

# *Bacilladnaviridae*: refined taxonomy and new insights into the biology and evolution of diatom-infecting DNA viruses

Arvind Varsani<sup>1,2,\*</sup>, Joy M. Custer<sup>1</sup>, Ilaria N. Cobb<sup>1</sup>, Ciara Harding<sup>1</sup>, Courtney L. Collins<sup>1</sup>, Crystal Suazo<sup>1</sup>, Joshua Schreck<sup>1</sup>, Rafaela S. Fontenele<sup>1</sup>, Daisy Stainton<sup>3</sup>, Anisha Dayaram<sup>4</sup>, Sharyn Goldstein<sup>3</sup>, Darius Kazlauskas<sup>5</sup>, Simona Kraberger<sup>1</sup> and Mart Krupovic<sup>6,\*</sup>

## Abstract

Bacilladnaviruses are single-stranded DNA viruses that infect diatoms that, so far, have been primarily identified in marine organisms and environments. Using a viral metagenomics approach, we discovered 13 novel bacilladnaviruses originating from samples of mud-flat snail (*Amphibola crenata*;  $n=3$  genomes) and benthic sediments ( $n=10$  genomes) collected from the Avon-Heathcote Estuary in New Zealand. Comparative genomics and phylogenetic analysis of the new bacilladnavirus sequences in the context of the previously classified members of the family helped refine and further expand the *Bacilladnaviridae* taxonomy. Here, based on the replication-associated protein phylogeny and pairwise identities, we established 4 new genera – *Aberdnavirus*, *Keisodnavirus*, *Puahadnavirus* and *Seawadnavirus* – and 13 new species within the family. Comparison of the bacilladnavirus capsid protein sequences suggests that the positively charged N-terminal region (R-arm) is required for encapsidation of the larger genomes, whereas the smaller bacilladnavirus genomes can be packaged in the absence of the R-arm subdomain. Furthermore, analysis of the bacilladnavirus genomes revealed that members of three genera encode a highly derived variant of a phospholipase A1, which is predicted to be involved in the lysis of the infected diatoms and/or facilitates the entry of the virions into the host cells. Collectively, our results allow refining of the taxonomy of bacilladnaviruses and provide new insights into the biology and evolution of this understudied group of diatom viruses.

## INTRODUCTION

Bacilladnaviruses are single-stranded DNA viruses in the phylum *Cressdnaviricota* (class *Arfviricetes*, order *Baphyvirales*) [1], with genomes ranging from ~3.5 to 6 kb. Bacilladnavirus genomes encode at least four major ORFs, but only two of them can currently be assigned functions, namely the replication-associated protein (Rep) and capsid protein (CP). As all other members of the *Cressdnaviricota*, bacilladnaviruses appear to replicate by a rolling circle mechanism and encode a characteristic Rep protein with two domains, the N-terminal HUH superfamily endonuclease domain and the C-terminal superfamily 3 helicase domain [2]. The two domains display congruent phylogenies within the family, suggesting coevolution [3]. The CP of bacilladnaviruses is structurally most closely related to the CPs of positive-sense RNA viruses of the *Nodaviridae* family [2, 4]. Notably, bacilladnaviruses are the only viruses in the *Cressdnaviricota* that have been shown to form T=3 icosahedral capsids [4], with all other members in this phylum having smaller T=1 capsids. Bacilladnaviruses cause lysis of their hosts, but the mechanism of cell lysis is unknown.

Received 23 November 2024; Accepted 12 February 2025; Published 12 March 2025

**Author affiliations:** <sup>1</sup>The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA; <sup>2</sup>Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, 7925, Cape Town, South Africa; <sup>3</sup>School of Biological Sciences, University of Canterbury, Christchurch, New Zealand; <sup>4</sup>Institute of Neurophysiology, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany; <sup>5</sup>Institute of Biotechnology, Vilnius University, Saulėtekio av. 7, Vilnius 10257, Lithuania; <sup>6</sup>Institut Pasteur, Université Paris Cité, CNRS UMR6047, Archaeal Virology Unit, Paris, France.

\*Correspondence: Arvind Varsani, arvind.varsani@asu.edu; Mart Krupovic, mart.krupovic@pasteur.fr

**Keywords:** algal viruses; *Bacilladnaviridae*; *Cressdnaviricota*; diatoms; single jelly-roll capsid proteins; ssDNA viruses.

**Abbreviations:** CP, capsid protein; ICTV, International Committee on Taxonomy of Viruses; PDB, Protein Data Bank; PLA1, phospholipase A1; RCA, rolling circle amplification; Rep, replication-associated protein.

All the sequences have been deposited in GenBank under accession numbers OM154938-OM154950, and the short reads mapping to the genomes have been deposited under BioProject no. PRJNA792815, BioSamples no. SAMN24466648-SAMN24466649 and SRA no. SRR17399949-SRR17399950.

Three supplementary figures, two supplementary tables and a supplementary excel file are available with the online version of this article.

002084 © 2025 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License. The Microbiology Society waived the open access fees for this article.

All bacilladnaviruses for which hosts have been experimentally determined infect diatoms, primarily *Haslea ostrearia* and members of the *Chaetoceros* genus [5–12]. Bacilladnaviruses have also been identified in mud-flat snail/tītiko (*Amphibola crenata*) [2] and red snapper (*Lutjanus* sp.) [13] tissues as well as in various marine/estuarine animal/environmental samples, such as ocean water [14] and estuarine benthic sediments [2], where they are also likely to infect diatoms.

The family *Bacilladnaviridae* was created in 2018 [15] and included three genera, namely *Diatodnavirus* (one species), *Kieseladnavirus* (one species) and *Protobacilladnavirus* (seven species) [16]. The genera were established based on the phylogeny of the HUH superfamily Rep, the most conserved protein encoded by bacilladnaviruses, with a 75% amino acid identity threshold chosen as a species demarcation criterion. Notably, the Rep sequences of bacilladnaviruses are highly divergent compared to those of other cressdnaviricots, resulting in the unstable position of the bacilladnavirus clade in global phylogenies [17]. Further exploration of the bacilladnavirus diversity is expected to help mitigate this situation.

Here, we describe the identification of 13 bacilladnavirus genomes from mud-flat snail ( $n=3$ ) and benthic sediments ( $n=10$ ), both collected from Avon-Heathcote Estuary (Ihutai) in New Zealand (Aotearoa). We further analyse these genomes with 17 other bacilladnavirus genomes available in GenBank to refine and update the taxonomy of the family *Bacilladnaviridae*.

## METHODS

### Sample collection and identification of complete viral genomes

As part of a study to identify viruses in estuarine ecosystems, we collected ~20 individuals of mud-flat snail/tītiko and ~20 g of benthic sediments from the Avon-Heathcote estuary (Ihutai; location 43.5578 s, 172.7058 E) in July 2012. Samples were refrigerated upon collection. The estuary forms the confluence of two rivers (the Avon and Heathcote), which flow through the city of Christchurch, New Zealand (Aotearoa). The mud-flat snail samples were removed from the shells and washed in sterile distilled water, and the whole bodies were then pooled. The pooled mud-flat snail samples and the benthic sediments were homogenized separately in SM buffer [0.1 M NaCl, 50 mM Tris/HCl (pH 7.4), 10 mM MgSO<sub>4</sub>] at a ratio of 10 ml SM buffer to 5 g of tissue/benthic sediments. Cellular debris was pelleted by centrifuging (10 000 g for 10 min), and the supernatant was sequentially filtered through 0.45 μm and 0.2 μm syringe filters (Sartorius Stedim Biotech, Germany). The viral DNA was extracted from 200 μl of filtrate using the High Pure Viral Nucleic Acid Kit (Roche, USA). To enrich circular DNA, we used rolling circle amplification (RCA) using TempliPhi 2000 (GE Healthcare, USA).

The RCA products of the two samples were used to generate 170 bp insert libraries, which were sequenced on an Illumina HiSeq 2000 (Illumina, USA) platform at the Beijing Genomics Institute (Hong Kong). The raw reads (for a 2×91 nt) were trimmed with Trimmomatic v0.39 [18] and *de novo* assembled using metaSPAdes 3.14.1 [19]. The *de novo* assembled contigs were analysed against a viral RefSeq protein database of bacilladnavirus protein sequences using BLASTX [20]. All contigs with hits to bacilladnavirus protein sequences were checked for terminal redundancy to determine if they were circular. For the circular bacilladnavirus-like contigs ( $n=13$ ), we used ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>) to identify and annotate ORFs.

### Viral genome sequence analysis

The full genome sequences of bacilladnaviruses available in GenBank ( $n=17$ ) were downloaded and analysed together with the 13 newly identified ones from mud-flat snail/tītiko ( $n=3$ ) and benthic sediments ( $n=10$ ). The Rep and CP sequences were extracted, translated and assembled into datasets for further analysis. Rep and CP sequence datasets were aligned using MAFFT [21], and these alignments were used to infer maximum likelihood phylogenetic trees using PhyML 3.0 [22] with best-fit model rtREV+G+I determined using ProtTest 3 [23]. Representative sequences from the *Circoviridae* family were used to root the Rep phylogenetic tree. The CP phylogenetic tree was midpoint rooted. PLA1 sequences were aligned using PROMALS3D [24], and the alignments visualized using Jalview v2 [25]. All sequence pairwise identities were determined using SDT v1.2 [26].

### Identification of the bacilladnavirus PLA1 homologues

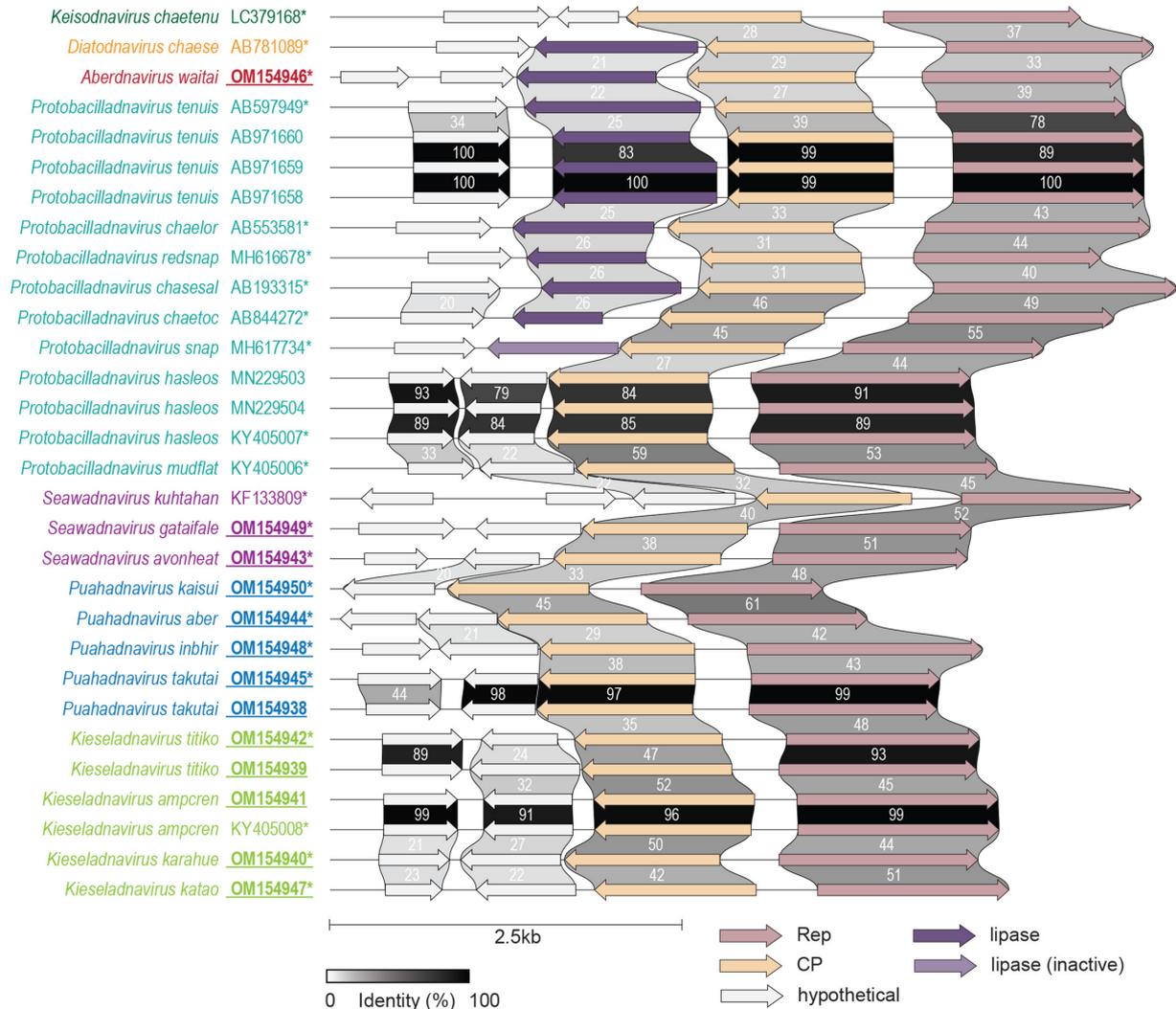
Sequences of hypothetical bacilladnavirus proteins (i.e. not CP or Rep) were analysed using HHsearch against the PFAM (Database of Protein Families), Protein Data Bank (PDB), Conserved Domain Database, and uniprot\_sprot\_vir70 databases [27]. The structures of representative bacilladnaviral PLA1 homologues were modelled using ColabFold v1.5.5 [28] ‘alphafold2\_multimer\_v3’ model with six recycles. Structures were searched against the PDB database using DALI [29]. Protein structures and structural models were visualized using UCSF ChimeraX v1.7.1 [30].

## RESULTS AND DISCUSSION

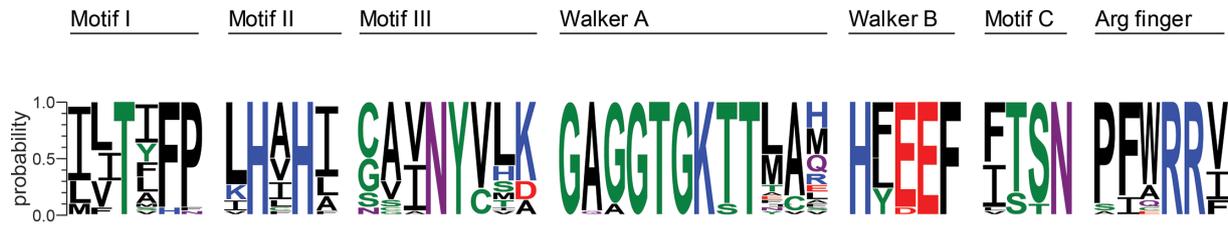
### Identification of new bacilladnavirus genomes

We identified 13 bacilladnavirus genomes that range in size from 3492 to 5608 nt (Fig. 1). Three of these were from mud-flat snail and ten from benthic sediments sampled from the Avon-Heathcote Estuary in New Zealand. It is most likely that all these bacilladnaviruses infect diatoms that have been indirectly sampled. Consistently, the mud-flat snail is a detritus or deposit feeder, which is known to feed on diatoms and other microbes [31]. Notably, the genome of avonheates virus Gas\_1207 (OM154941) from mud-flat snail shares 99.6% genome-wide pairwise identity with Avon-Heathcote estuary associated bacilladnavirus AHEaBavV1 (KY405008), previously identified in benthic sediments at the same location [2]. The two viruses belong to the same species, further supporting that the mud-flat snail-associated viruses infect diatom hosts.

The Rep sequences of bacilladnaviruses contain the three characteristic RCR motifs I, II and III [32–34] and the superfamily 3 helicase motifs Walker A, Walker B, motif C [35, 36], and the Arg finger [17] (Fig. 2). The conserved motifs display signatures that distinguish bacilladnaviruses from other members of the phylum *Cressdnaviricota* [1]. A distinct feature of bacilladnaviruses is the FP and PF residues in the motif I and Arg finger, respectively, as well as the two glutamates in the Walker B motif, instead of the aspartates present in the Reps of most other cressdnaviricots.



**Fig. 1.** Genome organization of all the bacilladnaviruses (complete genomes) available in GenBank. Genome comparisons were generated using Clinker [67]. The species names and accession numbers are provided on the left. The genera that the viruses belong to are colour coded. The accession numbers of the new sequences identified from mud-flat snail and benthic sediments sampled from the Avon-Heathcote Estuary (Ihuta) in New Zealand are in bold font and underlined. An asterisk (\*) next to the accession number denotes representative sequences for each species.



**Fig. 2.** Sequence probability logos, generated using WebLogo 3 [68], of the motifs conserved in bacilladnavirus Reps. The rolling circle replication endonuclease domain includes motifs I, II and III, whereas the superfamily 3 helicase domain includes Walker A and B, motif C and the Arg finger.

### Classification and nomenclature of new bacilladnaviruses

Apart from avonheates virus SG2\_28 (OM154938) and avonheates virus SG\_28 (OM154945), which share 91.8% genome-wide pairwise identity, and Gas\_1207 (OM154941) which belongs to the same species as AHEaBavV1 (KY405008), the other new bacilladnaviruses share 56.2–64.9% genome-wide pairwise identities to the previously classified bacilladnaviruses (Supplementary Data S1, available in the online Supplementary Material), suggesting that new taxa are needed for their classification. To extend the bacilladnavirus taxonomy, we searched the GenBank for other unclassified putative members of this family and identified three such bacilladnavirus genomes, namely Bacilladnaviridae sp. ctia23 (MH617734) and Bacilladnaviridae sp. ctdc18 (MH616678) from red snapper tissue [13] and Chaetoceros tenuissimus DNA virus SS12-43V (LC379168). Maximum likelihood phylogenetic analysis of the Rep sequences revealed four reasonably supported clades (>60% bootstrap support) and three singletons (Fig. 3a). Thus, we created four new genera to classify the new viruses and moved one species (accession no. KF133809) from *Protobacilladnavirus* to a new genus, *Seawadnavirus*.

The new genus names have been derived as follows:

- (1) *Aberdnavirus*: Derived from aber which is an estuary in Welsh
- (2) *Keisodnavirus*: Derived from Keisō (珪藻) which is a diatom in Japanese
- (3) *Puahadnavirus*: Derived from pūahatanga which is an estuary in Māori
- (4) *Seawadnavirus*: Derived from seawater

Genome as well as Rep and CP amino acid sequence comparisons showed that most members display 56–66% (median 59.5%), 35–63% (median 47.7%) and 23–66% (median 31.6) pairwise identities, respectively, with only a small fraction showing > 78% (Fig. 4a). A 75% Rep amino acid pairwise identity as a species demarcation threshold was proposed for members of the genus *Protobacilladnavirus* [15] (because other genera were represented by singletons at the time). Based on the distribution of the pairwise identities of all 30 bacilladnavirus Rep amino acid sequences, we extend and apply the same demarcation criterion to all genera with more than one species. Notably, the species demarcation within *Bacilladnaviridae* differs from most other families within *Cressdnaviricota*, where genome-wide pairwise nucleotide identity threshold of 75–78% is used instead (e.g. families *Circoviridae*, *Smacoviridae*, *Genomoviridae*, *Naryaviridae*, *Nenyaviridae* and *Vilyaviridae*) [37–40].

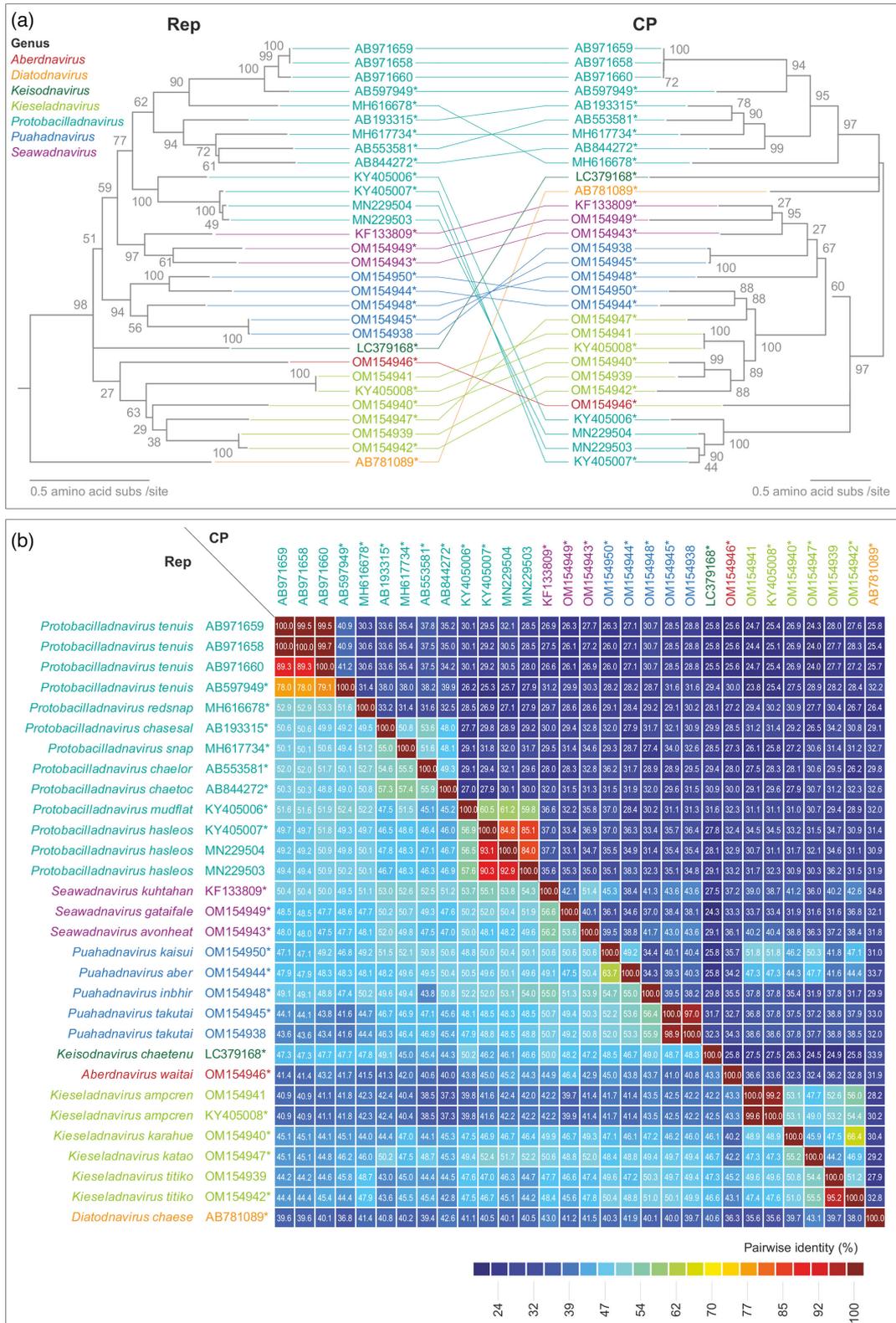
Sequences with accession numbers MH617734 and MH616678 represent two new species in the genus *Protobacilladnavirus*, and LC379168 represents a new species in a new genus, *Keisodnavirus*. The new bacilladnavirus sequences from Avon-Heathcote Estuary with accessions OM154943 and OM154949 represent two new species in the new genus *Seawadnavirus*; OM154940, OM154941, OM154942 and OM154947 represent new species in the genus *Kieseladnavirus*; and OM15494 is a member of the same species as the sequence with accession KY405008. Finally, the sequence with accession OM154938 (together with OM154945), OM154944, OM154948 and OM154950 represent four new species in the new genus *Puahadnavirus*, whereas OM154946 is the sole representative of a new genus *Aberdnavirus* (Table 1).

To verify whether our species demarcation criterion is robust, we generated a dataset of representative sequences from each species. Analysis of the genome sequences, as well as the Rep and CP amino acid sequences of these representative sequences ( $n=22$ ), shows that they share 56–65% (median 59.4%), 35–64% (median 47.7%) and 24–61% (median 31.8%), respectively (Fig. 4a), supporting the chosen species demarcation threshold.

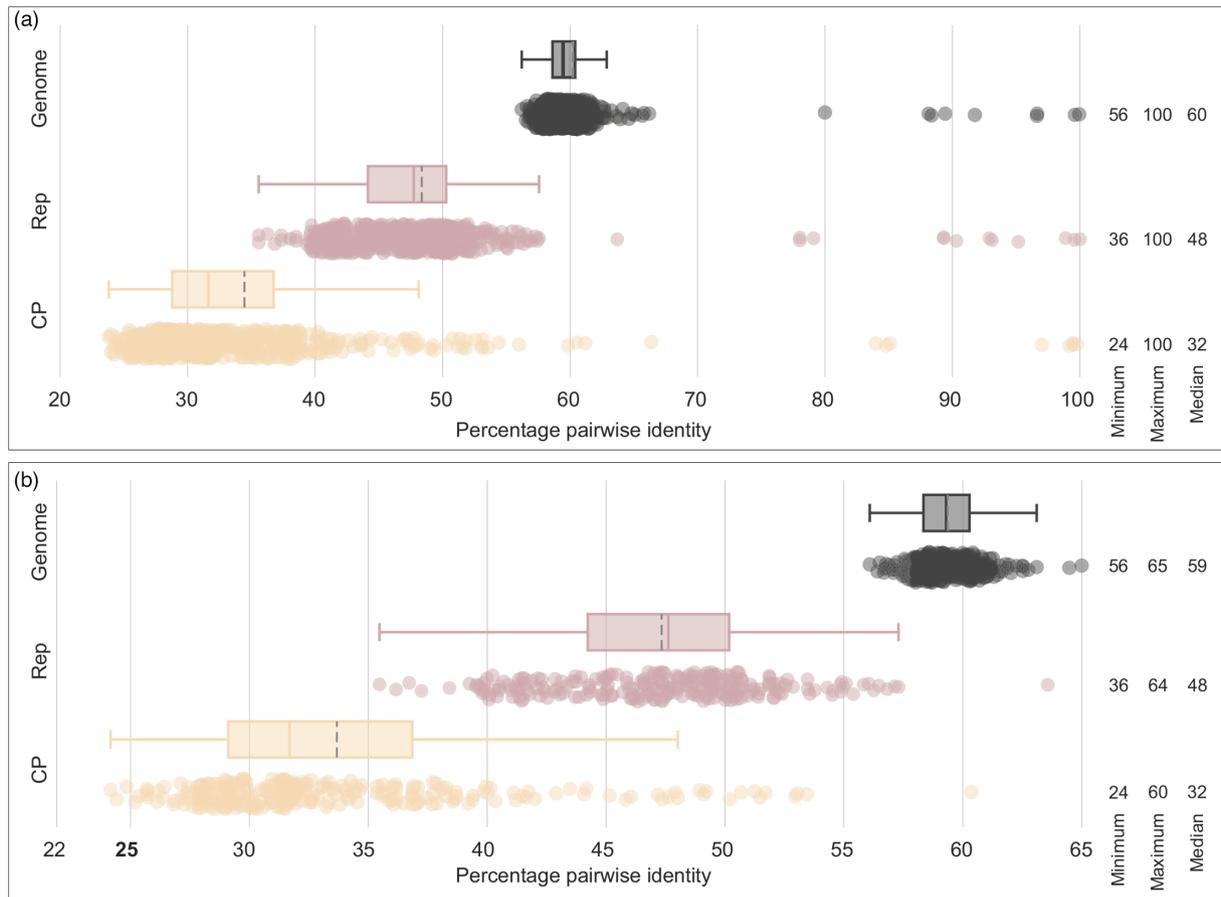
To comply with the mandated binomial species naming format [41], we have renamed all existing species using binomial nomenclature with a freeform epithet. All epithets are either derivatives of host species or isolation sources using names in local languages of people that inhabit that region or settlers (Table 1).

### Increasing diversity of bacilladnaviruses provides insight into their evolution

Although, by definition, Rep sequences within each genus are monophyletic, this is not the case for the CPs. The tanglegram of the Rep and CP phylogenies (Fig. 3a) showed that the CPs of members of the genera *Protobacilladnavirus* and *Puahadnavirus*



**Fig. 3.** Evolutionary histories of Rep and CP coding genes within bacilladnavirus genomes. (a) Tanglegram of maximum likelihood phylogenetic trees of the Rep and CP amino acid sequences. The Rep phylogenetic tree is rooted with representative sequences of circoviruses, whereas the CP phylogenetic tree is midpoint rooted. (b) Pairwise amino acid sequence identity matrix of the Rep and CP sequences. The representative sequences for each species are denoted with \* next to the accession number.



**Fig. 4.** Boxplots showing the distribution of the genome nucleotide and Rep and CP amino acid pairwise identities. Pairwise identities were calculated for (a) all bacilladnavirus sequences ( $n=30$ ) and (b) representative members from each species in the family *Bacilladnaviridae* ( $n=22$ ). The range and median are provided on the right of the plot. The grey dotted line in the box plots represents the mean, whereas the solid line across the box represents the median. All pairwise comparisons were calculated using SDT v1.2 [26].

are polyphyletic, suggesting that recombination has likely played a role in their evolution, as is the case for many other ssDNA viruses [37, 42–47]. Nevertheless, given the relatively high sequence conservation of the CPs within the family ( $\geq 24\%$  identity in any pairwise comparison, Fig. 3b), the recombination appears to have taken place within the family, with no evidence of non-orthologous gene replacements in the currently known bacilladnaviruses. Notably, whereas bacilladnavirus Reps are generally more conserved than CP at the sequence level, in several cases, this pattern is inverted (Fig. 3b), further pointing to potential recombination events.

Analysis of the genome length distributions revealed a considerable variation among members of the family (Fig. 5a). Although genomes of bacilladnaviruses from most genera fall in the range of 4.5–6 kb, the largest genomes, those in the genus *Protobacilladnavirus* [up to 6000 nt for *Chaetoceros salsugineum* DNA virus (AB193315)], are nearly twice the size of certain viruses in the newly created genus *Puahadnavirus* [as small as 3492 nt for *avonheates* virus SG\_479 (OM154950)]. Such size variation is likely to have consequences for the genome packaging density or the size and internal volume of the respective capsids. Comparison of the CPs encoded by viruses with the largest and smallest genomes, *Chaetoceros salsugineum* DNA virus and *avonheates* virus SG\_479, respectively, showed that the former is 59 aa longer than the latter (392 vs. 333 aa). Analysis of the CP multiple sequence alignment showed that the extended N-terminal region in the CP of *Chaetoceros salsugineum* DNA virus accounts for most of the extra residues (Fig. S1). By contrast, the central part of the CP encompassing the jelly-roll and projection domains (Fig. 6a), which form the capsid shell and spike-like structures that point away from the capsid surface [2, 4], was generally conserved, with only short insertions in some of the loops (Fig. S1). The largely unstructured N-terminal regions in the CP of *Chaetoceros salsugineum* DNA virus and other protobacilladnaviruses are enriched in positively charged amino acid residues and are oriented into the capsid lumen, where they are implicated in genome compaction. Similar N-terminal extensions, commonly known as the R-arm [[48]], are also found in the CPs of other eukaryotic ssDNA viruses, such as circoviruses [49] and anelloviruses [50–52]. Notably, the CP encoded by *avonheates* virus SG\_19 (OM154944), another member of the *Puahadnavirus*

**Table 1.** Summary of the taxonomy of bacilladnaviruses with additional information

Text in bold indicates new genera and species. Bold underlined text in the epithet notes highlights the source of the species epithet.

Genus	Accession	Virus name	Source	Country	Species	Epithet note	Reference
<b><i>Keisodnavirus</i></b>	LC379168	Chaetoceros tenuissimus DNA virus SS12-43V	<i>Chaetoceros tenuissimus</i>	Japan: Hiroshima Bay	<b><i>Keisodnavirus chaetenu</i></b>	<u><i>Chaetoceros tenuissimus</i></u>	–
<i>Protobacilladnavirus</i>	AB193315	Chaetoceros salsugineum DNA virus	<i>Chaetoceros salsugineum</i>	Japan: Fukuoka, Ariake Sea	<b><i>Protobacilladnavirus chasesal</i></b>	<u><i>Chaetoceros salsugineum</i></u>	[5]
	MH617734	Bacilladnaviridae sp. isolate ctia23	Red snapper tissue	USA	<b><i>Protobacilladnavirus snap</i></b>	Red <b>snapper</b>	[13]
	AB844272	Chaetoceros sp. DNA virus 7	<i>Chaetoceros</i> sp. SS628-11	Japan: Hiroshima, Hiroshima Bay	<b><i>Protobacilladnavirus chaetoc</i></b>	<u><i>Chaetoceros</i> sp.</u>	[8]
	AB553581	Chaetoceros lorenzianus DNA virus	<i>Chaetoceros lorenzianus</i>	Japan: Hiroshima, Itsukaichi Fishing Port	<b><i>Protobacilladnavirus chaelor</i></b>	<u><i>Chaetoceros lorenzianus</i></u>	[9]
	MH616678	Bacilladnaviridae sp. isolate ctdc18	Red snapper tissue	USA	<b><i>Protobacilladnavirus redsnap</i></b>	<b>Redsnapper</b>	[13]
	AB971660	Chaetoceros tenuissimus DNA virus type-II	<i>Chaetoceros tenuissimus</i> strain 2–10	Japan: Hiroshima, Hiroshima Bay	<b><i>Protobacilladnavirus tenuis</i></b>	<u><i>Chaetoceros tenuissimus</i></u>	[12]
	AB971659	Chaetoceros tenuissimus DNA virus type-II	<i>Chaetoceros tenuissimus</i> strain 2–10	Japan: Hiroshima, Hiroshima Bay	<b><i>Protobacilladnavirus tenuis</i></b>	<u><i>Chaetoceros tenuissimus</i></u>	[12]
	AB971658	Chaetoceros tenuissimus DNA virus type-II	<i>Chaetoceros tenuissimus</i> strain 2–10	Japan: Hiroshima, Hiroshima Bay	<b><i>Protobacilladnavirus tenuis</i></b>	<u><i>Chaetoceros tenuissimus</i></u>	[12]
	AB597949	Chaetoceros tenuissimus DNA virus	<i>Chaetoceros tenuissimus</i>	Japan: Fukuoka, Ariake Sound	<b><i>Protobacilladnavirus tenuis</i></b>	<u><i>Chaetoceros tenuissimus</i></u>	[7]
	KY405007	Amphibola crenata associated bacilladnavirus 2	<i>Amphibola crenata</i>	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<b><i>Protobacilladnavirus hasleos</i></b>	<u><i>Haslea ostrearia</i></u>	[2]
	MN229504	Haslea ostrearia associated bacilladnavirus	<i>Haslea ostrearia</i>	France: Bay of Bourgneuf	<b><i>Protobacilladnavirus hasleos</i></b>	<u><i>Haslea ostrearia</i></u>	[10]
	MN229503	Haslea ostrearia associated bacilladnavirus	<i>Haslea ostrearia</i>	France: Bay of Bourgneuf	<b><i>Protobacilladnavirus hasleos</i></b>	<u><i>Haslea ostrearia</i></u>	[10]
	KY405006	Amphibola crenata associated bacilladnavirus 1	<i>Amphibola crenata</i>	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<b><i>Protobacilladnavirus mudflat</i></b>	<b>Mud-flat</b> snail common name for <i>Amphibola crenata</i>	[2]
<b><i>Puahadnavirus</i></b>	OM154950	Avonheates virus SG_479	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<b><i>Puahadnavirus kaisui</i></b>	Seawater in Japanese	This study
	OM154944	Avonheates virus SG_19	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<b><i>Puahadnavirus aber</i></b>	Estuary in Welsh	This study
	OM154948	Avonheates virus SG_146	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<b><i>Puahadnavirus inbhir</i></b>	Estuary in Scottish	This study
	OM154945	Avonheates virus SG_28	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<b><i>Puahadnavirus takutai</i></b>	Coast in Māori	This study

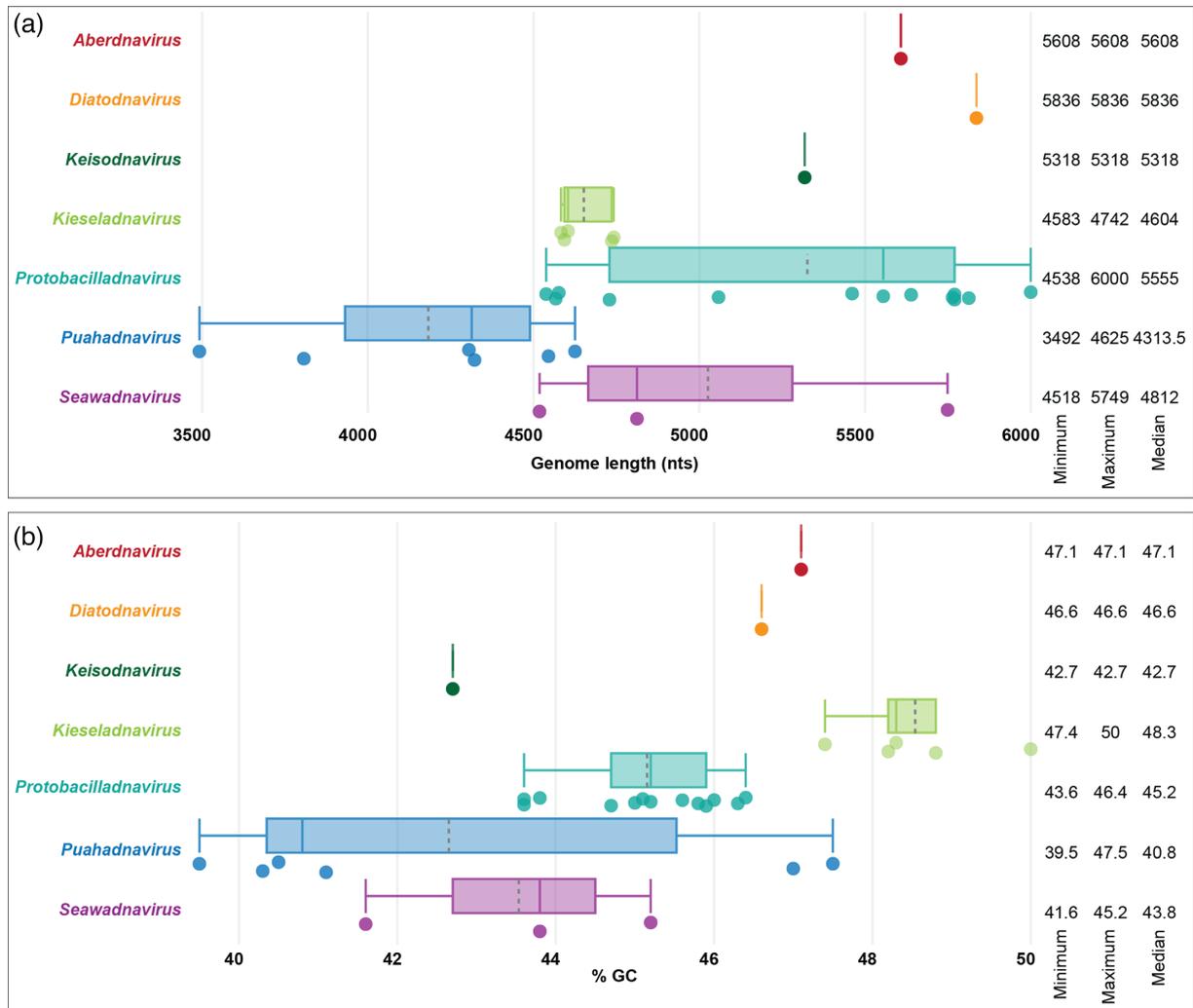
Continued

Table 1. Continued

Genus	Accession	Virus name	Source	Country	Species	Epithet note	Reference
<i>Seawadnavirus</i>	OM154938	Avonheates virus SG2_28	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<i>Puahadnavirus takutai</i>	Coast in Māori	This study
	OM154949	Avonheates virus SG_154	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<i>Seawadnavirus gataifale</i>	Estuary/coast in Samoan	This study
	KF133809	Bacillariodnavirus LDMD-2013	Ocean water	–	<i>Seawadnavirus kuhtahan</i>	Seawater in Massachusetts	[14]
<i>Kieseladnavirus</i>	OM154943	Avonheates virus SG_924	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<i>Seawadnavirus avonheat</i>	<u>Avon-Heathcote</u> estuary	This study
	OM154947	Avonheates virus SG_120	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<i>Kieseladnavirus katao</i>	Water in Māori	This study
	KY405008	Avon-Heathcote estuary associated bacilladnavirus	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<i>Kieseladnavirus ampcren</i>	<i>Amphibola Crenata</i>	[2]
	OM154942	Avonheates virus SG_4_10	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<i>Kieseladnavirus titiko</i>	Amphibola crenata in Māori	This study
	OM154939	Avonheates virus Gas_102	<i>Amphibola crenata</i>	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<i>Kieseladnavirus titiko</i>	Amphibola crenata in Māori	This study
	OM154941	Avonheates virus Gas_1207	<i>Amphibola crenata</i>	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<i>Kieseladnavirus ampcren</i>	<i>Amphibola Crenata</i>	This study
	OM154940	Avonheates virus Gas_1078	<i>Amphibola crenata</i>	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<i>Kieseladnavirus karahue</i>	Common name for <i>Amphibola crenata</i> in Māori	This study
	<i>Aberdnavirus</i>	OM154946	Avonheates virus SG_61	Estuary benthic sediments	New Zealand (Aotearoa): Christchurch, Avon-Heathcote estuary	<i>Aberdnavirus waitai</i>	Seawater in Māori
<i>Diatodnavirus</i>	AB781089	Chaetoceros setoense DNA virus	<i>Chaetoceros setoensis</i>	Japan: Hiroshima Bay	<i>Diatodnavirus chaese</i>	<i>Chaetoceros setoense</i>	[11]

genus with a relatively small genome (3807 nt), also lacks the R-arm. By contrast, the CPs of puahadnaviruses with genomes larger than 4 kb, similar to other bacilladnaviruses, such as avonheates virus SG\_4\_10 (OM154942), which is shown in Fig. 6(a) for comparison, contain the N-terminal regions enriched in arginine (R) and lysine (K) residues. Thus, the lack of the R-arm in the CPs of puahadnaviruses avonheates virus SG\_479 and avonheates virus SG\_19 suggests a relaxed necessity for the compaction of their smaller genomes. Consistently, the number of R and K residues in the R-arm (N-terminus) in the bacilladnavirus capsid proteins positively correlates with the genome length (Pearson correlation  $R=0.68$ , Fig. 6b). This result corroborates the conclusions of Requião *et al.* [48] that there is a positive correlation in the number of positively charged amino acids in the termini of capsid proteins of diverse DNA and RNA viruses with virus genome size. Indeed, given the conservation of the jelly-roll domains of bacilladnavirus CPs irrespective of their genome size (Fig. S1) the internal capsid volume of all bacilladnaviruses is expected to be similar, and hence, the accommodation of smaller genomes would require less compaction.

Notably, viruses in the *Puahadnavirus* genus displayed the largest variation in the %GC content (Fig. 5b). Although the exact hosts for bacilladnaviruses discovered by metagenomics remain unknown, the differences in %GC could reflect different host ranges for the corresponding viruses, as has been inferred for protist-infecting ssDNA viruses of the families *Naryaviridae*, *Nenyaviridae* and *Vilyaviridae* [45, 53].

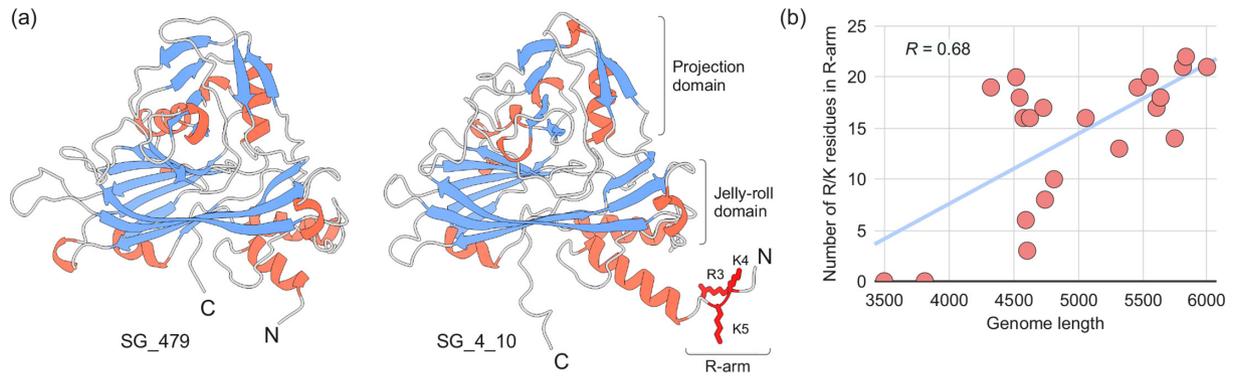


**Fig. 5.** Boxplots showing the distribution of genome size (a) and percentage GC (b) in the genomes for members of each genus. The range and median are provided on the right of the plot, and the grey dotted line in the box plots represents the mean, whereas the solid line across the box represents the median.

### Bacilladnaviruses encode a distinct phospholipase A1

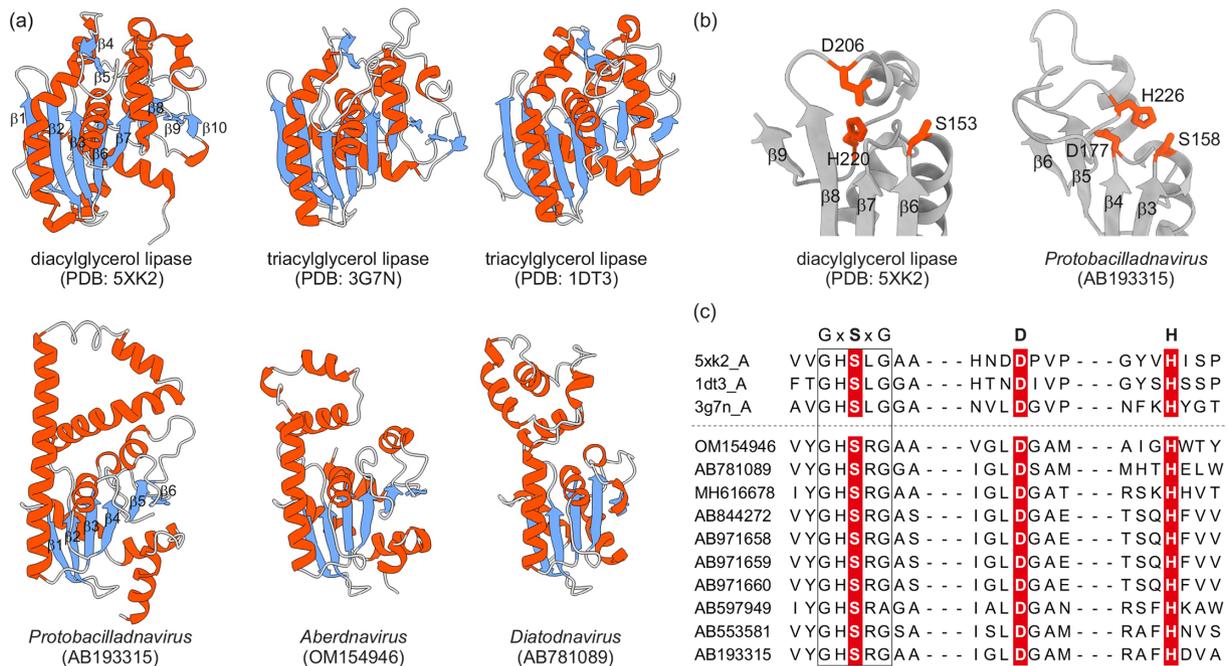
Owing to their larger genomes afforded by T=3 capsids, the coding capacity of bacilladnaviruses is larger compared to other members of *Cressdnaviricota*. However, the functions encoded by bacilladnaviruses besides the CP and Rep remain unknown. The expanded bacilladnavirus genome diversity prompted us to reassess the functional annotation of hypothetical ORFs using sensitive profile-profile comparisons with HHsearch [27]. This analysis revealed a group of orthologs encoded by a subset of protobacilladnaviruses, diatodnavirus and aberdnavirus, which yielded multiple significant hits to phospholipase A1 (PLA1) (Table S1), a family of enzymes that catalyse the cleavage at the *sn*-1 position of phospholipids, forming a fatty acid and a lysophospholipid [54].

The HHsearch hits to PLA1 encompassed the signature GxSxG (x, any residue) motif, which includes the catalytic Ser residue of the PLA1 active site but does not extend to the catalytic Asp and His residues [54]. Thus, to further assess the possibility that bacilladnaviruses encode an active PLA1, we built structural models of the corresponding proteins representing the three genera using ColabFold [28] (Fig. S2). Searches queried with the obtained models against the protein structures available at the PDB database using DALI [[29]] produced best hits to PLA1 structures (*Z*-scores 7.2–8.2, Table S2). Comparison of the PLA1 structures and bacilladnavirus homologues indeed revealed structural similarities but also highlighted notable differences (Fig. 7a). In particular, the first ( $\beta$ 1) and the last ( $\beta$ 10)  $\beta$ -strand of the core 8-stranded  $\beta$ -sheet conserved in PLA1 members is missing in the bacilladnavirus proteins. Additionally, the characteristic  $\beta$ -hairpin ( $\beta$ 4– $\beta$ 5) which follows  $\beta$ 3 of the core  $\beta$ -sheet, is also lacking in bacilladnavirus homologues (Fig. 7a). Notably, however, although the GxSxG motif is present at equivalent positions (after  $\beta$ 6



**Fig. 6.** Capsid proteins of bacilladnaviruses. (a) Comparison of the modelled capsid protein structures of avonheates virus SG\_479 (OM154950, species *Puahadnavirus kaisui*) and avonheates virus SG\_4\_10 (OM154942, species *Kieseladnavirus ampren*), which lack the R-arm and contain a short R-arm, respectively. The  $\beta$ -strands are shown in blue and the  $\alpha$ -helices in red. The arginine (r) and lysine (k) residues in the R-arm (N-terminus) are shown using stick representation. (b) Plot of the number of arginine (r) and lysine (k) in the R-arm of the bacilladnavirus capsid proteins relative to their genome size showing a positive Pearson correlation ( $R=0.68$ ).

in cellular PLA1 and  $\beta 3$  of bacilladnavirus proteins), the location of the two other predicted active site residues, Asp and His, is radically different. Whereas in cellular PLA1 homologues, the catalytic Asp and His are present in the loop connecting  $\beta$ -strands  $\beta 8$  and  $\beta 9$  [55–57], bacilladnavirus proteins contain invariable Asp and His positions in the loops following  $\beta$ -strands  $\beta 4$  and  $\beta 6$ , respectively (equivalent to  $\beta 7$  and  $\beta 9$  of cellular PLA1) (Fig. 6b). Importantly, despite being located within different loops in the bacilladnavirus proteins compared to the cellular PLA1 homologues, the three putative catalytic residues assume similar spatial positions (Fig. 7b). Furthermore, all three active site residues are invariable in the corresponding bacilladnavirus proteins



**Fig. 7.** Cellular and viral PLA1 homologues. (a) Structures of cellular (top row) and bacilladnavirus (bottom row) PLA1 homologues. Bacilladnavirus structures were modelled using ColabFold [28], whereas cellular PLA1 X-ray structures were downloaded from PDB, and the corresponding accession numbers are provided in parentheses. Structures are coloured according to the secondary structure elements:  $\beta$ -strands, blue;  $\alpha$ -helices, red; random coil, grey.  $\beta$ -strands are numbered for one cellular ( $\beta 1$ – $\beta 10$ ) and one viral ( $\beta 1$ – $\beta 6$ ) homolog (shown on the left). (b) Zoom-in on the core region encompassing the active site of cellular and viral PLA1 homologues. The experimentally determined cellular and predicted viral active site residues (red) are indicated using stick representation. (c) Alignment of the regions encompassing the putative active site residues (highlighted on red background) of bacilladnavirus PLA1 homologues. The catalytic Ser is located within a characteristic GxSxG motif (boxed), where x is any residue. Active site residues of the three cellular PLA1 proteins shown on the top were determined experimentally. Full-length alignment is provided in Fig. S3.

(Fig. S3), suggesting that the latter are functional, catalytically active phospholipases. Notably, one protobacilladnavirus (MH617734) encodes an inactivated homologue of PLA1, which lacks all of the active site residues (Table S1).

The role of PLA1 homologues in the life cycle of bacilladnaviruses is unclear. Notably, phospholipases have been previously described in viruses with small RNA and DNA genomes. Among these, the best understood are PLA2 homologues encoded by certain ssDNA viruses of the *Parvoviridae* family [58]. In parvoviruses, PLA2 is expressed as an N-terminal extension of a minor capsid protein and is deployed during the entry process to liberate the internalized virions from the endosomes [59, 60]. Subsequently, PLA2 domains were detected in RNA viruses, where they are also fused to the CP genes, suggesting a similar function as in parvoviruses [61]. The presence of PLA1 genes in viruses is less common, but not unprecedented. For instance, certain polinton-like viruses (*Aquintoviricetes*) and virophages (*Maviviridae*) sporadically encode PLA1 homologues [62, 63]. In the case of virophage mavirus, the virus-encoded PLA1 was detected in the virions [64], suggesting that it plays a role during virion entry, like in parvoviruses. Notably, SDS-PAGE analysis of the bacilladnavirus virions consistently revealed the presence of two major structural proteins [2, 5, 7–10, 12]. However, the identity of the two proteins was not determined. Thus, although structural studies showed that the capsid shell is built from a single CP [4], it cannot be excluded that the PLA1 is present in the capsid lumen, as in the case of parvoviruses, and is involved in the virion liberation from the endosomes during entry. Alternatively or additionally, given that bacilladnaviruses lyse their diatom hosts at the end of the infection cycle, the PLA1 homologues could facilitate the virion egress. Notably, it has been recently demonstrated that the infection of *Chaetoceros tenuissimus* with protobacilladnavirus CtenDNAV resulted in a more statistically significant response in the lipidome than infection with a diatom RNA virus. In particular, virus-mediated lysis led to the production of chemical defence compounds called oxylipins, which are toxic chemical signalling molecules that are made by oxidizing fatty acids from membrane lipids [65]. We hypothesize that viral PLA1 could play a role in this process.

## CONCLUDING REMARKS

Here, we describe the identification of three bacilladnavirus genomes from mud-flat snail/titiko and ten from benthic sediments sampled from the Avon-Heathcote Estuary (Ihutai) in New Zealand (Aotearoa). We analysed these genomes along with other unclassified and classified bacilladnavirus genomes available in GenBank and based on the phylogeny of the Rep sequences, established four new genera (*Aberdnavirus*, *Keisodnavirus*, *Puahadnavirus* and *Seawadnavirus*) to classify the new viruses and move one species from the genus *Protobacilladnavirus* (accession no. KF133809) to the new genus, *Seawadnavirus*. The following are based on the 75% Rep amino acid sequence pairwise identity threshold we established:

- (1) One new species in the new genus *Aberdnavirus*
- (2) Two new species in the genus *Kieseladnavirus*
- (3) One new species in the new genus *Keisodnavirus*
- (4) Two new species in the genus *Protobacilladnavirus*
- (5) Five new species in the new genus *Puahadnavirus*
- (6) Two new species in the new genus *Seawadnavirus* and moved one species from *Protobacilladnavirus* to this genus

Furthermore, we renamed existing species and named the new species using a binomial format with a freeform epithet to comply with the mandated binomial species naming format. All taxonomic changes described above have been approved and officially ratified by the International Committee on Taxonomy of Viruses (ICTV) [16].

As more bacilladnaviruses are discovered, it is likely that the taxonomy will change. Thus, we discourage the community from using names of taxonomic ranks as part of the virus names, so that subsequent reassignment of a virus into a different taxon, as in the case of KF133809 which was moved from *Protobacilladnavirus* to genus *Seawadnavirus*, does not render the virus names misleading (e.g. protobacilladnavirus 1 becoming a member of genus *Seawadnavirus*). To avoid such situations, here we have named all 13 newly identified viruses as avonheates viruses (e.g. avonheates virus SG\_61 that is deposited under OM154946), which stands for **Avon Heathcote estuary** followed by a unique code. Further, we also keep in mind that there is a clear distinction between a species name and a virus name; see Zerbini *et al.* [66].

All members of the community studying bacilladnaviruses are encouraged to submit taxonomy proposals for new species and genera to the ICTV (see <https://ictv.global/>). The Virus Metadata Resource (<https://ictv.global/vmr>) is updated annually and is an excellent resource for metadata on species and representative members of each species.

Finally, the analysis of the expanded diversity of bacilladnavirus genome and protein sequences provided new insights into the biology and evolution of this understudied group of viruses. Comparison of the bacilladnavirus CP sequences suggests that the R-arm-mediated compaction is only required for encapsidation of the larger genomes, whereas smaller genomes can be packaged in the absence of the R-arm subdomain. The presence of a highly derived PLA1 in members of three *Bacilladnaviridae* genera suggests a mechanism of penetration through the cellular membranes during the entry and/or egress stages. Notably, bacilladnaviruses, which do not encode recognizable PLA1 homologues, carry hypothetical genes at the equivalent position. We hypothesize that the products of these genes play analogous roles during the infection. Testing the activities of the divergent

bacilladnavirus PLA1 homologues, as well as those of their positional counterparts, appears to be a promising research direction that can uncover enzymes with new functionalities and biotechnological potential.

#### Funding information

Part of the work presented here was supported by The Brian Mason Scientific and Technical Trust funds awarded to AV, AD, and SG.

#### Conflicts of interest

The authors declare there are no conflicts of interest.

#### References

- Krupovic M, Varsani A, Kazlauskas D, Breitbart M, Delwart E, et al. *Cressdnaviricota*: a virus phylum unifying seven families of re-encoding viruses with single-stranded, circular DNA genomes. *J Virol* 2020;94:e00582-20.
- Kazlauskas D, Dayaram A, Kraberger S, Goldstien S, Varsani A, et al. Evolutionary history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA nodaviruses. *Virology* 2017;504:114–121.
- Kazlauskas D, Varsani A, Krupovic M. Pervasive chimerism in the replication-associated proteins of uncultured single-stranded DNA viruses. *Viruses* 2018;10:187.
- Munke A, Kimura K, Tomaru Y, Wang H, Yoshida K, et al. Primordial capsid and spooled ssDNA genome structures unravel ancestral events of eukaryotic viruses. *mBio* 2022;13:e0015622.
- Nagasaki K, Tomaru Y, Takao Y, Nishida K, Shirai Y, et al. Previously unknown virus infects marine diatom. *Appl Environ Microbiol* 2005;71:3528–3535.
- Tomaru Y, Shirai Y, Suzuki H, Nagasaki T, Nagumo T. Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom *Chaetoceros debilis*. *Aquat Microb Ecol* 2008;50:103–112.
- Tomaru Y, Shirai Y, Toyoda K, Nagasaki K. Isolation and characterisation of a single-stranded DNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus*. *Aquat Microb Ecol* 2011;64:175–184.
- Kimura K, Tomaru Y. Isolation and characterization of a single-stranded DNA virus infecting the marine diatom *Chaetoceros* sp. strain SS628-11 isolated from western Japan. *PLoS One* 2013;8:e82013.
- Tomaru Y, Takao Y, Suzuki H, Nagumo T, Koike K, et al. Isolation and characterization of a single-stranded DNA virus infecting *Chaetoceros lorenzianus* grunow. *Appl Environ Microbiol* 2011;77:5285–5293.
- Gastineau R, Lemieux C, Turmel M, Grypioti E, Verret F, et al. Two new bacilladnaviruses associated with the diatom *Haslea ostrearia*. *Eur J Phycol* 2020;55:444–453.
- Tomaru Y, Toyoda K, Suzuki H, Nagumo T, Kimura K, et al. New single-stranded DNA virus with a unique genomic structure that infects marine diatom *Chaetoceros setoensis*. *Sci Rep* 2013;3:3337.
- Kimura K, Tomaru Y. Discovery of two novel viruses expands the diversity of single-stranded DNA and single-stranded RNA viruses infecting a cosmopolitan marine diatom. *Appl Environ Microbiol* 2015;81:1120–1131.
- Tisza MJ, Pastrana DV, Welch NL, Stewart B, Peretti A, et al. Discovery of several thousand highly diverse circular DNA viruses. *Elife* 2020;9:e51971.
- McDaniel LD, Rosario K, Breitbart M, Paul JH. Comparative metagenomics: natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences. *Environ Microbiol* 2014;16:570–585.
- King AMQ, Lefkowitz EJ, Mushegian AR, Adams MJ, Dutilh BE, et al. Changes to taxonomy and the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2018). *Arch Virol* 2018;163:2601–2631.
- Zerbini FM, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, et al. Changes to virus taxonomy and the ICTV statutes ratified by the international committee on taxonomy of viruses (2023). *Arch Virol* 2023;168:175.
- Kazlauskas D, Varsani A, Koonin EV, Krupovic M. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat Commun* 2019;10:3425.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinform* 2014;30:2114–2120.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017;27:824–834.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
- Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinform* 2011;27:1164–1165.
- Pei J, Grishin NV. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol Biol* 2014;1079:263–271.
- Procter JB, Carstairs GM, Soares B, Mourão K, Ofoegbu TC, et al. Alignment of biological sequences with jalview. *Methods Mol Biol* 2021;2231:203–224.
- Muhire BM, Varsani A, Martin DP. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 2014;9:e108277.
- Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinform* 2019;20:473.
- Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, et al. ColabFold: making protein folding accessible to all. *Nat Methods* 2022;19:679–682.
- Holm L. Dali server: structural unification of protein families. *Nucleic Acids Res* 2022;50:W210–W215.
- Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, et al. UCSF chimeraX: tools for structure building and analysis. *Protein Sci* 2023;32:e4792.
- McClatchie S, Juniper SK, Knox GA. Structure of a mudflat diatom community in the avon-heathcote estuary, New Zealand. *N Z J Mar Freshwater Res* 1982;16:299–309.
- Chandler M, de la Cruz F, Dyda F, Hickman AB, Moncalian G, et al. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol* 2013;11:525–538.
- Ilyina TV, Koonin EV. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria. *Nucleic Acids Res* 1992;20:3279–3285.
- Koonin EV, Ilyina TV. Computer-assisted dissection of rolling circle DNA replication. *Biosystems* 1993;30:241–268.

35. Gorbalenya AE, Koonin EV. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol* 1993;3:419–429.
36. Gorbalenya AE, Koonin EV, Wolf YI. A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett* 1990;262:145–148.
37. Krupovic M, Varsani A. *Naryaviridae*, *Nenyaviridae*, and *Vilyaviridae*: three new families of single-stranded DNA viruses in the phylum *Cressdnaviricota*. *Arch Virol* 2022;167:2907–2921.
38. Varsani A, Harrach B, Roumagnac P, Benkő M, Breitbart M, et al. 2024 taxonomy update for the family *Circoviridae*. *Arch Virol* 2024;169:176.
39. Varsani A, Krupovic M. Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family *Genomoviridae*. *Virus Evol* 2017;3:vev037.
40. Varsani A, Krupovic M. Smacoviridae: a new family of animal-associated single-stranded DNA viruses. *Arch Virol* 2018;163:2005–2015.
41. Siddell SG, Walker PJ, Lefkowitz EJ, Mushegian AR, Dutilh BE, et al. Binomial nomenclature for virus species: a consultation. *Arch Virol* 2020;165:519–525.
42. Lefevre P, Lett JM, Varsani A, Martin DP. Widely conserved recombination patterns among single-stranded DNA viruses. *J Virol* 2009;83:2697–2707.
43. Martin DP, Biagini P, Lefevre P, Golden M, Roumagnac P, et al. Recombination in eukaryotic single stranded DNA viruses. *Viruses* 2011;3:1699–1738.
44. de la Higuera I, Kasun GW, Torrance EL, Pratt AA, Maluenda A, et al. Unveiling crucivirus diversity by mining metagenomic data. *mBio* 2020;11:e01410-20.
45. Kinsella CM, Bart A, Deijs M, Broekhuizen P, Kaczorowska J, et al. Entamoeba and giardia parasites implicated as hosts of CRESS viruses. *Nat Commun* 2020;11:4620.
46. Krupovic M, Zhi N, Li J, Hu G, Koonin EV, et al. Multiple layers of chimerism in a single-stranded DNA virus discovered by deep sequencing. *Genome Biol Evol* 2015;7:993–1001.
47. Roux S, Enault F, Bronner G, Vaulot D, Forterre P, et al. Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. *Nat Commun* 2013;4:2700.
48. Requião RD, Carneiro RL, Moreira MH, Ribeiro-Alves M, Rossetto S, et al. Viruses with different genome types adopt a similar strategy to pack nucleic acids based on positively charged protein domains. *Sci Rep* 2020;10:5470.
49. Sarker S, Terrón MC, Khandokar Y, Aragão D, Hardy JM, et al. Structural insights into the assembly and regulation of distinct viral capsid complexes. *Nat Commun* 2016;7:13014.
50. Butkovic A, Kraberger S, Smeele Z, Martin DP, Schmidlin K, et al. Evolution of anelloviruses from a circovirus-like ancestor through gradual augmentation of the jelly-roll capsid protein. *Virus Evol* 2023;9:vead035.
51. Liou S-H, Boggavarapu R, Cohen NR, Zhang Y, Sharma I, et al. Structure of anellovirus-like particles reveal a mechanism for immune evasion. *Nat Commun* 2024;15:7219.
52. De Koch MD, Krupovic M, Fielding R, Smith K, Schiavone K, et al. Novel lineage of anelloviruses with large genomes identified in dolphins. *J Virol* 2025;99:e0137024.
53. Varsani A, Hopkins A, Lund MC, Krupovic M. 2024 taxonomic update for the families *Naryaviridae*, *Nenyaviridae*, and *Vilyaviridae*. *Arch Virol* 2024;170:18.
54. Richmond GS, Smith TK. Phospholipases A<sub>1</sub>. *Int J Mol Sci* 2011;12:588–612.
55. Bian C, Yuan C, Chen L, Meehan EJ, Jiang L, et al. Crystal structure of a triacylglycerol lipase from *Penicillium expansum* at 1.3 Å determined by sulfur SAD. *Proteins* 2010;78:1601–1605.
56. Brzozowski AM, Savage H, Verma CS, Turkenburg JP, Lawson DM, et al. Structural origins of the interfacial activation in *Thermomyces (Humicola) lanuginosa* Lipase. *Biochemistry* 2000;39:15071–15082.
57. Lan D, Zhao G, Holzmann N, Yuan S, Wang J, et al. Structure-guided rational design of a mono- and diacylglycerol lipase from *Aspergillus oryzae*: a single residue mutant increases the hydrolysis ability. *J Agric Food Chem* 2021;69:5344–5352.
58. Cotmore SF, Agbandje-McKenna M, Canuti M, Chiorini JA, Eis-Hubinger A-M, et al. ICTV virus taxonomy profile: Parvoviridae. *J Gen Virol* 2019;100:367–368.
59. Cotmore SF, Tattersall P. Parvoviral host range and cell entry mechanisms. *Adv Virus Res* 2007;70:183–232.
60. Farr GA, Zhang L, Tattersall P. Parvoviral virions deploy a capsid-tethered lipolytic enzyme to breach the endosomal membrane during cell entry. *Proc Natl Acad Sci USA* 2005;102:17148–17153.
61. Wolf YI, Silas S, Wang Y, Wu S, Bocek M, et al. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat Microbiol* 2020;5:1262–1270.
62. Koonin EV, Fischer MG, Kuhn JH, Krupovic M. The polinton-like supergroup of viruses: evolution, molecular biology, and taxonomy. *Microbiol Mol Biol Rev* 2024;88:e0008623.
63. Yutin N, Shevchenko S, Kapitonov V, Krupovic M, Koonin EV. A novel group of diverse Polinton-like viruses discovered by metagenome analysis. *BMC Biol* 2015;13:95.
64. Born D, Reuter L, Mersdorf U, Mueller M, Fischer MG, et al. Capsid protein structure, self-assembly, and processing reveal morphogenesis of the marine virophage mavirus. *Proc Natl Acad Sci USA* 2018;115:7332–7337.
65. Edwards BR, Thamatrakoln K, Fredricks HF, Bidle KD, VanMooy BAS. Viral infection leads to a unique suite of allelopathic chemical signals in three diatom host-virus pairs. *Mar Drugs* 2024;22:228.
66. Zerbini FM, Siddell SG, Mushegian AR, Walker PJ, Lefkowitz EJ, et al. Differentiating between viruses and virus species by writing their names correctly. *Arch Virol* 2022;167:1231–1234.
67. Gilchrist CLM, Chooi YH. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinform* 2021;37:2473–2475.
68. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* 2004;14:1188–1190.

**The Microbiology Society is a membership charity and not-for-profit publisher.**

**Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.**

**Find out more and submit your article at [microbiologyresearch.org](https://microbiologyresearch.org)**