This form should be used for all taxonomic proposals. Please complete all those modules that are applicable.

For guidance, see the notes written in blue and the separate document “Help with completing a taxonomic proposal”

Please try to keep related proposals within a single document.

Part 1: **TITLE, AUTHORS, etc**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Code assigned:** | ***2017.001F*** | | | | (to be completed by ICTV officers) |
| **Short title:** (e.g. 6 new species in the genus *Zetavirus*) Create two new genera within a new family *Bacilladnaviridae*; also rename an existing genus and move from unassigned to the new family. | | | | | |
| **Modules attached**  (Modules 1, 4 and either 2 or 3 are required. | | **1**  **2  3  4** | | | |
| **Author(s):** | | | | | |
| Mart Krupovic  Darius Kazlauskas  Arvind Varsani | | | | | |
| **Corresponding author with e-mail address:** | | | | | |
| Mart Krupovic; E-mail: [krupovic@pasteur.fr](mailto:krupovic@pasteur.fr) | | | | | |
| **List the ICTV study group(s) that have seen this proposal:** | | | | | |
| A list of study groups and contacts is provided at <http://www.ictvonline.org/subcommittees.asp> . If in doubt, contact the appropriate subcommittee chair (there are six virus subcommittees: animal DNA and retroviruses, animal ssRNA-, animal ssRNA+, fungal and protist, plant, bacterial and archaeal) | | | **Fungal and Protist Viruses Subcommittee** | | |
| **ICTV Study Group comments (if any) and response of the proposer:** | | | | | |
|  | | | | | |
|  | | | | | |
| Date first submitted to ICTV: | | | | 02.05.2017 | |
| Date of this revision (if different to above): | | | |  | |

|  |
| --- |
| **ICTV-EC comments and response of the proposer:** |
|  |

**Part 2**: **PROPOSED TAXONOMY**

|  |
| --- |
| Present the proposed new taxonomy on accompanying spreadsheet |
| **Name of accompanying spreadsheet:** **2017.001F.N.v1.Bacilladnaviridae.xlsx** |

Please display the taxonomic changes you are proposing on the accompanying spreadsheet module 2017\_TP\_Template\_Excel\_module. Submit both this and the spreadsheet to the appropriate ICTV Subcommittee Chair.

**Part 3:** **NON-STANDARD**

Template for any proposal regarding ICTV procedures, rules or policy, not involving the creation of new taxonomy.

|  |
| --- |
| non-standard proposal |
| **Title of proposal:** |
| **Text of proposal:** |
|  |

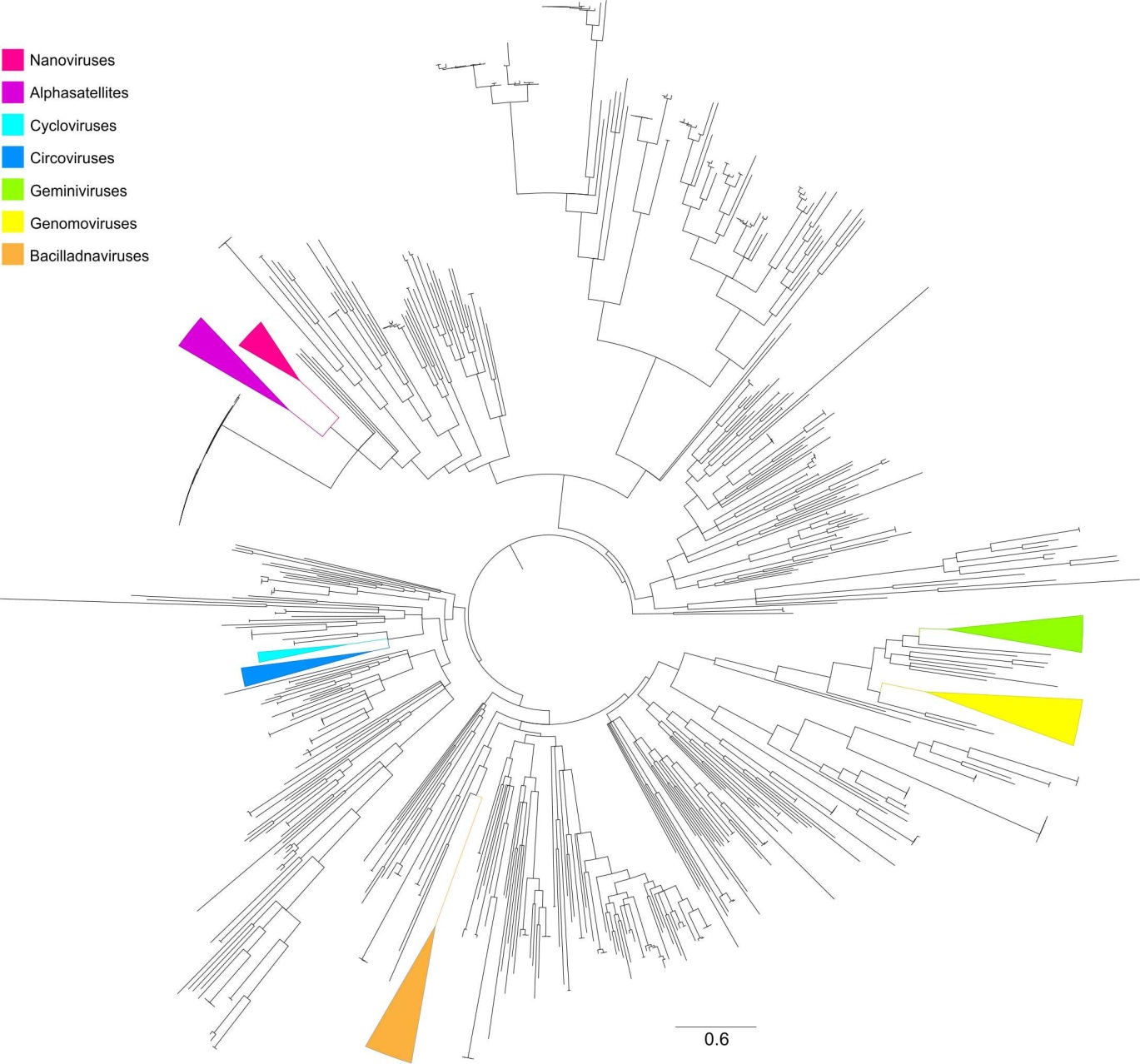
**Part 4:** **APPENDIX**: supporting material

Chaetoceros salsugineum DNA virus 01 (CsalDNAV) infects diatoms, a major group of unicellular algae (class Bacillariophyceae), and is currently the only officially classified member of the unassigned genus Bacilladnavirus (Nagasaki et al., 2005). However, several potential members of the genus infecting different diatom species have been isolated and their genomes completely sequenced. These include Chaetoceros sp. DNA virus 7 (Csp07DNAV) infecting Chaetoceros sp. strain SS628-11 (Kimura and Tomaru, 2013), Chaetoceros lorenzianus DNA virus (ClorDNAV; AB553581) infecting Ch. lorenzianus (Tomaru et al., 2011b), Chaetoceros setoensis DNA virus (CsetDNAV) of Ch. setoensis (Tomaru et al., 2013), and Chaetoceros tenuissimus DNA viruses (CtenDNAV) type I (Tomaru et al., 2011a) and type II (Kimura and Tomaru, 2015) infecting Ch. tenuissimus. Partial genome sequences are available for Chaetoceros virus YT-2008 (Csp05DNAV) of Chaetoceros sp. strain TG07-C28 (Toyoda et al., 2012), Chaetoceros debilis DNA virus (CdebDNAV) infecting Chaetoceros debilis (Tomaru et al., 2008) and Thalassionema nitzschioides DNA virus (TnitDNAV) infecting Thalassionema nitzschioides (Tomaru et al., 2012). In addition, 4 putative bacilladnavirus genomes have been assembled from metagenomic sequences (Kazlauskas et al., 2017; McDaniel et al., 2014). Characterized bacilladnaviruses carry circular ssDNA genomes of ~4.5-6 kb encapsidated within isometric particles of 33-38 nm in diameter (Kimura and Tomaru, 2015). It has been demonstrated that digestion of the bacilladnavirus genomes with S1 nuclease results in short linear dsDNA fragments of variable sizes (<1kb), suggesting that the genomes might be partially double-stranded. The number and location of such fragments within the viral genomes varies (Kimura and Tomaru, 2015; Tomaru et al., 2013), whereas their function, if any, remains unresolved.

Here, we propose a classification scheme for all previously unclassified putative bacilladnaviruses for which complete genome sequences are available.

**Bacilladnaviruses represent a unique group of ssDNA viruses**

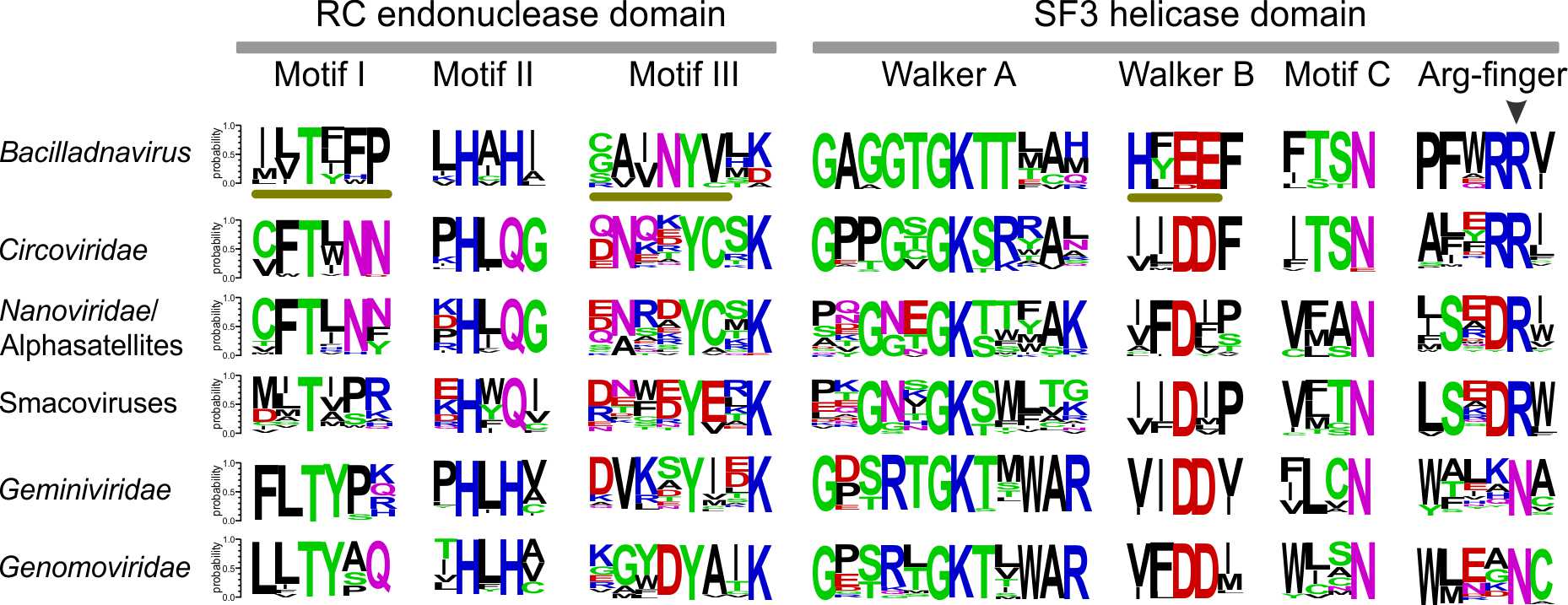
To investigate the affinity of the bacilladnavirus group to other CRESS DNA viruses, we performed phylogenetic analysis of their replication-initiation proteins (Reps) in the context of Rep sequences from a large collection of classified and unclassified CRESS DNA viruses (for circular Rep-encoding ssDNA viruses) available in GenBank (downloaded on the 14th Dec 2016) (Fig. 1). Phylogenetic analysis has unequivocally shown that bacilladnavirus Reps form a well-supported monophyletic clade, separated from other groups of ssDNA viruses.



**Figure 1.** Approximate Maximum-likelihood phylogenetic tree (mid-point rooted) of the Rep amino acid sequences of CRESS DNA viruses inferred using FastTree. The classified virus groups are indicated with colored triangles and the key is provided in the top left corner of the figure. Branches with less than 50% SH (Shimodaira–Hasegawa)-like support were collapsed. The scale bar represents the number of substitutions per site. The figure is reproduced from Kazlauskas et al., 2017.

The Rep sequences of bacilladnaviruses display a set of conserved motifs distinct from those typical of Reps of other ssDNA viruses. The nuclease domain of Rep contains three conserved motifs (Ilyina and Koonin, 1992). Motif I, which is thought to be involved in the recognition of iterative sequences associated with the origin of replication, and Motif III, which contains the catalytic tyrosine residue, are unique to bacilladnaviruses (Fig. 2). By contrast, Motif II of bacilladnaviruses contains two invariant histidine residues required for coordination of catalytically-important divalent metal ions (Mg2+ or Mn2+) and is thus most similar to the corresponding motif of geminiviruses and genomoviruses (Fig. 2).

Among the three conserved motifs previously defined in the helicase domain of viral Reps (Krupovic, 2013; Rosario et al., 2012; Gorbalenya et al., 1990), Walker A and Walker B motifs of bacilladnaviruses are distinct from those found in other viral Reps (Fig. 2). Notably, bacilladnavirus Walker B motif contains two conserved glutamate residues, which are typically not observed in other CRESS DNA viral Reps (Fig. 2). By contrast, Motif C and the Arg-finger of bacilladnaviruses are most similar to those of circoviruses (Kazlauskas et al., 2017).



**Figure 2.** Sequence motifs of Rep proteins from eukaryotic ssDNA viruses. Bacilladnavirus-specific motifs are underlined. Residues are colored according to their chemical properties (polar, green; basic, blue; acidic, red; hydrophobic, black; neutral, purple). The figure is reproduced from Kazlauskas et al., 2017.

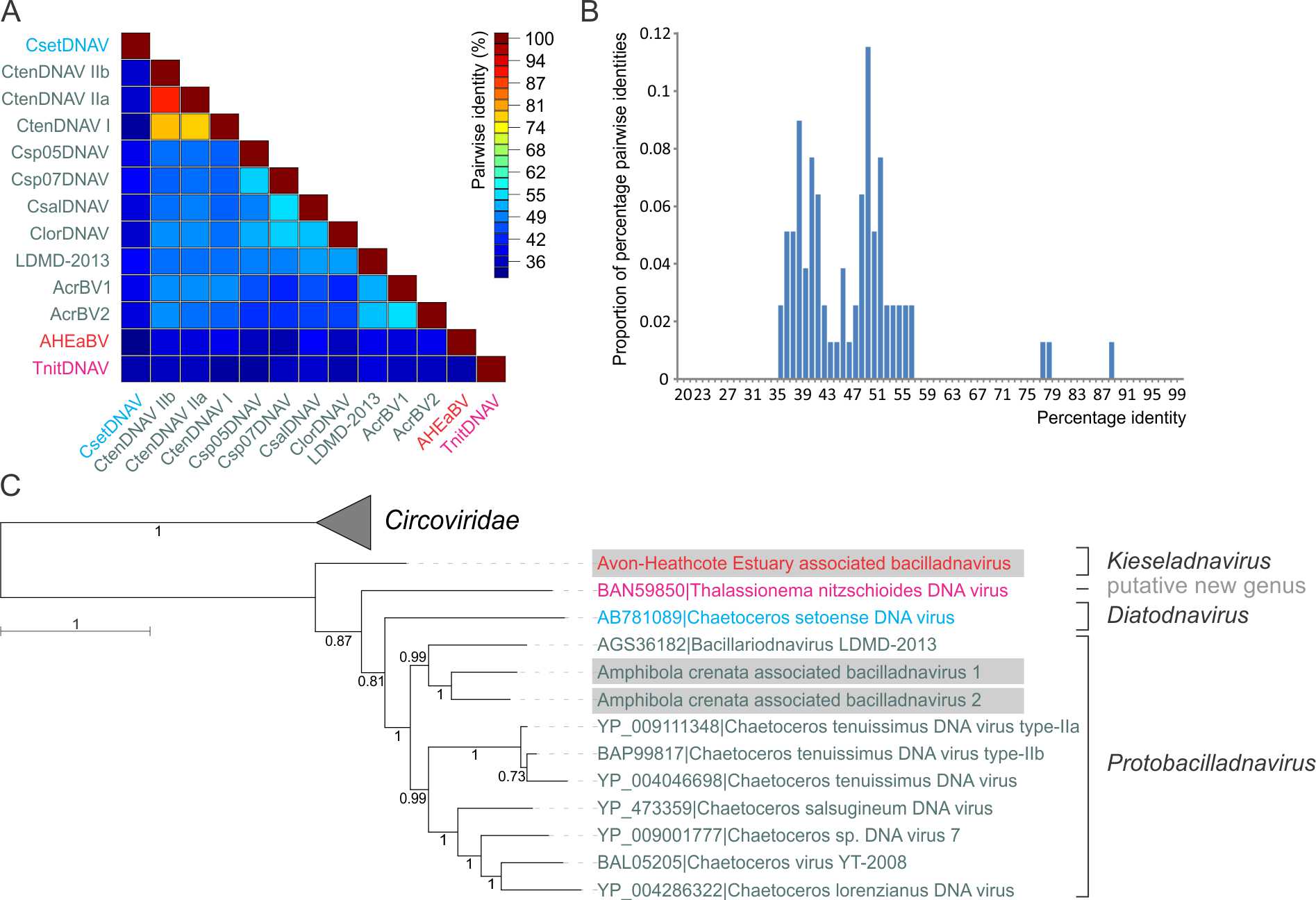
Finally, the capsid protein of bacilladnaviruses is unrelated to those of other known ssDNA viruses. Sensitive profile-profile comparisons show that the bacilladnavirus CP is most closely related to the CP of positive-sense RNA viruses of the *Nodaviridae* family (Kazlauskas et al., 2017).

Based on the above, it is clear that members of the floating genus *Bacilladnavirus* are distinct from all other classified ssDNA viruses and cannot be included into any of the currently existing virus families.

**Sequence diversity within the ‘Bacilladnavirus’ group**

To investigate the extent of sequence diversity within and the taxonomic structure of the Bacilladnavirus group, we focused on the Reps, which are the most conserved proteins encoded by bacilladnaviruses. We analyzed the distribution of pairwise identities (one minus Hamming distances of pairwise aligned sequences with pairwise deletion of gaps) across all available bacilladnavirus Rep sequences (Figure 3A, Table S3). Most of the Reps in our dataset share 36-58% pairwise identities and only 3 Reps display higher identity values (≥78% identity; Figure 3B), indicating that sequence diversity among bacilladnaviruses remains largely unexplored.

To better understand the relationships among bacilladnaviruses, we constructed a phylogenetic tree of their Reps (Figure 3C). The tree was rooted with the sequences of circoviruses, a group of viruses with Reps that are most similar to those of bacilladnaviruses. Consistent with results of the pairwise comparisons shown in Figure 3A, the phylogenetic analysis revealed 4 potential groups within the Bacilladnavirus assemblage (Figure 3C). The largest group includes all viruses, except for the CsetDNAV, infecting Chaetoceros species as well as uncultured viruses AcrBV1, AcrBV2 and Bacillariodnavirus LDMD-2013. By contrast, CsetDNAV, TnitDNAV, and AHEaBV are only distantly related to each other and to the major clade of Chaetoceros-infecting viruses. To reflect this diversity within the Bacilladnavirus group, the three singletons and the major clade should be considered as different genera. Notably, however, the complete genome sequence for TnitDNAV is unavailable, precluding its classification at this point. Consequently, the classification of bacilladnaviruses has to be revised and the current genus Bacilladnavirus, unavoidably, upgraded to the family level.



**Figure 3.** Sequence diversity and phylogenetic analysis of bacilladnavirus Reps. **A.** Matrix of pairwise identities between Rep sequences of all available bacilladnaviruses. Actual pairwise identity values are provided in Table 1. Abbreviations: CsalDNAV, Chaetoceros salsugineum DNA virus; Csp07DNAV, Chaetoceros sp. DNA virus 7; ClorDNAV, Chaetoceros lorenzianus DNA virus; CsetDNAV, Chaetoceros setoensis DNA virus; CtenDNAV I, IIa and IIb, Chaetoceros tenuissimus DNA viruses type I, IIa and IIb (SS10-35V), respectively; TnitDNAV, Thalassionema nitzschioides DNA virus; Csp05DNAV, Chaetoceros virus YT-2008; LDMD-2013, Bacillariodnavirus LDMD-2013; AcrBV1 and AcrBV2, Amphibola crenata associated bacilladnaviruses 1 and 2; AHEaBV, Avon-Heathcote Estuary associated bacilladnavirus. The accession numbers of the corresponding proteins are provided in panel C. **B.** Distribution of Rep amino acid pairwise identities among bacilladnaviruses. **C.** Phylogenetic tree of bacilladnavirusRep proteins, rooted with the Rep sequences of circoviruses. Newly sequenced genomes are shown in grey background. Numbers at the branch points represent the Bayesian-like transformation of aLRT (aBayes) local support values. The scale bar represents the number of substitutions per site. The major group of bacilladnaviruses and the 3 singletons are indicated with different colors in panels A and C. Panels A and C are reproduced from Kazlauskas et al., 2017.

**Table 1.** Matrix of pairwise identities between Rep amino acid sequences of all available bacilladnaviruses.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Virus name\* |  | Pairwise identity value | | | | | | | | | | | | |
| AB781089|CsetDNAV |  | 100 |  |  |  |  |  |  |  |  |  |  |  |  |
| BAP99817|CtenDNAV IIb |  | 40 | 100 |  |  |  |  |  |  |  |  |  |  |  |
| BAP99811|CtenDNAV IIa |  | 39 | 89 | 100 |  |  |  |  |  |  |  |  |  |  |
| BAJ40167|CtenDNAV I |  | 36 | 79 | 78 | 100 |  |  |  |  |  |  |  |  |  |
| BAL05205|Csp05DNAV |  | 42 | 50 | 50 | 48 | 100 |  |  |  |  |  |  |  |  |
| BAO48208|Csp07DNAV |  | 42 | 48 | 50 | 49 | 55 | 100 |  |  |  |  |  |  |  |
| BAE79193|CsalDNAV |  | 40 | 49 | 50 | 49 | 51 | 57 | 100 |  |  |  |  |  |  |
| BAJ79016|ClorDNAV |  | 39 | 51 | 52 | 50 | 54 | 55 | 54 | 100 |  |  |  |  |  |
| AGS36182|LDMD-2013 |  | 43 | 50 | 50 | 49 | 50 | 51 | 53 | 52 | 100 |  |  |  |  |
| AQA27289|AcrBV1 |  | 41 | 51 | 51 | 52 | 47 | 45 | 47 | 44 | 53 | 100 |  |  |  |
| AQA27293|AcrBV2 |  | 40 | 51 | 49 | 49 | 46 | 46 | 46 | 46 | 55 | 57 | 100 |  |  |
| AQA27298|AHEaBV |  | 35 | 41 | 40 | 41 | 39 | 37 | 42 | 38 | 42 | 40 | 41 | 100 |  |
| BAN59850|TnitDNAV |  | 37 | 38 | 38 | 36 | 36 | 39 | 39 | 37 | 40 | 38 | 38 | 38 | 100 |

\* - Full virus names are provided in the legend to Figure 3.

**Proposed taxonomy**

Due to the unique features of ssDNA viruses infecting diatoms, we propose to create a family *Bacilladnaviridae* for their classification.

Origin of the family name: the name is derived by adding a family ending -*viridae* to the originally proposed name of the *Bacilladnavirus* genus (*Bacillariophytes* infecting ssDNA virus).

Based on the phylogenetic analysis of the bacilladnavirus Rep sequences, we propose to rename the existing genus *Bacilladnavirus* to *Protobacilladnavirus* and to establish 2 new genera, *Diatodnavirus* and *Kieseladnavirus*.

Genus ***Protobacilladnavirus***will include the following species:

* *Chaetoceros protobacilladnavirus 1* (Chaetoceros salsugineum DNA virus 1),
* *Chaetoceros protobacilladnavirus 2* (Chaetoceros sp. DNA virus 7),
* *Chaetoceros protobacilladnavirus 3* (Chaetoceros lorenzianus DNA virus),
* *Chaetoceros protobacilladnavirus 4* (Chaetoceros tenuissimus DNA viruses type I, type IIa, and IIb),
* *Snail associated protobacilladnavirus 1* (Amphibola crenata associated bacilladnavirus 1)[[1]](#footnote-1),
* *Snail associated protobacilladnavirus 2* (Amphibola crenata associated bacilladnavirus 2)1,
* *Marine protobacilladnavirus 1* (Bacillariodnavirus LDMD-2013).

All of the proposed species (n=7; 9 isolates) within the genus *Protobacilladnavirus* share between 44% and 55% sequence identity with Rep sequences of other isolates within the same genus. Isolates within the 7 species cluster with 100% branch support within phylogenetic tree constructed from Rep sequences (Figure 3).

Based on comparison of pairwise identities of protobacilladnavirus Rep amino acid sequences (Figure 3), we propose a 75% pairwise identity as a species cut-off. That is, isolates whose Rep sequences share < 75% pairwise identity should be considered as putative new species.

Genus ***Diatodnavirus*** will include a single species:

* *Chaetoceros diatodnavirus 1* (Chaetoceros setoensis DNA virus).

Origin of the genus name: Diatom-infecting ssDNA virus = *Diatodnavirus*.

Genus ***Kieseladnavirus*** will include a single species:

* *Avon-Heathcote Estuary associated kieseladnavirus 1* (Avon-Heathcote Estuary associated bacilladnavirus)1.

Origin of the genus name: Kieselalge (*diatom* in German)-infecting ssDNA virus = *Kieseladnavirus*.

**References**

Gorbalenya, A.E., Koonin, E.V., Wolf, Y.I. (1990) A new superfamily of putative NTPbinding domains encoded by genomes of small DNA and RNA viruses. FEBS Lett. 262, 145–148.

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 archaebacteria. Nucleic acids research *20*, 3279-3285.

Kazlauskas, D., Dayaram, A., Kraberger, S., Goldstien, S., Varsani, A., and Krupovic, M. (2017). Evolutionary history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA nodaviruses. Virology *504*, 114-121.

Kimura, K., and Tomaru, Y. (2013). Isolation and characterization of a single-stranded DNA virus infecting the marine diatom Chaetoceros sp. strain SS628-11 isolated from western Japan. PloS one *8*, e82013.

Kimura, K., and Tomaru, Y. (2015). 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 *81*, 1120-1131.

Krupovic, M. (2013). Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Curr Opin Virol *3*, 578-586.

McDaniel, L.D., Rosario, K., Breitbart, M., and Paul, J.H. (2014). Comparative metagenomics: natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences. Environ Microbiol *16*, 570-585.

Nagasaki, K., Tomaru, Y., Takao, Y., Nishida, K., Shirai, Y., Suzuki, H., and Nagumo, T. (2005). Previously unknown virus infects marine diatom. Appl Environ Microbiol *71*, 3528-3535.

Rosario, K., Duffy, S., and Breitbart, M. (2012). A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch Virol *157*, 1851-1871.

Tomaru, Y., Shirai, Y., Suzuki, H., Nagumo, T., and Nagasaki, K. (2008). Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom *Chaetoceros debilis*. Aquat Microb Ecol *50*, 103-112.

Tomaru, Y., Shirai, Y., Toyoda, K., and Nagasaki, K. (2011a). Isolation and characterisation of a single-stranded DNA virus infecting the marine planktonic diatom Chaetoceros tenuissimus. Aquat Microb Ecol *64*, 175-184.

Tomaru, Y., Takao, Y., Suzuki, H., Nagumo, T., Koike, K., and Nagasaki, K. (2011b). Isolation and characterization of a single-stranded DNA virus infecting Chaetoceros lorenzianus Grunow. Appl Environ Microbiol *77*, 5285-5293.

Tomaru, Y., Toyoda, K., Kimura, K., Hata, N., Yoshida, M., and Nagasaki, K. (2012). First evidence for the existence of pennate diatom viruses. The ISME journal *6*, 1445-1448.

Tomaru, Y., Toyoda, K., Suzuki, H., Nagumo, T., Kimura, K., and Takao, Y. (2013). New single-stranded DNA virus with a unique genomic structure that infects marine diatom Chaetoceros setoensis. Sci Rep *3*, 3337.

Toyoda, K., Kimura, K., Hata, N., Nakayama, N., Nagasaki, K., and Tomaru, Y. (2012). Isolation and characterization of a single-stranded DNA virus infecting the marine planktonic diatom *Chaetoceros* sp. (strain TG07-C28). Plankton Benthos Res *7*, 20-28.

1. Although initially detected by high-throughput sequencing the genome was subsequently PCR amplified from the original sample, cloned and sequenced using Sanger method to ensure high quality of the genomic data (Kazlauskas et al., 2017). [↑](#footnote-ref-1)