 2011; Nguyen et al. 2012; Cadar et al. 2013; Roux et al. 2013), and relative ease of genome amplification, an incredible diversity of these viruses has been discovered through metagenomics studies in all conceivable habitats. ssDNA viruses infect cells from all three domains of life and are currently classified by the International Committee on Taxonomy of Viruses (ICTV) into eleven families and one unassigned genus. Members of the families *Microviridae* and *Inoviridae* infect bacteria, viruses of the families *Spiraviridae* and *Pleolipoviridae* prey on archaea, whereas eukaryotes host viruses classified into the families *Anelloviridae*, *Bidnaviridae*, *Circoviridae*, *Geminiviridae*, *Genomoviridae*, *Nanoviridae*, and *Parvoviridae*, and the unassigned genus *Bacilladnavirus*. In addition, several widespread groups of uncultured viruses discovered by viral metagenomics remain unclassified, predominantly those that are circular replication-associated protein encoding single-stranded (CRESS) DNA viruses (Simmonds et al. 2017).

The *Genomoviridae* family is one of the most recently established families of ssDNA viruses (Adams et al. 2016; Krupovic et al. 2016). The family currently includes a single genus *Gemycircularvirus*, which contains a single species, *Sclerotinia gemycircularvirus* 1, encompassing a single isolate, *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1 (SsHADV-1). SsHADV-1 was isolated from a plant-pathogenic fungus *Sclerotinia sclerotiorum* and is the only ssDNA virus known to infect fungi (Yu et al. 2010, 2013). Recently, Liu et al. (2016) have shown that SsHADV-1 is able to infect a mycophagous insect (*Lycoriella ingenua*) which acts as a transmission vector. SsHADV-1 virions are non-enveloped, isometric, 20–22 nm in diameter, and assembled from a single capsid protein (CP) (Yu et al. 2010). The genome is a circular ssDNA molecule of 2,166 nucleotides and contains two genes—for CP and rolling-circle replication initiation protein (Rep). Like in many other ssDNA viruses with circular genomes, the large intergenic region of SsHADV-1 contains a potential stem-loop structure with a nonanucleotide (TAAATATTAT) motif at its apex, which is likely to be important for rolling-circle replication initiation. The CP of SsHADV-1 is not recognizably similar to the corresponding proteins from viruses in other taxa. Although SsHADV-1 remains the only isolated and classified member of the *Genomoviridae*, 121 viral genomes with varying degree of similarity to that of SsHADV-1 have been recovered and sequenced from various environmental, plant- and animal-associated samples, indicating that these viruses are widespread and abundant in the environment (Table 1). However, a proper taxonomic framework and demarcation criteria necessary to accommodate these viruses within the family *Genomoviridae* are lacking. Here, we explore the diversity and evolution of uncultured SsHADV-1-like viruses and attempt to establish a framework for their classification based on sequence data alone.

## 2. Genomoviridae diversity and species classification

At the time of the analysis (August, 2016), there were 121 SsHADV-1-like genome sequences in the GenBank database.

Each of these genomes encodes two putative proteins homologous to the CP and Rep of SsHADV-1, highlighting strong coherence of this virus assemblage. Nevertheless, there is a considerable sequence divergence within the group (Supplementary Fig. S1). To investigate the extent of genomoviral sequence diversity, we analyzed the distribution of genome-wide pairwise identities (one minus Hamming distances of pairwise aligned sequences with pairwise deletion of gaps) across all 121 available genomes (Fig. 1A) using SDT v1.2 (Muhire, Varsani, and Martin 2014). Most of the virus genomes in our dataset share 56–66% genome-wide pairwise identities and only a handful contained nearly identical relatives ( $\geq 98\%$  identity), indicating that sequence diversity among SsHADV-1-like viruses remains largely unexplored.

Pairwise comparison of the Rep and CP protein sequences revealed a broader distribution of identity values (Fig. 1B and C). Notably, the CPs were considerably more divergent than the Reps, with the highest proportion of pairwise identities being  $\sim 33\%$  (versus  $\sim 48\%$  for the Rep). This observation is in line with functional differences of the two proteins and the fact that viral CPs often encompass host recognition determinants which are under constant pressure to co-evolve with the cellular receptors (Kolawole et al. 2014; Shangjin, Cortey, and Segales, 2009). Based on the analysis of distribution of the pairwise identities across genomes, CPs and Reps, we consider a threshold of 78% to be a conservative value for species demarcation. Thus, all viral genomes showing identities higher than this value should be considered as variants of the existing species. Nonetheless, there may be situations where it is difficult to assign species because a particular new sequence is

1.  $>78\%$  similar to sequences from a particular species but is  $<78\%$  similar to other variants of that same species;
2.  $>78\%$  similar to sequences from two or more different species.

To resolve the above conflicts, we suggest adopting a similar approach proposed for geminiviruses (Muhire et al. 2013; Varsani et al. 2014a, b; Brown et al. 2015). To resolve conflict 1, we suggest that the new sequence be classified within any species in which it shares  $>78\%$  identity to any one variant formerly classified as belonging to that species, even if it is  $<78\%$  identical to other viruses within that species. To resolve conflict 2, we suggest that the new sequence be considered as belonging to the species with sequences with which it shares the highest degree of similarity.

## 3. Rep-based approach for creation of genera

Maximum likelihood phylogenetic analyses based on the Rep of 121 genomoviruses revealed several well-supported clades that could be considered as genera within the family (Fig. 2). We note that the clades obtained in the Rep-based phylogeny are not fully consistent with those obtained in the phylogenetic analysis of the full genome or the more diverse CP sequences (Figs 3 and 4). This is most explicit in the case of the newly proposed genus *Gemykolovirus* (see below). In the Rep-based tree corresponding sequences form a sister clade to the single representative of the genus *Gemyduguivirus* (Fig. 2). In contrast, in the whole-genome-based phylogeny, *gemykoloviruses* form a sister group to members of the genus *Gemycircularvirus* (Fig. 3). The reason for this incongruence is likely to be intra-familial recombination between different genomovirus genomes resulting in chimeric entities encoding Rep and CP with different evolutionary histories (Kraberger et al. 2015a). Indeed, in the CP-based

Table 1. Details of all members of the Genomoviridae.

| Genus             | Species                                    | Accession # | Sequence ID                 | Isolation source         | Common name            | Sample type               | Country     | Reference                    |
|-------------------|--------------------------------------------|-------------|-----------------------------|--------------------------|------------------------|---------------------------|-------------|------------------------------|
| Gemycircularvirus | Blackbird associated gemycircularvirus 1   | KF371641    | P9                          | Turdus merula            | Blackbird              | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                   | Blackbird associated gemycircularvirus 1   | KF371642    | P22                         | Turdus merula            | Blackbird              | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                   | Blackbird associated gemycircularvirus 1   | KF371643    | as41                        | Ovis aries               | Sheep                  | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                   | Bovine associated gemycircularvirus 1      | KT862253    | 52 Fec78023 cow             | Bos taurus               | Cow                    | Faeces                    | New Zealand | Steel et al. (2016)          |
|                   | Bromus associated gemycircularvirus 1      | KM510192    | BasCV-3 NZ-NZG01 Sef-2012   | Bromus hordeaceus        | Soft brome/Bull grass  | Leaf                      | New Zealand | Kraberger et al. (2015b)     |
|                   | Cassava associated gemycircularvirus 1     | JQ412056    | G14                         | Manihot esculenta        | Cassava                | Leaf                      | Ghana       | Dayaram et al. (2012)        |
|                   | Cassava associated gemycircularvirus 1     | JQ412057    | G5                          | Manihot esculenta        | Cassava                | Leaf                      | Ghana       | Dayaram et al. (2012)        |
|                   | Chickadee associated gemycircularvirus 1   | KT309029    | 254065908                   | Poecile atricapillus     | Black-capped chickadee | Buccal and cloacal swab   | USA         | Hanna et al. (2015)          |
|                   | Chicken associated gemycircularvirus 1     | KT862243    | 27 Fec79971 chicken         | Gallus gallus domesticus | Chicken                | Faeces                    | New Zealand | Steel et al. (2016)          |
|                   | Chicken associated gemycircularvirus 1     | KT862244    | 29 Fec79971 llama           | Lama glama               | Llama                  | Faeces                    | New Zealand | Steel et al. (2016)          |
|                   | Chicken associated gemycircularvirus 1     | KT862246    | 30 Fec79971 horse           | Equus ferus caballus     | Horse                  | Faeces                    | New Zealand | Steel et al. (2016)          |
|                   | Chicken associated gemycircularvirus 2     | KT862242    | 27 Fec16497 chicken         | Gallus gallus domesticus | Chicken                | Faeces                    | New Zealand | Steel et al. (2016)          |
|                   | Dragonfly associated gemycircularvirus 1   | JX185429    | FL2-5X-2010                 | Erythemis simplicicollis | Dragonfly              | Abdomen                   | USA         | Rosario et al. (2012)        |
|                   | Equine associated gemycircularvirus 1      | KT862248    | 30 Fec80061 horse           | Equus ferus caballus     | Horse                  | Faeces                    | New Zealand | Steel et al. (2016)          |
|                   | Fur seal associated gemycircularvirus 1    | KF371638    | as50                        | Arctocephalus forsteri   | New Zealand fur seal   | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                   | Fur seal associated gemycircularvirus 1    | KT862241    | 27 Fec1 chicken             | Gallus gallus domesticus | Chicken                | Faeces                    | New Zealand | Steel et al. (2016)          |
|                   | Gerygone associated gemycircularvirus 1    | KF371636    | P24a                        | Gerygone albofrontata    | Chatham Island warbler | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                   | Gerygone associated gemycircularvirus 2    | KF371637    | P24b                        | Gerygone albofrontata    | Chatham Island warbler | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                   | Gerygone associated gemycircularvirus 3    | KF371639    | P24c                        | Gerygone albofrontata    | Chatham Island warbler | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                   | Hypericum associated gemycircularvirus 1   | KF413620    | VNHJ1W                      | Hypericum japonicum      | Hypericum              | Leaf                      | Vietnam     | Du et al. (2014)             |
|                   | Lama associated gemycircularvirus 1        | KT862245    | 29 Fec80018 llama           | Lama glama               | Llama                  | Faeces                    | New Zealand | Steel et al. (2016)          |
|                   | Lama associated gemycircularvirus 1        | KT862247    | 30 Fec80018 horse           | Equus ferus caballus     | Horse                  | Faeces                    | New Zealand | Steel et al. (2016)          |
|                   | Mallard associated gemycircularvirus 1     | KF371635    | as24                        | Anas platyrhynchos       | Mallard duck           | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                   | Miniopterus associated gemycircularvirus 1 | KJ641719    | BtMF-CV-23/GD2012           | Miniopterus fuliginosus  | Bat                    | Pharyngeal & rectal swabs | China       | Wu et al. (2016)             |
|                   | Mongoose associated gemycircularvirus 1    | KP263547    | 478d                        | Herpestes ichneumon      | Egyptian mongoose      | Faeces                    | Portugal    | Conceicao-Neto et al. (2015) |
|                   | Mosquito associated gemycircularvirus 1    | HQ335086    | SDBVL G                     | Culex erythrorhox        | Mosquito               | Mosquito samples          | USA         | Ng et al. (2011)             |
|                   | Odonata associated gemycircularvirus 1     | KM598385    | OdaGmV-1-US-260BC-12        | Ischnura posita          | Damselfly              | Abdomen                   | USA         | Dayaram et al. (2015)        |
|                   | Odonata associated gemycircularvirus 1     | KM598386    | OdaGmV-1-US-260SR1-12       | Pantala hymenaea         | Dragonfly              | Abdomen                   | USA         | Dayaram et al. (2015)        |
|                   | Odonata associated gemycircularvirus 2     | KM598387    | OdaGmV-2-US-1642KW-12       | Aeshna multicolor        | Dragonfly              | Abdomen                   | USA         | Dayaram et al. (2015)        |
|                   | Odonata associated gemycircularvirus 2     | KM598388    | OdaGmV-2-US-1634LM2-12      | Libellula saturata       | Dragonfly              | Abdomen                   | USA         | Dayaram et al. (2015)        |
|                   | Poaceae associated gemycircularvirus 1     | KT253577    | PaGmV-1 TO STO14-29204 2014 | Brachiaria deflexa       | Signalgrass            | Leaf                      | Tonga       | Male et al. (2015)           |
|                   | Poaceae associated gemycircularvirus 1     | KT253578    | PaGmV-1 TO STO15-29204 2014 | Brachiaria deflexa       | Signalgrass            | Leaf                      | Tonga       | Male et al. (2015)           |
|                   | Poaceae associated gemycircularvirus 1     | KT253579    | PaGmV-1 TO STO18-29204 2014 | Brachiaria deflexa       | Signalgrass            | Leaf                      | Tonga       | Male et al. (2015)           |
|                   | Porcine associated gemycircularvirus 1     | KT862250    | 49 Fec80061 pig             | Sus scrofa domestica     | Pig                    | Faeces                    | New Zealand | Steel et al. (2016)          |
|                   | Porcine associated gemycircularvirus 2     | KF371640    | as5                         | Sus scrofa               | Domestic pig           | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                   | Pteropus associated gemycircularvirus 1    | KT732804    | That 45285                  | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 1    | KT732805    | That 47364                  | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 2    | KT732792    | That 103791                 | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 2    | KT732793    | That A 103791               | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 3    | KT732797    | That A 103852               | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 4    | KT732814    | That H 103806               | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 5    | KT732801    | That 12377                  | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 5    | KT732802    | That H 12377                | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 6    | KT732796    | That H 103639               | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 6    | KT732803    | That 103951                 | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 7    | KT732807    | That A 103746               | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |
|                   | Pteropus associated gemycircularvirus 7    | KT732808    | That A 103909               | Pteropus tonganus        | Bat                    | Faeces                    | Tonga       | Male et al. (2016)           |

(continued)

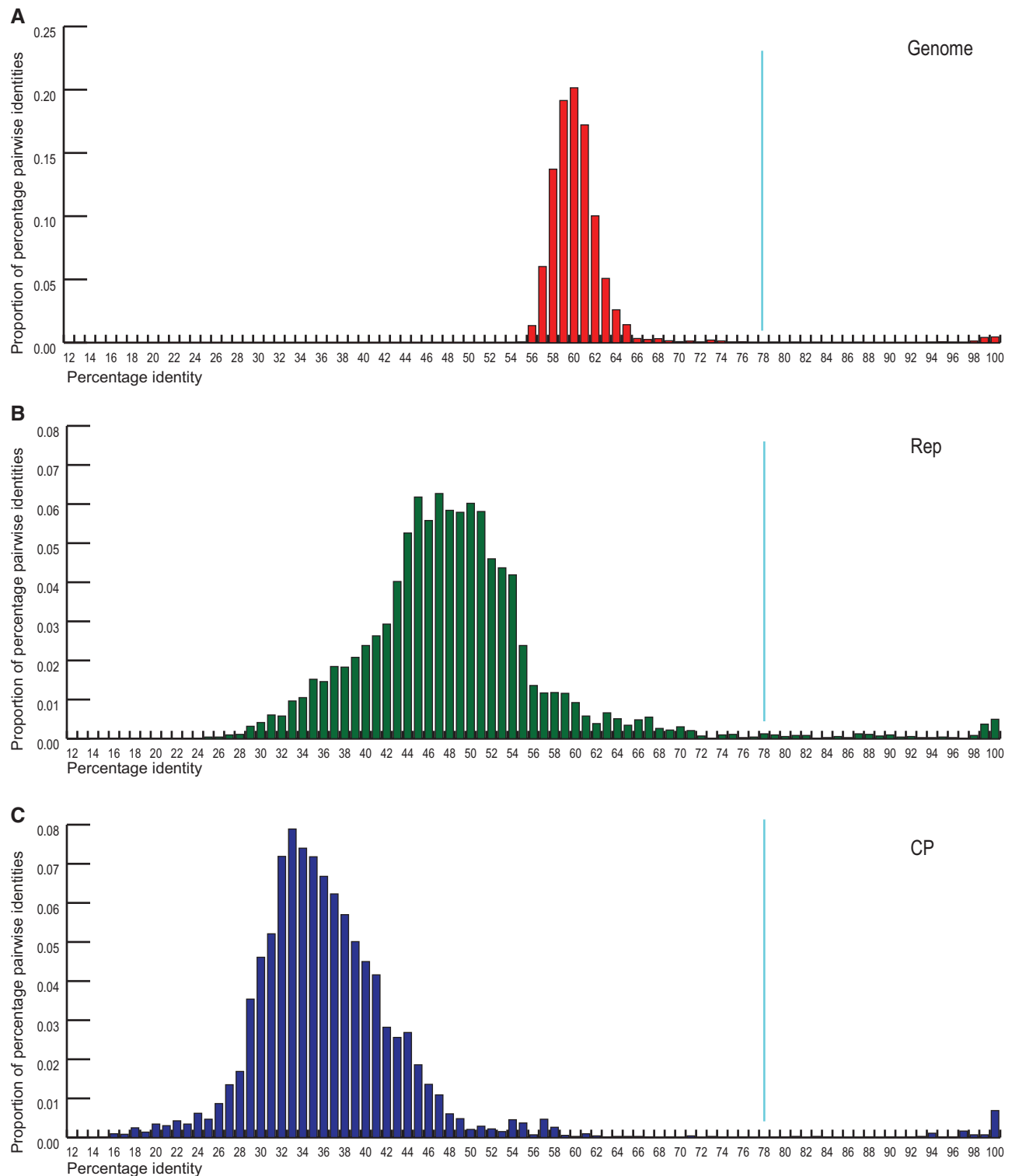
Table 1. Continued

| Genus                         | Species                                  | Accession # | Sequence ID            | Isolation source         | Common name                | Sample type         | Country      | Reference                    |
|-------------------------------|------------------------------------------|-------------|------------------------|--------------------------|----------------------------|---------------------|--------------|------------------------------|
| Gemydguivirus<br>Gemygrovirus | Pteropus associated gemycircularvirus 7  | KT732809    | Tbat H 103746          | Pteropus tonganus        | Bat                        | Faeces              | Tonga        | Male et al. (2016)           |
|                               | Pteropus associated gemycircularvirus 7  | KT732810    | Tbat H 103909          | Pteropus tonganus        | Bat                        | Faeces              | Tonga        | Male et al. (2016)           |
|                               | Pteropus associated gemycircularvirus 7  | KT732811    | Tbat L 103746          | Pteropus tonganus        | Bat                        | Faeces              | Tonga        | Male et al. (2016)           |
|                               | Pteropus associated gemycircularvirus 8  | KT732806    | Tbat 31579             | Pteropus tonganus        | Bat                        | Faeces              | Tonga        | Male et al. (2016)           |
|                               | Pteropus associated gemycircularvirus 9  | KT732795    | Tbat 21383             | Pteropus tonganus        | Bat                        | Faeces              | Tonga        | Male et al. (2016)           |
|                               | Pteropus associated gemycircularvirus 10 | KT732794    | Tbat H 103958          | Pteropus tonganus        | Bat                        | Faeces              | Tonga        | Male et al. (2016)           |
|                               | Rat associated gemycircularvirus 1       | KR912221    | Ch-zjrat-01            | Rattus norvegicus        | Rat                        | Blood               | China        | Li et al. (2015)             |
|                               | Sclerotinia gemycircularvirus 1          | GQ365709    | SsHADV-1 CN            | Sclerotinia sclerotiorum | Sclerotinia                | Mycelial samples    | China        | Yu et al. (2010)             |
|                               | Sclerotinia gemycircularvirus 1          | KF268025    | SsHADV-1 NZ H6 2012    | River Sediments          | -                          | River Sediments     | New Zealand  | Kraberger et al. (2013)      |
|                               | Sclerotinia gemycircularvirus 1          | KF268026    | SsHADV-1 NZ SR1 2012   | River Sediments          | -                          | River Sediments     | New Zealand  | Kraberger et al. (2013)      |
|                               | Sclerotinia gemycircularvirus 1          | KF268027    | SsHADV-1 NZ SR3 2012   | River Sediments          | -                          | River Sediments     | New Zealand  | Kraberger et al. (2013)      |
|                               | Sclerotinia gemycircularvirus 1          | KF268028    | SsHADV-1 NZ SR5 2012   | River Sediments          | -                          | River Sediments     | New Zealand  | Kraberger et al. (2013)      |
|                               | Sclerotinia gemycircularvirus 1          | KM598382    | SsHADV-1-US-549LB-12   | Ischnura ramburii        | Damselfly                  | Abdomen             | USA          | Dayaram et al. (2015)        |
|                               | Sclerotinia gemycircularvirus 1          | KM598383    | SsHADV-1-US-549DFS-12  | Erythemis simplicicollis | Dragonfly                  | Abdomen             | USA          | Dayaram et al. (2015)        |
|                               | Sclerotinia gemycircularvirus 1          | KM598384    | SsHADV-1-US-549SR-12   | Pantala hymenaea         | Dragonfly                  | Abdomen             | USA          | Dayaram et al. (2015)        |
|                               | Sewage derived gemycircularvirus 1       | KJ547638    | BS3917                 | Sewage oxidation pond    | -                          | Sewage              | New Zealand  | Kraberger et al. (2015a)     |
|                               | Sewage derived gemycircularvirus 1       | KM821747    | SaGmV-1 NZ-BS3970-2012 | Sewage oxidation pond    | -                          | Sewage              | New Zealand  | Kraberger et al. (2015a)     |
|                               | Sewage derived gemycircularvirus 2       | KJ547641    | BS4117                 | Sewage oxidation pond    | -                          | Sewage              | New Zealand  | Kraberger et al. (2015a)     |
|                               | Sewage derived gemycircularvirus 3       | KJ547636    | BS4014                 | Sewage oxidation pond    | -                          | Sewage              | New Zealand  | Kraberger et al. (2015a)     |
|                               | Sewage derived gemycircularvirus 4       | KJ547637    | BS3939                 | Sewage oxidation pond    | -                          | Sewage              | New Zealand  | Kraberger et al. (2015a)     |
|                               | Sewage derived gemycircularvirus 4       | KJ547640    | BS3972                 | Sewage oxidation pond    | -                          | Sewage              | New Zealand  | Kraberger et al. (2015a)     |
|                               | Sewage derived gemycircularvirus 5       | KJ547639    | BS3970                 | Sewage oxidation pond    | -                          | Sewage              | New Zealand  | Kraberger et al. (2015a)     |
| Gemydguivirus<br>Gemygrovirus | Sheep associated gemycircularvirus 1     | KT862249    | 47 Fec8064 sheep       | Ovis aries               | Sheep                      | Faeces              | New Zealand  | Steel et al. (2016)          |
|                               | Sheep associated gemycircularvirus 1     | KT862251    | 51 Fec8064 sheep       | Ovis aries               | Sheep                      | Faeces              | New Zealand  | Steel et al. (2016)          |
|                               | Soybean associated gemycircularvirus 1   | KT598248    | SLaGemV1-1             | Glycine max              | Soybean                    | Leaf                | USA          | Marzano and Domier, (2016)   |
|                               | Canine associated gemydguivirus 1        | JX185428    | T0-DFS3B2-2010         | Pantala flavescens       | Dragonfly                  | Abdomen             | Tonga        | Rosario et al. (2012)        |
|                               | Mallard associated gemygrovirus 1        | JN704610    | 4 Fec7 duck            | Canis lupus familiaris   | Dog                        | Faeces              | New Zealand  | Steel et al. (2016)          |
|                               | Mallard associated gemygrovirus 1        | KT862238    | 4 Fec7 duck            | Meles meles              | Duck                       | Rectal swab         | Netherlands  | van den Brand et al. (2012)  |
|                               | Mallard associated gemygrovirus 1        | KT862239    | 24 Fec7 duck           | Anas platyrhynchos       | Duck                       | Faeces              | New Zealand  | Steel et al. (2016)          |
|                               | Pteropus associated gemygrovirus 1       | KT732790    | Tbat A 103952          | Anas platyrhynchos       | Bat                        | Faeces              | New Zealand  | Steel et al. (2016)          |
|                               | Pteropus associated gemygrovirus 1       | KT732791    | Tbat H 103952          | Pteropus tonganus        | Bat                        | Faeces              | Tonga        | Male et al. (2016)           |
|                               | Sewage derived gemygrovirus 1            | KJ413144    | 349                    | Pteropus tonganus        | Bat                        | Faeces              | Tonga        | Male et al. (2016)           |
| Gemykibivirus                 | Sewage derived gemygrovirus 1            | KJ547635    | BS3963                 | Homo sapiens             | Human                      | Cervical sample     | South Africa | Kraberger et al. (2015a)     |
|                               | Starling associated gemygrovirus 1       | KF371632    | P14                    | Sewage oxidation pond    | -                          | Sewage              | New Zealand  | Sikorski et al. (2013)       |
|                               | Badger associated gemykibivirus 1        | KF263543    | 588t                   | Sturnus vulgaris         | European starling          | Faeces              | New Zealand  | Sikorski et al. (2013)       |
|                               | Black robin associated gemykibivirus 1   | KF371634    | P21                    | Meles meles              | European badger            | Faeces              | Portugal     | Conceicao-Neto et al. (2015) |
|                               | Blackbird associated gemykibivirus 1     | KF371633    | P22                    | Petroica traversi        | Chatham Island black robin | Faeces              | New Zealand  | Sikorski et al. (2013)       |
|                               | Bovine associated gemykibivirus 1        | LK931483    | HCB18.215              | Turdus merula            | Blackbird                  | Faeces              | New Zealand  | Sikorski et al. (2013)       |
|                               | Dragonfly associated gemykibivirus 1     | JX185430    | FL1-2X-2010            | Bos taurus               | Cow                        | Serum               | Germany      | Lamberto et al. (2014)       |
|                               | Human associated gemykibivirus 1         | KJ547644    | BS3980                 | Mithyria marcella        | Dragonfly                  | Abdomen             | USA          | Rosario et al. (2012)        |
|                               | Human associated gemykibivirus 1         | KJ547645    | BS3849                 | Sewage oxidation pond    | -                          | Sewage              | New Zealand  | Kraberger et al. (2015a)     |
|                               | Human associated gemykibivirus 1         | KP974694    | DB2                    | Sewage oxidation pond    | -                          | Sewage              | New Zealand  | Kraberger et al. (2015a)     |
| Gemykibivirus                 | Human associated gemykibivirus 1         | LK931485    | MSS12.225              | Homo sapiens             | Human                      | Plasma              | Germany      | Zhang et al. (2016)          |
|                               | Human associated gemykibivirus 1         | KP133075    | SL1                    | Homo sapiens             | Human                      | Blood               | Germany      | Lamberto et al. (2014)       |
|                               | Human associated gemykibivirus 2         | KP133076    | SL2                    | Homo sapiens             | Human                      | Cerebrospinal fluid | Germany      | Lamberto et al. (2014)       |
|                               | Human associated gemykibivirus 2         | KP133077    | SL3                    | Homo sapiens             | Human                      | Cerebrospinal fluid | Sri Lanka    | Phan et al. (2015)           |
|                               | Human associated gemykibivirus 2         | KP133077    | SL3                    | Homo sapiens             | Human                      | Cerebrospinal fluid | Sri Lanka    | Phan et al. (2015)           |
|                               | Human associated gemykibivirus 2         | KP133077    | SL3                    | Homo sapiens             | Human                      | Cerebrospinal fluid | Sri Lanka    | Phan et al. (2015)           |

(continued)

Table 1. Continued

| Genus           | Species                                | Accession # | Sequence ID         | Isolation source          | Common name       | Sample type               | Country     | Reference                    |
|-----------------|----------------------------------------|-------------|---------------------|---------------------------|-------------------|---------------------------|-------------|------------------------------|
| Gemyklovirus    | Human associated gemykibivirus 2       | KP133078    | BZ1                 | Homo sapiens              | Human             | Faeces                    | Brazil      | Phan et al. (2015)           |
|                 | Human associated gemykibivirus 2       | KP133079    | BZ2                 | Homo sapiens              | Human             | Faeces                    | Brazil      | Phan et al. (2015)           |
|                 | Human associated gemykibivirus 2       | KP133080    | NP                  | Untreated sewage          | –                 | Sewage                    | Nepal       | Phan et al. (2015)           |
|                 | Human associated gemykibivirus 3       | KP263546    | 541c                | Herpestes ichneumon       | Egyptian mongoose | Faeces                    | Portugal    | Conceicao-Neto et al. (2015) |
|                 | Human associated gemykibivirus 3       | KP987887    | GemyC1c             | Homo sapiens              | Human             | Plasma                    | France      | Uch et al. (2015)            |
|                 | Human associated gemykibivirus 4       | KT363839    | GeT21               | Homo sapiens              | Human             | Cerebrospinal fluid       | China       | Zhou et al. (2015)           |
|                 | Human associated gemykibivirus 5       | KU343137    | HV-GcV2             | Homo sapiens              | Human             | Pericardial fluid         | France      | Halary et al. (2016)         |
|                 | Mongoose associated gemykibivirus 1    | KP263545    | 160b                | Herpestes ichneumon       | Egyptian mongoose | Faeces                    | Portugal    | Conceicao-Neto et al. (2015) |
|                 | Pteropus associated gemykibivirus 1    | KT732813    | Tbat A 64418        | Pteropus tonganus         | Bat               | Faeces                    | Tonga       | Male et al. (2016)           |
|                 | Rhinolophus associated gemykibivirus 1 | KJ641737    | BtRh-CV-6/Tibet2013 | Rhinolophus hipposideros  | Bat               | Pharyngeal & rectal swabs | China       | Wu et al. (2016)             |
| Gemykrovirus    | Rhinolophus associated gemykibivirus 1 | KP263544    | 181a                | Herpestes ichneumon       | Egyptian mongoose | Faeces                    | Portugal    | Conceicao-Neto et al. (2015) |
|                 | Rhinolophus associated gemykibivirus 2 | KJ641726    | BtRF-CV-8/NM2013    | Rhinolophus ferrumequinum | Bat               | Pharyngeal & rectal swabs | China       | Wu et al. (2016)             |
|                 | Sewage derived gemykibivirus 1         | KJ547643    | BS4149              | Sewage oxidation pond     | –                 | Sewage                    | New Zealand | Kraberger et al. (2015a)     |
|                 | Sewage derived gemykibivirus 1         | KT862240    | 27 BS14149 chicken  | Gallus gallus domesticus  | Chicken           | Faeces                    | New Zealand | Steel et al. (2016)          |
|                 | Sewage derived gemykibivirus 1         | KT862252    | 52 BS14149 cow      | Bos taurus                | Cow               | Faeces                    | New Zealand | Steel et al. (2016)          |
|                 | Sewage derived gemykibivirus 1         | KT862255    | 56 BS14149 hare     | Lepus europaeus           | Hare              | Faeces                    | New Zealand | Steel et al. (2016)          |
|                 | Sewage derived gemykibivirus 2         | KJ547642    | BS3911              | Sewage oxidation pond     | –                 | Sewage                    | New Zealand | Kraberger et al. (2015a)     |
|                 | Pteropus associated gemykibivirus 1    | KT732798    | Tbat A 103779       | Pteropus tonganus         | Bat               | Faeces                    | Tonga       | Male et al. (2016)           |
|                 | Pteropus associated gemykibivirus 1    | KT732799    | Tbat H 103779       | Pteropus tonganus         | Bat               | Faeces                    | Tonga       | Male et al. (2016)           |
|                 | Pteropus associated gemykibivirus 2    | KT732800    | Tbat H 103921       | Pteropus tonganus         | Bat               | Faeces                    | Tonga       | Male et al. (2016)           |
| Gemykroznavirus | Bovine associated gemykrovirus 1       | LK931484    | HCB19.212           | Bos taurus                | Cow               | Serum                     | Germany     | Lamberto et al. (2014)       |
|                 | Caribou associated gemykrovirus 1      | KJ938717    | FaGmCV-13           | Rangifer tarandus         | Caribou           | Faeces                    | Canada      | Ng et al. (2014)             |
|                 | Sewage derived gemykrovirus 1          | KJ547634    | BS3913              | Sewage oxidation pond     | –                 | Sewage                    | New Zealand | Kraberger et al. (2015a)     |
|                 | Rabbit associated gemykroznavirus 1    | KF371631    | as35                | Oryctolagus cuniculus     | Rabbit            | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                 | Gemykroznavirus 1                      | KF371630    | as3                 | Struthio camelus          | Ostrich           | Faeces                    | New Zealand | Sikorski et al. (2013)       |
|                 | Human associated gemykrovirus 1        | KP974693    | DB1                 | Homo sapiens              | Human             | Plasma                    | Germany     | Zhang et al. (2016)          |



**Figure 1.** Distribution of (A) genome-wide, (B) Rep and (C) CP pairwise identities (121 taxa) of genomoviruses calculated using SDT v1.2 (Muhire, Varsani, and Martin 2014).

tree gemykoloviruses are firmly nested within the large clade including the majority of gemycircularviruses (Fig. 4). Given that CP sequences of genomoviruses are considerably more divergent than the Rep sequences (Fig. 1), it appears reasonable to establish a higher (i.e., above the species level) taxonomic framework using the Rep (Fig. 2). The latter protein is also

conserved in other eukaryotic ssDNA viruses (which is not the case for the CP) and can thus be used to assess the place of genomoviruses within the larger community of ssDNA viruses.

To evaluate the taxonomic structure of the *Genomoviridae*, we took advantage of the fact that in Rep-based phylogenetic analyses, genomoviruses consistently form a sister group to



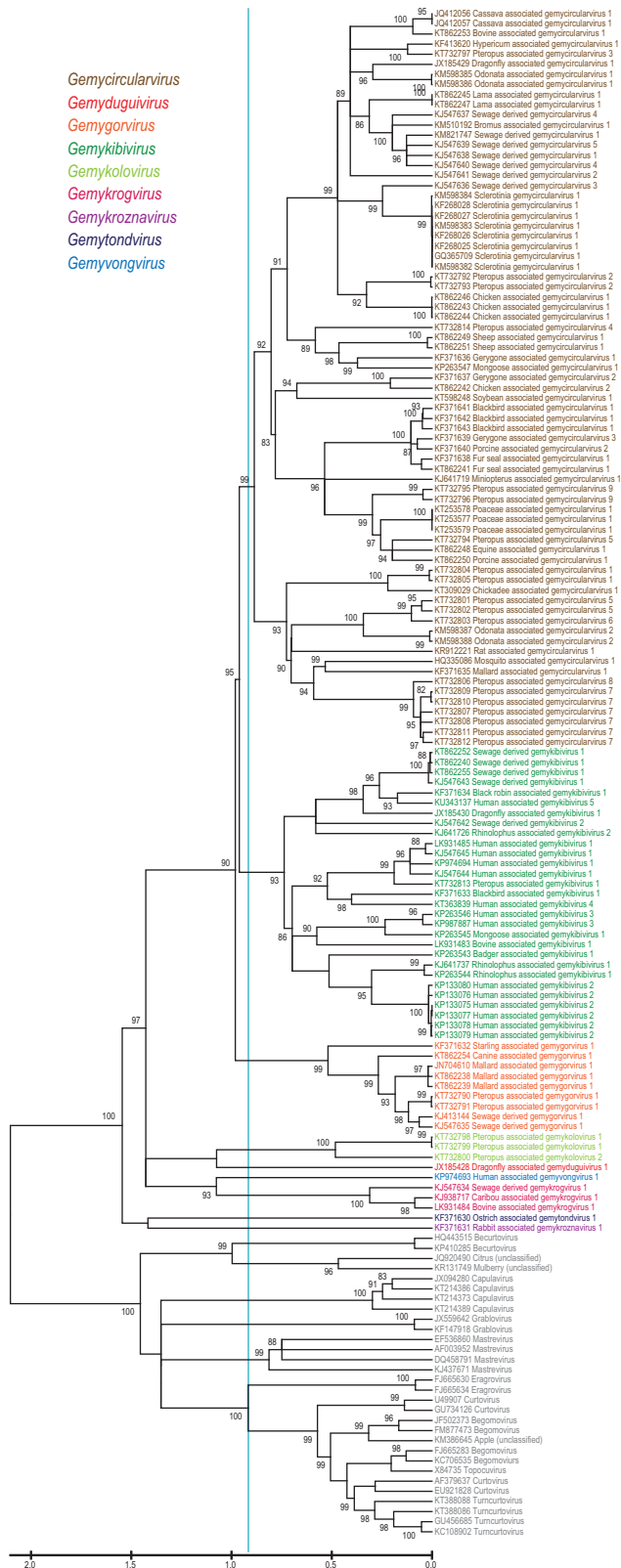


Figure 2. Maximum likelihood phylogenetic tree of the Rep amino acid sequences inferred using PHYML (Guindon et al. 2010) with LG + G+I substitution model and rooted with geminivirus sequences. The sequences of geminiviruses labeled with the corresponding genera names are used as a guide to identify genera within the Genomoviridae family. The cyan line shows a proposed demarcation of genera for both Genomoviridae and Geminiviridae. Branches with <75% SH-like branch support have been collapsed.

members of the Geminiviridae (Krupovic et al. 2016), a comprehensively characterized family of plant viruses with circular ssDNA genomes (Varsani et al. 2014b). Thus, using the established taxonomic framework of the Geminiviridae overlaid on the Rep-based phylogeny as a guide, we could define five clades and four additional singletons within the Genomoviridae branch (Fig. 2). The defined groups displayed equivalent intra-family divergence as the established genera within the family Geminiviridae (Varsani et al. 2014b). The nine groups were supported in both nucleotide and protein sequence inferred phylogenies (Supplementary Fig. S2). Consequently, in addition to the existing genus *Gemyrcircularvirus*, we propose establishing eight new genera within the family Genomoviridae. The details of the nine genera are summarized in Fig. 5 and briefly outlined below.

### 3.1 Gemyrcircularvirus

This genus has the largest number of new species ( $n = 43$ ; seventy-three genomes; Table 1) and includes SsHADV-1, the founding member of the family. Members of the genus display 44% diversity. Viruses within the forty-three species cluster with 99 and 96% branch support values in phylogenetic trees constructed from either Rep or full genome sequences, respectively (Figs 2 and 3).

### 3.2 Gemykibivirus

This is the second most populated genus ( $n = 16$ ; twenty-nine genomes; Table 1) in the family with 43% diversity among its members. The name of the genus is an acronym of words geminivirus-like and myco-like kibi virus (kibi means circular in Amharic). Sequences within the fifteen species cluster with 93% branch support within phylogenetic trees constructed from Rep (Fig. 2) and two well-supported clades (100 and 96%) within trees constructed from full genome sequences (Fig 3), suggesting that recombination has played an important role in the evolution of this group.

### 3.3 Gemygorvirus

Members of this genus ( $n = 5$ ; nine genomes; Table 1) display 49% diversity. The name of the genus is an acronym of words geminivirus-like and myco-like gor virus (gor means round in Hindi). Sequences within the five species cluster with 100 and 99% branch support within phylogenetic trees constructed from either Rep or full genome sequences, respectively (Figs 2 and 3).

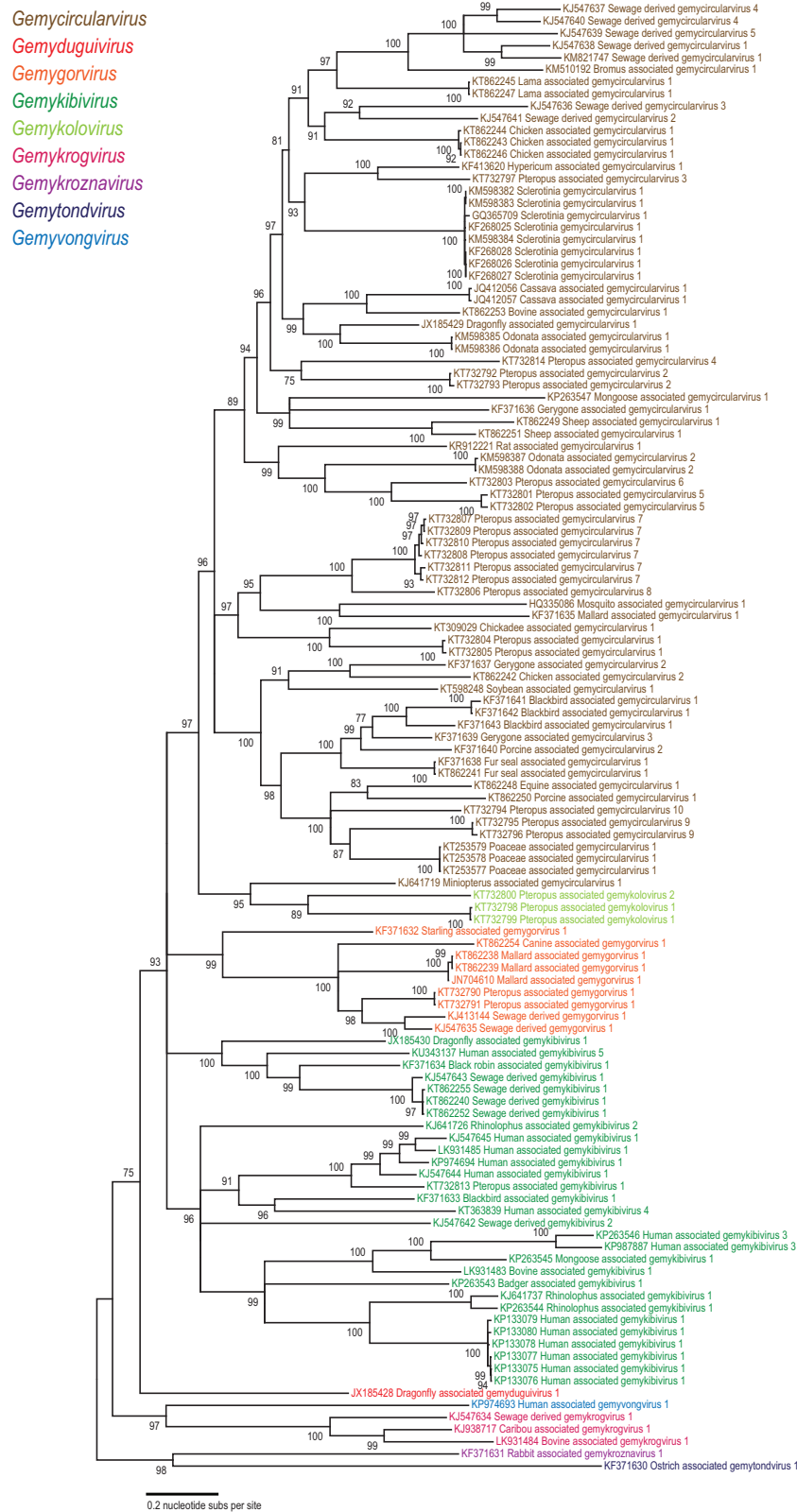
### 3.4 Gemykolovirus

Members of this genus ( $n = 2$ ; three genomes; Table 1) display 37% diversity. The name of the genus is an acronym of words geminivirus-like and myco-like kolo virus (kolo means round in Czech). Sequences within the two species cluster with 100 and 89% branch support within phylogenetic trees constructed from either Rep or full genome sequences, respectively (Figs 2 and 3).

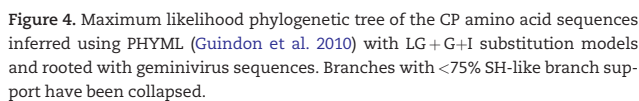
### 3.5 Gemykrogvirus

Members of this genus ( $n = 3$ ; three genomes; Table 1) display 33% diversity. The name of the genus is an acronym of words geminivirus-like and myco-like krog virus (krog means round in Slovenian). Sequences within the three species cluster with 99 and 100% branch support within phylogenetic trees constructed from either Rep or full genome sequences respectively (Figs 2 and 3).





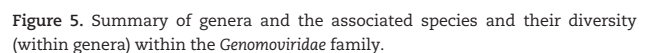
**Figure 3.** Maximum likelihood phylogenetic tree of the genomes of viruses in the *Genomoviridae* family. The tree was inferred using FastTree (Price, Dehal, and Arkin 2010) (GTR + CAT). The numbers at the branches indicate SH-like support values. The topology of tree supports the proposed genera demarcation at the genome level, despite there being evidence of recombination within the genomes. Branches with <75% SH-like branch support have been collapsed.



The name of the genus is an acronym of words geminivirus-like and myco-like vong virus (vong means circular in Lao). The single species *Human associated gemyvongvirus 1* (Table 1) within the genus shares between 56 and 62% genome-wide sequence similarity with viruses in other genera and is a divergent taxon in the phylogenetic trees constructed from either Rep or full genome sequences (Figs 2 and 3).

The name of the genus is an acronym of words geminivirus-like and myco-like *tond virus* (*tond* means round in Maltese). The single species *Ostrich associated gemytondovirus 1* (Table 1) within the genus shares between 53 and 61% genome-wide sequence similarity with viruses in other genera and is a divergent taxon in the phylogenetic trees constructed from either Rep or full genome sequences (Figs 2 and 3).

The name of the genus is an acronym of words geminivirus-like and myco-like krozna virus (krozna means circular in Slovenian). The single species Rabbit associated gemykroznavirus 1 (Table 1)



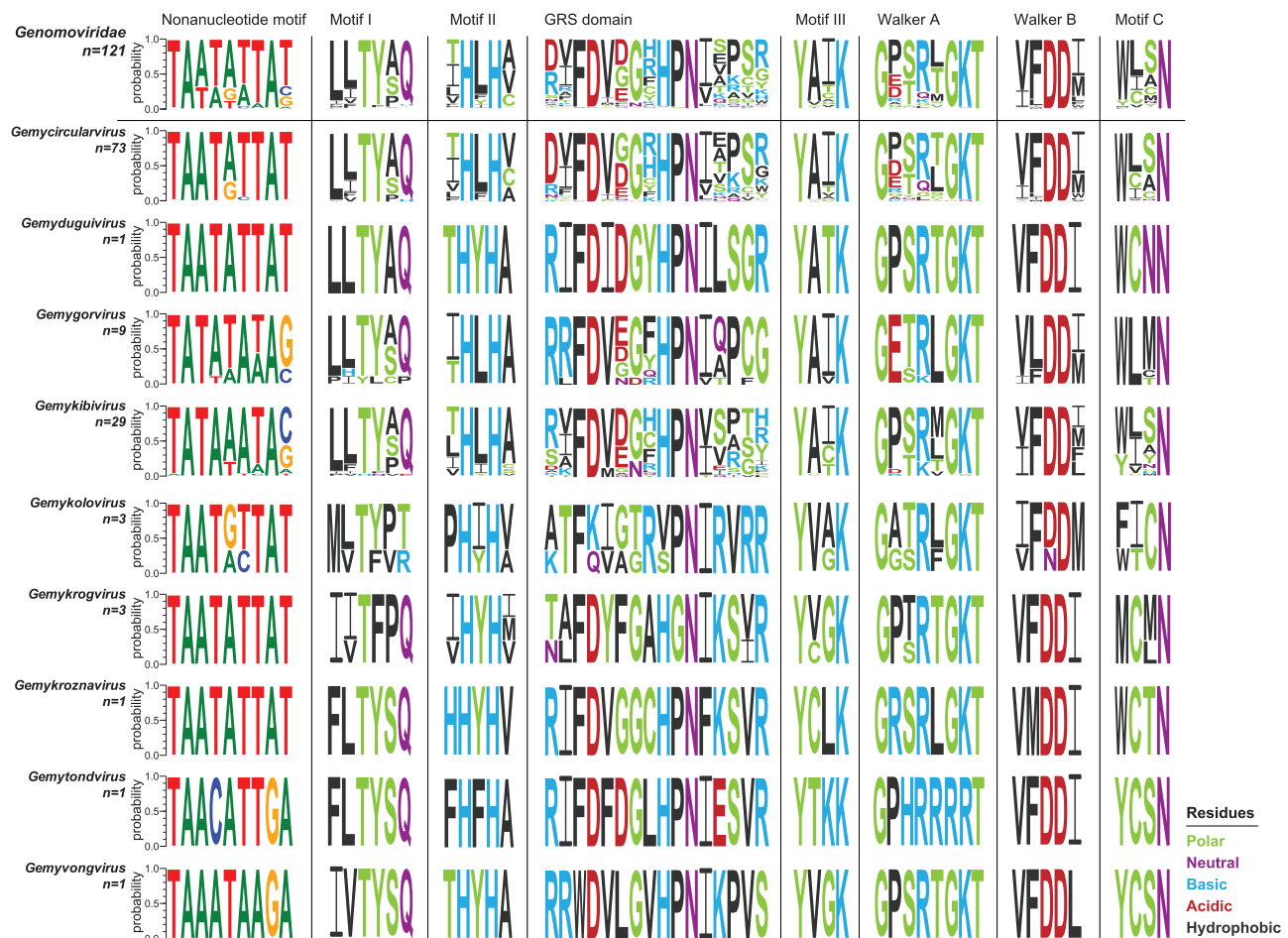


Figure 6. Summary of conserved motifs, that is nonanucleotide and Rep motifs illustrated using WebLogo3 (Crooks et al. 2004) identified in the family Genomoviridae as a whole and its nine genera separately. Note the highly derived Walker A motif (GPHRRRT) in the sole member of the genus Gemytondovirus.

within the genus shares between 56 and 61% genome-wide sequence similarity with other sequences in other genera and is a divergent taxon in the phylogenetic trees constructed from either Rep or full genome sequences (Figs 2 and 3).

### 3.8 Gemyduguvirus

The name of the genus is an acronym of words geminivirus-like and myco-like dugui virus (dugui means circular in Mongolian). The single species *Dragonfly associated gemyduguvirus 1* (Table 1) within the genus shares between 57 and 62% genome-wide sequence similarity with viruses in other genera and is a divergent taxon in the phylogenetic trees constructed from either Rep or full genome sequences (Figs 2 and 3).

## 4. Conserved sequence motifs in the Genomoviridae

CRESS DNA viruses replicate through the rolling circle replication (RCR) mechanism which is similar to that used by bacterial plasmids (Khan 1997; Chandler et al. 2013; Ruiz-Maso et al. 2015). RCR is initiated by the Rep, encoded by CRESS DNA viruses, cleaving the dsDNA between positions 7 and 8 of a nonanucleotide sequence located at a putative stem-loop structure at the origin of replication (Heyraud-Nitschke et al. 1995; Laufs et al. 1995b; Timchenko et al. 1999; Rosario, Duffy, and Breitbart, 2012). In the case of genomoviruses, this

nonanucleotide is variable ('TAWWDWRN') with 'TAATWYAT' being the consensus nonanucleotide for gemycircularviruses, whereas gemykibiruses display the greatest variation in this motif—'WATAWWHAN' (Fig. 6; Supplementary Data S1). In contrast, we note that within the Geminiviridae family, including all recently described geminiviruses (Varsani et al. 2009; Briddon et al. 2010; Krenz et al. 2012; Loconsole et al. 2012; Bernardo et al. 2013; Heydamejad et al. 2013; Ma et al. 2015; Bernardo et al. 2016), the consensus nonanucleotide motif is 'TRAKATTRC'.

The N terminus of the Rep contains motifs that are important for initiating RCR and it is not surprising that some of these motifs are well conserved across many ssDNA viruses, phages, and plasmids that replicate using the RCR mechanism (Ilyina and Koonin, 1992; Vega-Rocha et al. 2007a; Rosario, Duffy, and Breitbart, 2012; Krupovic, 2013). The presence of a single catalytic tyrosine residue in the RCR motif III classifies genomovirus, geminivirus, bacilladnavirus, circovirus and nanovirus Reps as members of superfamily II (Ilyina and Koonin, 1992; Krupovic, 2013).

In genomoviruses, the conserved sequence of the RCR motif I, which is thought to be involved in the recognition of iterative sequences associated with the origin of replication, is predominantly 'uuTYxQ' (u denotes hydrophobic residues and x any residue) (Fig. 6; Supplementary Data S1), with the exception of the Reps of currently known gemykoloviruses and gemykrogviruses. The genomovirus RCR motif II, 'xHxHx'

(Fig. 6; Supplementary Data S1), resembles that found in geminiviruses, and early work has shown that histidines in this motif coordinate divalent metal ions,  $Mg^{2+}$  or  $Mn^{2+}$ , which are important cofactors for endonuclease activity at the origin of replication (Koonin and Ilyina 1992; Laufs et al. 1995b). Genomoviruses have an RCR motif III of 'YxxK' and based on other Rep studies, this motif is involved in the dsDNA cleavage and subsequent covalent attachment of Rep through the catalytic tyrosine residue to the 5' end of the cleaved product (Laufs et al. 1995a, b; Orozco and Hanley-Bowdoin, 1998; Timchenko et al. 1999; Steinfeldt, Finsterbusch, and Mankertz, 2006; Rosario, Duffy, and Breitbart, 2012). The conserved lysine residue in the RCR motif III (Fig. 6; Supplementary Data S1) is proposed to mediate binding and positioning during catalysis (Vega-Rocha et al. 2007a, b). A fourth conserved motif, the geminivirus Rep sequence (GRS), is only found in geminiviruses and genomoviruses (Fig. 6). In geminiviruses, it enables appropriate spatial arrangements of RCR motifs II and III (Nash et al. 2011). Site-directed mutagenesis of the GRS domain in tomato golden mosaic virus yielded non-infectious clones, demonstrating that the GRS is essential for geminivirus replication (Nash et al. 2011) and it is likely this is also the case for genomoviruses.

Rep is a multifunctional protein, with both endonuclease and helicase activities. Rep helicase activity is mediated by conserved motifs known as Walker A, Walker B and motif C located in a C-terminal NTP-binding domain (Fig. 6; Supplementary Data S1) (Gorbalenya, Koonin, and Wolf 1990; Koonin, 1993; Choudhury et al. 2006; Clerot and Bernardi 2006). The helicase domain found in Rep proteins of eukaryotic ssDNA viruses belongs to the helicase superfamily 3 (Gorbalenya, Koonin, and Wolf 1990; Koonin 1993). The conserved Walker A motif of genomoviruses is 'GxxxxGKT', with the exception of gemytondivirus which contains a highly derived variant of this motif (GPHRRRT; Fig. 6). Previous studies have shown that during synthesis of progeny strands, Rep helicase activity unwinds the dsDNA intermediate in the 3'-5' direction using nucleotide triphosphates as an energy source (Choudhury et al. 2006; Clerot and Bernardi 2006). Walker A motif forms part of the 'P-loop' structure in the NTP-binding domain that facilitates ATP recognition and binding with a conserved lysine residue (Desbiez et al. 1995; Timchenko et al. 1999; Choudhury et al. 2006; Clerot and Bernardi 2006; Rosario, Duffy, and Breitbart 2012; George et al. 2014). The Walker B of genomoviruses is predominantly 'uuDDu' (Fig. 6; Supplementary Data S1), whereas the motif C is 'uxxN' (u denotes hydrophobic residues and x any residue; Fig. 6, Supplementary Data S1). The hydrophobic residues in Walker B motif contribute to ATP binding and are essential for ATP hydrolysis, whereas the one in motif C (Fig. 6; Supplementary Data S1) interacts with the gamma phosphate of ATP and the nucleophilic water molecule via a conserved asparagine residue (Choudhury et al. 2006; George et al. 2014).

Genomoviruses from different genera display distinct signatures within the nonanucleotide as well as conserved nuclease and helicase motifs, which are generally consistent with the proposed taxa (Fig. 6; Supplementary Data S1).

## 5. Concluding remarks

The Reps of genomoviruses are most closely related to those of geminiviruses and hence here we used a geminivirus taxonomy-informed approach to classify 121 genomoviruses into Rep sequence-based genera. Within the *Genomoviridae* family we establish eight new genera in addition to the one created previously (Krupovic et al. 2016). Detailed analysis of sequence

motifs conserved within the genomoviral genomes further supports the validity of the proposed genera. We also define a species demarcation criterion of 78% genome-wide identity, that is sequences that share >78% pairwise identity with other genomovirus sequences belong to the same species and those that share <78% can be considered as new species. It is worth noting that despite the fact that geminiviruses have been studied for over two decades, the sequence diversity of all known geminiviruses is similar to that of the recently discovered genomoviruses (46 vs 47%, respectively). This observation strongly suggests that the extent of sequence diversity within this expansive virus group remains largely unexplored.

Although the guidelines presented here are tailored for the classification of viral genomes in the family *Genomoviridae*, a similar sequence-based framework can be easily adapted for other virus clusters identified through metagenomics studies and lacking a pre-existing taxonomic framework, in particular for novel CRESS DNA viruses. We do acknowledge that this approach deviates from a previous norm that used a set of criteria including biological properties such as host range, pathology, vectors, etc. coupled with sequence data. However, given that the rate at which genome sequences of uncultivated viruses are being identified from various sources, we need to establish more robust classification approaches that can easily be implemented on the bases of sequence data alone. Indeed, this necessity is acknowledged by the ICTV which encourages submissions of taxonomic proposals for classification of viruses that are known exclusively from their genome sequences (Simmonds et al. 2017). This new tide in virus taxonomy is expected to catalyze the comprehension of the diversity, ecology and evolution of the global virome.

## Supplementary data

Supplementary data are available at *Virus Evolution* online.

## Disclaimer

This article is based on the taxonomic proposal 2016.001a-agF.U.v5.*Genomoviridae* which has been considered and approved by the Executive Committee (EC) of the ICTV. AV and MK are elected members of the ICTV EC.

Conflict of interest: None declared.

## References

- Adams, M. J., et al. (2016) 'Ratification Vote on Taxonomic Proposals to the International Committee on Taxonomy of Viruses', *Archives of Virology*, 161: 2921–49.
- Bernardo, P., et al. (2013) 'Identification and Characterisation of a Highly Divergent Geminivirus: Evolutionary and Taxonomic Implications', *Virus Research*, 177: 35–45.
- , et al. (2016) 'Molecular Characterization and Prevalence of Two Capulaviruses: Alfalfa Leaf Curl Virus From France and Euphorbia Caput-Medusae Latent Virus From South Africa', *Virology*, 493: 142–53.
- Briddon, R. W., et al. (2010) 'Turnip Curly Top Virus, a Highly Divergent Geminivirus Infecting Turnip in Iran', *Virus Research*, 152: 169–75.
- Brown, J. K., et al. (2015) 'Revision of Begomovirus Taxonomy Based on Pairwise Sequence Comparisons', *Archives of Virology*, 160: 1593–619.



- Cadar, D., et al. (2013) 'Phylogeny, Spatio-Temporal Phylodynamics and Evolutionary Scenario of Torque teno sus virus 1 (TTSuV1) and 2 (TTSuV2) in Wild Boars: Fast Dispersal and High Genetic Diversity', *Veterinary Microbiology*, 166: 200–13.
- Chandler, M., et al. (2013) 'Breaking and Joining Single-Stranded DNA: the HUH Endonuclease Superfamily', *Nature Reviews Microbiology*, 11: 525–38.
- Choudhury, N. R., et al. (2006) 'The Oligomeric Rep Protein of Mungbean Yellow Mosaic India Virus (MYMIV) Is a Likely Replicative Helicase', *Nucleic Acids Research*, 34: 6362–77.
- Clerot, D., and Bernardi, F. (2006) 'DNA Helicase Activity Is Associated with the Replication Initiator Protein Rep of Tomato Yellow Leaf Curl Geminivirus', *Journal of Virology*, 80: 11322–30.
- Conceicao-Neto, N., et al. (2015) 'Fecal Virome Analysis of Three Carnivores Reveals a Novel Nodavirus and Multiple Gemycircularviruses', *Virology Journal*, 12: 79.
- Crooks, G. E., et al. (2004) 'WebLogo: a Sequence Logo Generator', *Genome Research*, 14: 1188–90.
- Dayaram, A., et al. (2012) 'Molecular Characterisation of a Novel Cassava Associated Circular ssDNA Virus', *Virus Research*, 166: 130–5.
- , et al. (2015) 'Identification of Diverse Circular Single-Stranded DNA Viruses in Adult Dragonflies and Damselflies (Insecta: Odonata) of Arizona and Oklahoma, USA', *Infection, Genetics, and Evolution*, 30: 278–87.
- , et al. (2016) 'Diverse Circular Replication-Associated Protein Encoding Viruses Circulating in Invertebrates Within a Lake Ecosystem', *Infection, Genetics, and Evolution*, 39: 304–16.
- Desbiez, C., et al. (1995) 'Rep Protein of Tomato Yellow Leaf Curl Geminivirus Has an ATPase Activity Required for Viral DNA Replication', *Proceedings of the National Academy of Sciences of the United States of America*, 92: 5640–4.
- Du, Z., et al. (2014) 'Identification and Molecular Characterization of a Single-Stranded Circular DNA Virus with Similarities to *Sclerotinia sclerotiorum* Hypovirulence-Associated DNA Virus 1', *Archives of Virology*, 159: 1527–31.
- Duffy, S., and Holmes, E. C. (2008) 'Phylogenetic Evidence for Rapid Rates of Molecular Evolution in the Single-Stranded DNA Begomovirus Tomato Yellow Leaf Curl Virus', *Journal of Virology*, 82: 957–65.
- , and — (2009) 'Validation of High Rates of Nucleotide Substitution in Geminiviruses: Phylogenetic Evidence From East African Cassava Mosaic Viruses', *Journal of General Virology*, 90: 1539–47.
- Dutilh, B. E., et al. (2014) 'A Highly Abundant Bacteriophage Discovered in the Unknown Sequences of Human Faecal Metagenomes', *Nature Communications*, 5: 4498.
- Firth, C., et al. (2009) 'Insights into the Evolutionary History of an Emerging Livestock Pathogen: Porcine Circovirus 2', *Journal of Virology*, 83: 12813–21.
- George, B., et al. (2014) 'Mutational Analysis of the Helicase Domain of a Replication Initiator Protein Reveals Critical Roles of Lys 272 of the B' Motif and Lys 289 of the Beta-Hairpin Loop in Geminivirus Replication', *Journal of General Virology*, 95: 1591–602.
- Gorbalenya, A. E., Koonin, E. V., and Wolf, Y. I. (1990) 'A New Superfamily of Putative NTP-Binding Domains Encoded by Genomes of Small DNA and RNA Viruses', *FEBS Letters*, 262: 145–8.
- Grigoras, I., et al. (2010) 'High Variability and Rapid Evolution of a Nanovirus', *Journal of Virology*, 84: 9105–17.
- Guindon, S., et al. (2010) 'New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0', *Systems Biology*, 59: 307–21.
- Halary, S., et al. (2016) 'Novel Single-Stranded DNA Circular Viruses in Pericardial Fluid of Patient with Recurrent Pericarditis', *Emerging infectious diseases*, 22: 1839–41.
- Hanna, Z. R., et al. (2015) 'Isolation of a Complete Circular Virus Genome Sequence from an Alaskan Black-Capped Chickadee (*Parus atricapillus*) Gastrointestinal Tract Sample', *Genome Announcements*, 3: e01081\_15.
- Harkins, G. W., et al. (2009) 'Experimental Evidence Indicating that Mastreviruses Probably Did Not Co-Diverge with Their Hosts', *Virology Journal*, 6: 104.
- , et al. (2014) 'Towards Inferring the Global Movement of Beak and Feather Disease Virus', *Virology*, 450–451: 24–33.
- Heydarnejad, J., et al. (2013) 'Fulfilling Koch's Postulates for Beet Curly Top Iran Virus and Proposal for Consideration of New Genus in the Family Geminiviridae', *Archives of Virology*, 158: 435–43.
- Heyraud-Nitschke, F., et al. (1995) 'Determination of the Origin Cleavage and Joining Domain of Geminivirus Rep Proteins', *Nucleic Acids Research*, 23: 910–6.
- Ilyina, T. V., and Koonin, E. V. (1992) 'Conserved Sequence Motifs in the Initiator Proteins for Rolling Circle DNA Replication Encoded by Diverse Replicons from Eubacteria, Eucaryotes and Archaeobacteria', *Nucleic Acids Research*, 20: 3279–85.
- Khan, S. A. (1997) 'Rolling-Circle Replication of Bacterial Plasmids', *Microbiology and Molecular Biology Reviews*, 61: 442–55.
- Kolawole, A. O., et al. (2014) 'Flexibility in Surface-Exposed Loops in a Virus Capsid Mediates Escape From Antibody Neutralization', *Journal of Virology*, 88: 4543–57.
- Koonin, E. V. (1993) 'A Common Set of Conserved Motifs in a Vast Variety of Putative Nucleic Acid-Dependent ATPases Including MCM Proteins Involved in the Initiation of Eukaryotic DNA Replication', *Nucleic Acids Research*, 21: 2541–7.
- , and Ilyina, T. V. (1992) 'Geminivirus Replication Proteins Are Related to Prokaryotic Plasmid Rolling Circle DNA Replication Initiator Proteins', *Journal of General Virology*, 73: 2763–6.
- Kraberger, S., et al. (2013) 'Discovery of *Sclerotinia sclerotiorum* Hypovirulence-Associated Virus-1 in Urban River Sediments of Heathcote and Styx Rivers in Christchurch City, New Zealand', *Genome Announcements*, 1: e00559\_13.
- , et al. (2015a) '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.
- , et al. (2015b) 'Identification of Novel Bromus- and Trifolium-Associated Circular DNA Viruses', *Archives of Virology*, 160: 1303–11.
- Krenz, B., et al. (2012) 'Complete Genome Sequence of a New Circular DNA Virus From Grapevine', *Journal of Virology*, 86: 7715.
- Krupovic, M. (2013) 'Networks of Evolutionary Interactions Underlying the Polyphyletic Origin of ssDNA Viruses', *Current Opinion in Virology*, 3: 578–86.
- , et al. (2016) 'Genomoviridae: a New Family of Widespread Single-Stranded DNA Viruses', *Archives of Virology*, 161: 2633–43.
- Labonte, J. M., and Suttle, C. A. (2013) 'Previously Unknown and Highly Divergent ssDNA Viruses Populate the Oceans', *ISME Journal*, 7: 2169–77.

- Lamberto, I., et al. (2014) 'Mycovirus-Like DNA Virus Sequences From Cattle Serum and Human Brain and Serum Samples From Multiple Sclerosis Patients', *Genome Announcements*, 2: e00848\_14.
- Laufs, J., et al. (1995a) 'Identification of the Nicking Tyrosine of Geminivirus Rep Protein', *FEBS Letters*, 377: 258–62.
- , et al. (1995b) 'In Vitro Cleavage and Joining at the Viral Origin of Replication by the Replication Initiator Protein of Tomato Yellow Leaf Curl Virus', *Proceedings of the National Academy of Sciences of the United States of America*, 92: 3879–83.
- Li, W., et al. (2015) 'A Novel Gemycircularvirus From Experimental Rats', *Virus Genes*, 51: 302–5.
- Liu, S., et al. (2016) 'Fungal DNA Virus Infects a Mycophagous Insect and Utilizes It as a Transmission Vector', *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1608013113.
- Loconsole, G., et al. (2012) 'Identification of a Single-Stranded DNA Virus Associated with Citrus Chlorotic Dwarf Disease, a New Member in the Family Geminiviridae', *Virology*, 432: 162–72.
- Ma, Y., et al. (2015) 'Identification and Molecular Characterization of a Novel Monopartite Geminivirus Associated with Mulberry Mosaic Dwarf Disease', *Journal of General Virology*, 96: 2421–34.
- Male, M. F., et al. (2015) 'Genome Sequences of Poaceae-Associated Gemycircularviruses from the Pacific Ocean Island of Tonga', *Genome Announcements*, 3: e01144\_15.
- , et al. (2016) 'Cycloviruses, Gemycircularviruses and Other Novel Replication-Associated Protein Encoding Circular Viruses in Pacific flying fox (*Pteropus tonganus*) Faeces', *Infection, Genetics, and Evolution*, 39: 279–92.
- Martin, D. P., et al. (2011) 'Recombination in Eukaryotic Single Stranded DNA Viruses', *Viruses*, 3: 1699–738.
- Marzano, S. Y., and Domier, L. L. (2016) 'Novel Mycoviruses Discovered from Metatranscriptomics Survey of Soybean Phyllosphere Phytobiomes', *Virus Research*, 213: 332–42.
- Muhire, B., et al. (2013) 'A Genome-Wide Pairwise-Identity-Based Proposal for the Classification of Viruses in the Genus Mastrevirus (family Geminiviridae)', *Archives of Virology*, 158: 1411–24.
- Muhire, B. M., Varsani, A., and Martin, D. P. (2014) 'SDT: a Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation', *PLoS One*, 9: e108277.
- Nash, T. E., et al. (2011) 'Functional Analysis of a Novel Motif Conserved Across Geminivirus Rep Proteins', *Journal of Virology*, 85: 1182–92.
- Ng, T. F., et al. (2011) 'Broad Surveys of DNA Viral Diversity Obtained Through Viral Metagenomics of Mosquitoes', *PLoS One*, 6: e20579.
- , et al. (2014) 'Preservation of Viral Genomes in 700-y-old Caribou Feces from a Subarctic Ice Patch', *Proceedings of the National Academy of Sciences of the United States of America*, 111: 16842–7.
- Nguyen, V. G., et al. (2012) 'Population Dynamics and ORF3 Gene Evolution of Porcine Circovirus Type 2 Circulating in Korea', *Archives of Virology*, 157: 799–810.
- Orozco, B. M., and Hanley-Bowdoin, L. (1998) 'Conserved Sequence and Structural Motifs Contribute to the DNA Binding and Cleavage Activities of a Geminivirus Replication Protein', *Journal of Biological Chemistry*, 273: 24448–56.
- Phan, T. G., et al. (2015) 'Small Circular Single Stranded DNA Viral Genomes in Unexplained Cases of Human Encephalitis, Diarrhea, and in Untreated Sewage', *Virology*, 482: 98–104.
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) 'FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments', *PLoS One*, 5: e9490.
- Rosario, K., et al. (2012) 'Diverse Circular ssDNA Viruses Discovered in Dragonflies (Odonata: Epiprocta)', *Journal of General Virology*, 93: 2668–81.
- , Duffy, S., and Breitbart, M. (2012) 'A Field Guide to Eukaryotic Circular Single-Stranded DNA Viruses: Insights Gained From Metagenomics', *Archives of Virology*, 157: 1851–71.
- Roux, S., et al. (2012) 'Evolution and Diversity of the Microviridae Viral Family Through a Collection of 81 New Complete Genomes Assembled from Virome Reads', *PLoS One*, 7: e40418.
- , et al. (2013) 'Chimeric Viruses Blur the Borders Between the Major Groups of Eukaryotic Single-Stranded DNA Viruses', *Nature Communications*, 4: 2700.
- Ruiz-Maso, J. A., et al. (2015) 'Plasmid Rolling-Circle Replication', *Microbiology Spectrum*, 3: PLAS-0035-2014.
- Shangjin, C., Cortey, M., and Segales, J. (2009) 'Phylogeny and Evolution of the NS1 and VP1/VP2 Gene Sequences from Porcine Parvovirus', *Virus Research*, 140: 209–15.
- Sikorski, A., et al. (2013) 'Novel Myco-Like DNA Viruses Discovered in the Faecal Matter of Various Animals', *Virus Research*, 177: 209–16.
- Simmonds, P., et al. (2017) 'Virus Taxonomy in the Age of Metagenomics', *Nature Reviews Microbiology* (in press). DOI: 10.1038/nrmicro.2016.177.
- Steel, O., et al. (2016) 'Circular Replication-Associated Protein Encoding DNA Viruses Identified in the Faecal Matter of Various Animals in New Zealand', *Infection, Genetics, and Evolution*, 43: 151–64.
- Steinfeldt, T., Finsterbusch, T., and Mankertz, A. (2006) 'Demonstration of Nicking/Joining Activity at the Origin of DNA Replication Associated with the Rep and Rep' Proteins of Porcine Circovirus Type 1', *Journal of Virology*, 80: 6225–34.
- Streck, A. F., et al. (2011) 'High Rate of Viral Evolution in the Capsid Protein of Porcine Parvovirus', *Journal of General Virology*, 92: 2628–36.
- Timchenko, T., et al. (1999) 'A Single Rep Protein Initiates Replication of Multiple Genome Components of Faba Bean Necrotic Yellow Virus, a Single-Stranded DNA Virus of Plants', *Journal of Virology*, 73: 10173–82.
- Uch, R., et al. (2015) 'Divergent Gemycircularvirus in HIV-Positive Blood, France', *Emerging Infectious Diseases*, 21: 2096–8.
- van den Brand, J. M., et al. (2012) 'Metagenomic Analysis of the Viral Flora of Pine Marten and European Badger Feces', *Journal of Virology*, 86: 2360–5.
- Varsani, A., et al. (2009) 'A Highly Divergent South African Geminivirus Species Illuminates the Ancient Evolutionary History of This Family', *Virology Journal*, 6: 36.
- , et al. (2014a) 'Revisiting the Classification of Curtoviruses Based on Genome-Wide Pairwise Identity', *Archives of Virology*, 159: 1873–82.
- , et al. (2014b) 'Establishment of Three New Genera in the Family Geminiviridae: Becurtovirus, Eragrovirus and Turncortovirus', *Archives of Virology*, 159: 2193–203.
- Vega-Rocha, S., et al. (2007a) 'Solution Structure, Divalent Metal and DNA Binding of the Endonuclease Domain from the Replication Initiation Protein from Porcine Circovirus 2', *Journal of Molecular Biology*, 367: 473–87.
- , et al. (2007b) 'Solution Structure of the Endonuclease Domain from the Master Replication Initiator Protein of the Nanovirus Faba Bean Necrotic Yellow Virus and Comparison with the Corresponding Geminivirus and Circovirus Structures', *Biochemistry*, 46: 6201–12.



- Wu, Z., et al. (2016) 'Deciphering the Bat Virome Catalog to Better Understand the Ecological Diversity of Bat Viruses and the Bat Origin of Emerging Infectious Diseases', *ISME Journal*, 10: 609–20.
- Yau, S., et al. (2011) 'Virophage Control of Antarctic Algal Host-Virus Dynamics', *Proceedings of the National Academy of Sciences of the United States of America*, 108: 6163–8.
- Yu, X., et al. (2010) 'A Geminivirus-Related DNA Mycovirus that Confers Hypovirulence to a Plant Pathogenic Fungus', *Proceedings of the National Academy of Sciences of the United States of America*, 107: 8387–92.
- , et al. (2013) 'Extracellular Transmission of a DNA Mycovirus and Its Use as a Natural Fungicide', *Proceedings of the National Academy of Sciences of the United States of America*, 110: 1452–7.
- Yutin, N., et al. (2015) 'A Novel Group of Diverse Polinton-Like Viruses Discovered by Metagenome Analysis', *BMC Biology*, 13: 95.
- Zhang, W., et al. (2016) 'Viral Nucleic Acids in Human Plasma Pools', *Transfusion*, 56: 2248–55.
- Zhou, C., et al. (2015) 'A Novel Gemycircularvirus in an Unexplained Case of Child Encephalitis', *Virology Journal*, 12: 197.