viruses

MDPI

Communication

How to name and classify your phage: an informal guide

Evelien M. Adriaenssens [1,2,*] and J. Rodney Brister [3]

[1] Microbiology Research Group, Institute of Integrative Biology, University of Liverpool, UK; evelien.adriaenssens@liv.ac.uk
[2] Vice Chair of the Bacterial and Archaeal Virus Subcommittee of the International Committee on the Taxonomy of Viruses; evelien.adriaenssens@gmail.com
[3] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892, USA; jamesbr@ncbi.nlm.nih.gov
* Correspondence: evelien.adriaenssens@gmail.com; Tel.: +44-151-795-4576

Academic Editor: name
Received: date; Accepted: date; Published: date

Abstract: With this informal guide, we try to assist both new and experienced phage researchers through two important stages that follow phage discovery, i.e. naming and classification. Providing an appropriate name for a bacteriophage is not as trivial as it sounds and the effects might be long-lasting in databases and in official taxon names. Phage classification is the responsibility of the Bacterial and Archaeal Viruses Subcommittee (BAVS) of the International Committee on the Taxonomy of Viruses (ICTV). While the BAVS aims at providing a holistic approach to phage taxonomy, for individual researchers who have isolated and sequenced a new phage, this can be a little overwhelming. We are now providing these researchers with an informal guide to phage naming and classification, taking a "bottom-up" approach from the phage isolate level.

Keywords: Bacteriophages; phage taxonomy; phage classification; naming guide; classification guide

1. Introduction

Virus taxonomy is currently the responsibility of the International Committee on the Taxonomy of Viruses (ICTV, [1]), which published its first report in 1971. The Bacterial and Archaeal Viruses Subcommittee (BAVS) within ICTV holds the responsibility of classifying new prokaryotic viruses. New proposals can be submitted year round by the public. These Taxonomy Proposals (TaxoProps) are evaluated by relevant Study Groups (SGs) and the BAVS [2], and are then discussed and voted on by the Executive Committee (EC) during the yearly meeting. All ICTV-accepted proposals are finally ratified by the members of the IUMS (International Union of Microbiological Societies) Virology Division through an email vote.

Bacterial virus taxonomy has undergone a number of changes since the discovery of bacteriophages in the early 20th century. Electron microscopy, which lead to the recognition of different phage morphologies, and nucleic acid content provided the basis for the first classification scheme [3,4]. Ever since, genome composition and morphology have been the major criterion for classification at the family rank, with the current taxonomy comprising 22 families grouping bacterial or archaeal viruses.

For many years, the grouping of prokaryotic viruses in lower rank taxa such as genus and subfamily, happened at a minimal pace. Taking the tailed phage families as an example, the 5th Report of ICTV (1991) described one genus in each of the families *Myoviridae*, *Podoviridae* and *Siphoviridae* [5]. This increased to 16 genera spread over the three families by the 7th Report [6] and 18 genera by the 8th Report [7]. As nucleotide sequencing techniques improved the number of publically available

bacteriophage sequences, researchers started to question the large number of bacteriophage genomes that remained unclassified [8]. These concerns would prove prescient, and in the coming years next generation sequencing methods would spur an explosion in bacteriophage sequencing (Figure 1).

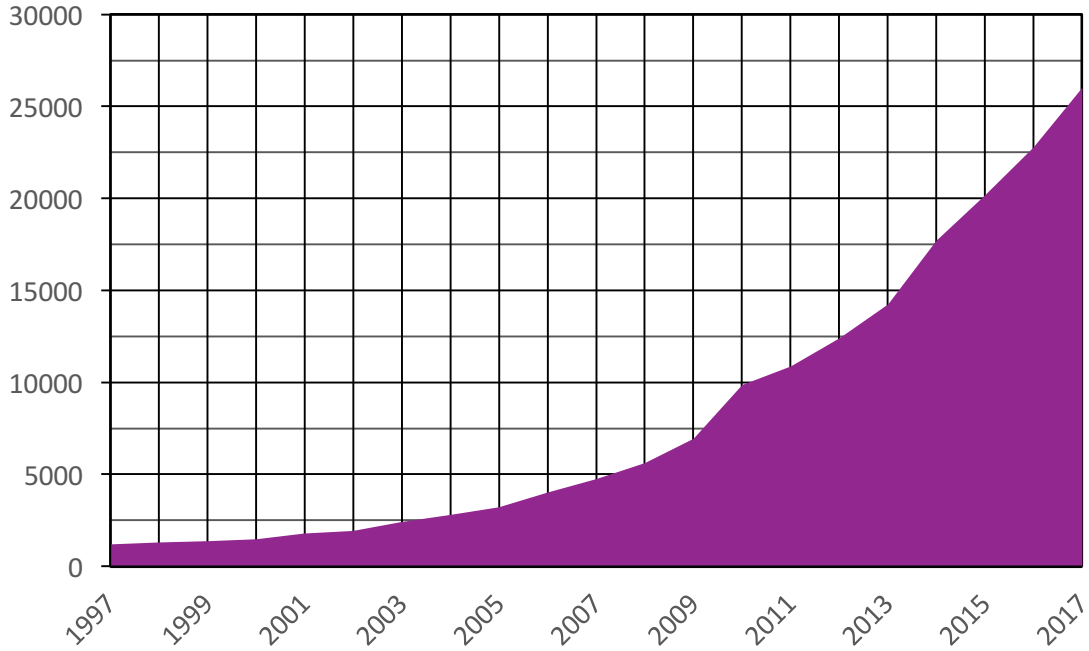

Figure 1. Number of bacteriophage nucleotide sequences deposited in INSDC databases. The number of nucleotide sequences publically available from INSDC databases was calculated by searching the GenBank nucleotide database with the term "vhost bacteria[filter]" and plotting the number of sequences available on January 1 of each year shown [9].

The availability of genome sequence data also gave rise to a range of potential classification or grouping schemes, such as the Phage Proteomic Tree [10], phage network clusters [11], kmer-based grouping [12], signature genes-based grouping [13] or whole genome nucleotide identity-based grouping [14], which were not always compatible with the rules laid out in the ICTV Code and/or the International Code of Virus Classification and Nomenclature (ICVCN). Since the 8th Report of ICTV, both genome and proteome-based methods have been used by the BAVS to classify phages into species, genera and subfamilies, resulting in 14 subfamilies, 204 genera and 873 species in the 2015 taxonomy release [15–20].

In this paper, we provide a naming and classification guide for researchers who have isolated and sequenced a novel bacteriophage isolate specifically, however, these guidelines can be applied to archaeal viruses as well. The guide will follow a "bottom-up" approach, i.e. starting at the species level, rather than the "top-down" approach which was used in the past to assign isolates to a family based on morphology.

2. A short, informal guide to naming and classifying your phage

2.1. Congratulations, you have isolated a bacteriophage!

You have just isolated and sequenced a bacteriophage, so what do you do next? Well, hopefully, you plan on publishing your finding and submitting the sequence to one of the public databases that are part of the International Nucleotide Sequence Database Consortium (INSDC), GenBank, the European Nucleotide Archive (ENA) or the DNA Data Bank of Japan (DDBJ) [21]. That's great, but now the big question is: "What do I name my phage?" Didn't think about that one, did you? Well, turns out the name you give to your new virus isolate is pretty important, as it will be used in

publications, mentioned in conversations among colleagues, and identify the sequence of your virus in public databases and other resources.

Perhaps the most important rule of bacteriophage naming is "don't use an existing name." There are already four dissimilar bacteriophages named N4, making it very difficult to distinguish between them. So before naming your bacteriophage – and definitely before publishing a report on it – please take the time to compare proposed names against those already used within the field. A good, if not dated, place to start is Bacteriophage Names 2000 [22]. A more up to date list of bacteriophage names can be found by searching the NCBI Nucleotide database [23] with the term "vhost bacteria[filter] AND ddbj_embl_genbank[filter]" [9]. This search will return all bacteriophage isolate names currently associated with sequences in INSDC databases – both those classified by ICTV, as well as those that have yet to undergo official classification.

The current approach to bacteriophage naming is a tripartite construct consisting of the bacterial host genus name, the word "phage," and a unique identifier, for example "Escherichia phage T4." Since the first two components of this naming construct are not unique, the third component is critical to the usability of the name. Leafing through a list of bacteriophage names, it is clear that there are a number of approaches to providing unique identifiers in names. For example, one approach to constructing unique identifiers includes information about phage morphology and host [24]. So the name Escherichia phage vB_EcoM-VR20 denotes a **v**irus of **B**acteria, infecting ***E****scherichia* ***co****li*, with **m**yovirus morphology. One caveat to this approach is that one needs to employ electron microscopy or computational methods to derive the correct morphotype. While there are few hard and fast rules for these terms, please be careful when choosing one, because it is likely to be used as shorthand in a variety of contexts for years to come.

Please use the following bacteriophage naming guidelines:

- Always use the complete host genus name, followed by a space, followed by the word "phage," followed by a space, followed by a unique identifier, e.g. Escherichia phage T4.
- Use only the isolation host genus in the name, rather than higher order taxa names - such as Enterobacteria, Pseudomonad, or the generic Bacteriophage – or lower order taxa names like *Staphylococcus aureus* DSM 1234.
- Do not combine the host genus and the word "phage" into a single word, for example, Mycobacteriophage, Mycophage, etc.
- Do not use an existing unique identifier in the name.
- Do not use Greek letters in the unique identifier.
- Do not start the unique identifier with a numeral and do not use only use only a single letter. Identifiers should include enough complexity to easily distinguish your bacteriophage form others.
- Do not use hyphens, slashes or any type of special character like %$@ etc. You may use underscores to separate parts of the designation, for example vs_p123_233, but these underscores cannot be carried over into official taxon names (see paragraph 2.4).
- Do not use controversial names/phases, profanity, names of prominent people, and trademarked names/phrases as unique identifiers.
- Please do contact the friendly folks on the BAVS if you have any questions.

2.2 What is the relationship between bacteriophage isolate names and taxa names?

The rules for naming taxa are described in the International Code of Virus Classification and Nomenclature [25]. Typically, the name of a species is based on the name of the first sequenced isolate, which then becomes the type isolate. Current bacteriophage species names replace "phage" in the tripartite isolate name construct with "virus," so the isolate Escherichia phage T4 belongs to the species *Escherichia virus T4*. Higher order taxa names are derived from unique identifiers used in isolate names as in the genus *T4virus*. Sometimes these unique identifiers are too similar to existing

taxon names, inappropriate, or do not otherwise conform to ICTV taxa naming standards, and a different taxon name must be chosen.

2.3 Now it is time to publish your phage sequence.

Once you have isolated, sequenced, and named your new bacteriophage, it is time to start thinking about sharing your data with the world. Today, sharing your results is not simply about publishing in a peer reviewed journal. While such descriptions are central to the scientific process, so too is the sequence of your new bacteriophage. Though often overlooked, submitting your new bacteriophage sequence to a public INSDC database such as GenBank, is critical to making your sequence publically available for generations to come. Please keep in mind that in this age of bioinformatics and computational biology, it is likely that over time the sequence record for your new bacteriophage will be accessed exponentially more often than a traditional publication. In other words, do your best to provide detailed and accurate information about your phage when you submit the sequence to an INSDC database. This includes providing the most accurate classification data possible.

If known, lineage information should be included in INSDC sequence submissions using the "lineage" field or in a source note. For example, if you have sequenced a new phage that belongs to the species *Escherichia virus T4*, provide the name of your new virus, e.g. "Escherichia phage My_New_Virus," and the lineage "Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae; Tevenvirinae; T4virus; Escherichia virus T4". In cases where your new phage cannot be placed in an existing species, provide a lineage that reflects classification. For example, if your new phage belongs to the genus *T4virus*, provide the lineage "Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae; Tevenvirinae; T4virus".

Please use the following guidelines when submitting to public databases:

- Do include lineage information for all submitted sequences. Even if your bacteriophage is novel and does not belong to a described species, provide the most accurate lineage information possible that places the sequence including genus and/or family using the criteria discussed in this manuscript.
- Do include accurate genomic composition information when no other lineage information is available or can be inferred. In most cases it should be possible to place a new isolate within the higher order dsDNA, ssDNA, dsRNA, or ssRNA lineage groupings.
- Do identify prophages using the "proviral" location descriptor.
- If you have questions about sequence submission to INSDC databases, please see The GenBank Submissions Handbook [26].
- If you still have questions, contact GenBank or another INSDC database.

2.4 Classifying bacteriophage.

So why does taxonomy even matter? Well, taxonomy offers a very useful way of aggregating genome sequence data around a collection of genetic and/or molecular attributes. In this way, rules describing taxa are effectively search terms that allow you to retrieve a set of sequences with similar characteristics. Taxonomy also provides context to sequences when searching for sequence similarities. Knowing that a newly sequenced virus is highly similar to a previously classified one, immediately tells you something about the new virus – the expected gene content, host range, environmental niche, etc.

Bacteriophage classification also supports the organization of genome sequence data within public databases. Each viral species is represented in by at least one "reference" genome in the NCBI Viral Refseq database. Other validated genomes belonging to the same species will be stored as so

called "genome neighbors" of the RefSeq genome [9,27]. This arrangement provides a compressed dataset where each species is represented by one (or more) representative sequences – typically from type isolates - as well as a method for retrieving a set of validated genomes from each viral species.

2.4.1 Does my phage represent a new species?

The first question you need to answer is basic one: "Does my newly sequenced phage belong to an existing species?" The main species demarcation criterion for bacterial and archaeal viruses is currently set at a genome sequence identity of 95%, meaning that two viruses belonging to the same species differ from each other by less than 5% at the nucleotide level. This can be calculated by comparing your sequence to existing phage genomes. There are several tools to do this (e.g. BLASTN [28], PASC [29], Gegenees [30] or EMBOSS Stretcher), but each comparison needs to be checked for genomic synteny. While it is common for larger dsDNA phages to differ in their genome organization, isolates showing high levels of rearrangements can never belong to the same species.

If your phage belongs to an existing species, be sure to specify that taxonomic lineage when depositing the sequence into GenBank or other INSDC databases. If results suggest that your phage represents a novel species, congratulations! The appropriate Study Groups and BAVS will, with your help in providing data, create a new species based on your phage. To place this new species in a higher taxon, we will move up to the genus level in the next section.

We recommend that you alert the appropriate BAVS Study Group chair or the Subcommittee Chair [2] who will advise you on how to proceed. This will generally involve filling out an ICTV Taxonomy Proposal Submission Template (TaxoProp for short) which is available here from the ICTV website [31]. The ICTV website includes examples of completed TaxoProps.

2.4.2 Is my phage a member of an existing genus?

The BAVS currently describes a genus as a cohesive group of viruses sharing a high degree of nucleotide sequence similarity (> 50%), which is distinct from viruses of other genera. For each genus, defining characteristics can be determined, such as average genome length, average number of coding sequences, percentage of shared coding sequences, average number of tRNAs, and the presence of certain signature genes in member viruses. The latter can in turn be used for phylogenetic analysis with other bacterial or archaeal viruses encoding this gene to find monophyletic groups as well as higher order relationships.

All the genera currently in the ICTV database have a taxonomy history (TaxoProp) accessible through the website, which can be used for researchers to assess the genus inclusion criteria. If your phage falls into an existing genus, the BAVS will define the new viral species within the existing genus. If the phage is sufficiently different from existing isolates, we can define a new genus, according to the characteristics described above. The minimum requirements for the creation of a new genus are the description of the average genome characteristics of its proposed members (size, GC content, tRNAs, coding sequences), a nucleotide identity comparison with visualization, a comparison of the predicted proteomes and phylogenetic analysis of at least one conserved gene, all of these with the appropriate outliers to demonstrate cohesiveness of the new genus.

While you can propose a new genus and species based upon a unique virus, the BAVS generally recommends that genera be established when two or more related viruses have been deposited in one of the INSDC databases.

2.4.3 What about subfamilies and families?

In the current taxonomy releases, bacterial and archaeal viruses are classified at the family rank according to the morphology of their virions, e.g., phages with short tails are placed in the family *Podoviridae*. This means that for proper classification, electron micrographs of the viral particles

should be taken. Based on the morphology and the genomic information necessary for classification in species and genus, we can now look whether your isolate falls in an existing subfamily of viruses. If your new phage, in its newly created genus, is genomically or proteomically similar to phages in an existing subfamily, the genus can be added to the subfamily. The criteria for inclusion can vary between subfamilies and should be consulted from the TaxoProps describing the respective subfamily.

At this time, subfamilies are only created when they add necessary hierarchical information (ICVCN Rule 3.2). In practical terms, this mean that a new subfamily is created when two or more genera show an obvious relation which is not adequately described at the family level. For instance, in the family *Podoviridae*, the subfamily *Autographivirinae* groups all podoviruses that contain an RNA polymerase gene in their genome [16]. The requirements for the creation of a new subfamily are not easily defined, but should definitely include the description of at least two clearly related genera within a family, with evidence that the new subfamily is cohesive.

In a genome-based taxonomy, the tailed phage families *Myoviridae*, *Siphoviridae* and *Podoviridae*, have become an artificial "ceiling" prohibiting the accurate description and grouping of the genomic diversity present among their member phages. For example, T4-related phages, infecting a wide range of host bacteria from different phyla, are characterized by the presence of a set of 30 conserved (core) proteins [32], but also have more distant cousins in the Far-T4 group sharing only 10 core proteins [33]. These phages are currently all classified within the family *Myoviridae*, but have the genomic diversity to represent a new order. Another example involves the "lambdoid phages", comprising both siphoviruses (Escherichia phage Lambda) and podoviruses (Salmonella phage P22), which cannot be grouped together in the same family at this time. The BAVS is therefore working on a new system that would abolish these families in favor of genome/proteome-based family descriptions.

2.4.4 My phage/virus does not fit in anywhere, what now?

In the very special circumstance that your new phage does not fit in with any known bacterial or archaeal virus, genomically or morphologically, it is the first representative of a new family. In this case, we strongly urge you to contact the BAVS Chair, or the chair of an appropriate Study Group, to work together to define the demarcation criteria for this new family.

2.5 Proposed software to use

This is a non-exhaustive list with suggested software to use. The BAVS as a subcommittee is not associated with the developers of the software described below.

- Nucleotide sequence comparison: NCBI BLASTn [28], Gegenees (uses BLASTn) [30], PASC [29], Gepard dotplot [34].
- Comparison of protein groups, predicted proteomes, identification of signature genes: CoreGenes 3.5 [35,36], Roary (core and accessory genome analysis) [37], prokaryotic Virus Orthologous Groups resource (pVOGs) [38].
- Multiple alignment and/or phylogenetic analyses: Clustal Omega [39], MUSCLE [40], phylogeny.fr [41], MEGA6 [42], FastTree [43].
- Visualization: progressiveMAUVE [44], Easyfig [45], BRIG [46].

3. Conclusion

Bacteriophage genomics, ecology, and evolution are quickly growing fields, and large numbers of new phages are being discovered, named, sequenced, and deposited into public databases. This poses semantic, logistical and taxonomical challenges that we have tried to address in this informal

guide. It is also important to understand that taxonomy is ever changing because of the unremitting flow of new information. The effort of classification is currently undertaken by a small group of dedicated scientists, but with input from the larger phage community – this means you, dear reader – we can increase the effort while keeping it manageable for each individual researcher.

Supplementary Materials: This article has no supplementary materials associated with it.

Acknowledgments: We would like to thank Andrew M. Kropinski for his invaluable support and comments on this manuscript, Linda Frisse and Detlef Leipe for discussions about sequence submission guidelines, and Jens H. Kuhn for applying his attention to detail and keen eye to these writings.

Author Contributions: E.M.A. conceived the study, E.M.A. and J.R.B. wrote the paper.

Conflicts of Interest: E.M.A. and J.R.B. are both members of the BAVS of ICTV. E.M.A. was funded by the National Environmental Research Council of the UK. Research by J.R.B. was supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine. The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. International Committee on Taxonomy of Viruses (ICTV). Available online: http://www.ictvonline.org/ (accessed Feb 23, 2017).
2. Bacterial and Archaeal Viruses Subcommittee. Available online: https://talk.ictvonline.org/information/w/members/441/bacterial-and-archaeal-viruses-subcommittee (accessed Feb 23, 2017).
3. Bradley, D. E. Ultrastructure of bacteriophage and bacteriocins. *Bacteriol. Rev.* **1967**, 31, 230–314. Available online: http://mmbr.asm.org/cgi/content/short/31/4/230 (accessed on Feb 23, 2017).
4. Ackermann, H.-W. Classification of bacteriophages. In *The Bacteriophages*, 2nd ed.; Calendar, R., Ed.; Oxford University Press: New York, NY, USA, 2006; pp. 8–17.
5. Francki, R.; Fauquet, C.; Knudson, D.; Brown, F. *Classification and nomenclature of viruses: Fifth report of the International Committee on the Taxonomy of Viruses*; Springer: Wien, NY, USA, 1991.
6. Fauquet, M. C.; Mayo, A. M. The 7th ICTV Report. *Arch. Virol.* **2001**, 146, 189–194, DOI 10.1007/s007050170203
7. Fauquet, C.; Mayo, M.; Maniloff, J.; Desselberger, U.; Ball, L. *Virus taxonomy: Eighth report of the International Committee on Taxonomy of Viruses*; Elsevier Academic Press: San Diego, CA, USA, 2005.
8. Nelson, D. Phage taxonomy: we agree to disagree. *J. Bacteriol.* **2004**, 186, 7029–7031, DOI 10.1128/JB.186.21.7029-7031.2004.
9. Brister, J. R.; Ako-Adjei, D.; Bao, Y.; Blinkova, O. NCBI viral genomes resource. *Nucleic Acids Res.* **2015**, 43, D571–D577, DOI 10.1093/nar/gku1207.
10. Rohwer, F.; Edwards, R. The Phage Proteomic Tree: a genome-based taxonomy for phage. *J. Bacteriol.* **2002**, 184, 4529–4535, DOI 10.1128/JB.184.16.4529.
11. Lima-Mendez, G.; Van Helden, J.; Toussaint, A.; Leplae, R. Reticulate representation of evolutionary and functional relationships between phage genomes. *Mol. Biol. Evol.* **2008**, 25, 762–777, DOI 0.1093/molbev/msn023.
12. Pride, D. T.; Wassenaar, T. M.; Ghose, C.; Blaser, M. J. Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. *BMC Genomics* **2006**, 7, 8, DOI 10.1186/1471-2164-7-8.
13. Asare, P. T.; Jeong, T.; Ryu, S.; Klumpp, J.; Loessner, M. J.; Merrill, B. D.; Kim, K. Putative type 1 thymidylate synthase and dihydrofolate reductase as signature genes of a novel bastille-like group of phages in the subfamily Spounavirinae. *BMC Genomics* **2015**, 16, 582, DOI 10.1186/s12864-015-1757-0.
14. Grose, J. H.; Casjens, S. R. Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family *Enterobacteriaceae*. *Virology* **2014**, 468–470, 421–443, DOI 10.1016/j.virol.2014.08.024.
15. Adriaenssens, E. M.; Edwards, R.; Nash, J. H. E.; Mahadevan, P.; Seto, D.; Ackermann, H.-W.; Lavigne, R.; Kropinski, A. M. Integration of genomic and proteomic analyses in the classification of the *Siphoviridae* family. *Virology* **2015**, 477, 144–154, DOI 10.1016/j.virol.2014.10.016.

16. Lavigne, R.; Seto, D.; Mahadevan, P.; Ackermann, H.-W.; Kropinski, A. M. Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. *Res. Microbiol.* **2008**, 159, 406–414, DOI 10.1016/j.resmic.2008.03.005.
17. Lavigne, R.; Darius, P.; Summer, E. J.; Seto, D.; Mahadevan, P.; Nilsson, A. S.; Ackermann, H.-W.; Kropinski, A. M. Classification of *Myoviridae* bacteriophages using protein sequence similarity. *BMC Microbiol.* **2009**, 9, 224, DOI 10.1186/1471-2180-9-224.
18. Krupovic, M.; Dutilh, B. E.; Adriaenssens, E. M.; Wittmann, J.; Vogensen, F. K.; Sullivan, M. B.; Rumnieks, J.; Prangishvili, D.; Lavigne, R.; Kropinski, A. M.; Klumpp, J.; Gillis, A.; Enault, F.; Edwards, R. A.; Duffy, S.; Clokie, M. R. J.; Barylski, J.; Ackermann, H.-W.; Kuhn, J. H. Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. *Arch. Virol.* **2016**, 4, 1095-1099, DOI 10.1007/s00705-015-2728-0.
19. Adriaenssens, E. M.; Krupovic, M.; Knezevic, P.; Ackermann, H.-W.; Barylski, J.; Brister, J. R.; Clokie, M. R. C.; Duffy, S.; Dutilh, B. E.; Edwards, R. A.; Enault, F.; Jang, H. Bin; Klumpp, J.; Kropinski, A. M.; Lavigne, R.; Poranen, M. M.; Prangishvili, D.; Rumnieks, J.; Sullivan, M. B.; Wittmann, J.; Oksanen, H. M.; Gillis, A.; Kuhn, J. H. Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses subcommittee. *Arch. Virol.* **2017**, in press, DOI 10.1007/s00705-016-3173-4.
20. Virus Taxonomy: 2015 Release. Available online: http://talk.ictvonline.org/taxonomy/ (accessed Feb 23, 2017).
21. Karsch-Mizrachi, I.; Nakamura, Y.; Cochrane, G. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.* **2012**, 40, D33–D37, DOI 10.1093/nar/gkr1006.
22. Ackermann, H.-W.; Abedon, S. T. Bacteriophage names 2000. Available online: http://www.phage.org/names/2000/ (accessed Feb 23, 2017).
23. NCBI Nucleotide Database. Available online: http://www.ncbi.nlm.nih.gov/nuccore (accessed Feb 24, 2017).
24. Kropinski, A. M.; Prangishvili, D.; Lavigne, R. Position paper: the creation of a rational scheme for the nomenclature of viruses of Bacteria and Archaea. *Environ. Microbiol.* **2009**, 11, 2775–2777, DOI 10.1111/j.1462-2920.2009.01970.x.
25. ICTV Code. Available online: http://talk.ictvonline.org/information/w/ictv-information/383/ictv-code (accessed Feb 23, 2017).
26. National Center for Biotechnology Information (US), The GenBank submissions handbook [Internet]. Available online: https://www.ncbi.nlm.nih.gov/books/NBK51157/ (accessed Feb 23, 2017).
27. O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C. M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina, W.; Joardar, V. S.; Kodali, V. K.; Li, W.; Maglott, D.; Masterson, P.; McGarvey, K. M.; Murphy, M. R.; O'Neill, K.; Pujar, S.; Rangwala, S. H.; Rausch, D.; Riddick, L. D.; Schoch, C.; Shkeda, A.; Storz, S. S.; Sun, H.; Thibaud-Nissen, F.; Tolstoy, I.; Tully, R. E.; Vatsan, A. R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M. J.; Kimchi, A.; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T. D.; Pruitt, K. D. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, 44, D733–D745, DOI 10.1093/nar/gkv1189.
28. Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; Madden, T. L. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **2008**, 36, 5–9, DOI 10.1093/nar/gkn201.
29. Bao, Y.; Chetvernin, V.; Tatusova, T. Improvements to pairwise sequence comparison (PASC): a genome-based web tool for virus classification. *Arch. Virol.* **2014**, 159, 3293–3304, DOI 10.1007/s00705-014-2197-x.
30. Ågren, J.; Sundström, A.; Håfström, T.; Segerman, B. Gegenees: Fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One* **2012**, 7, e39107, DOI 10.1371/journal.pone.0039107.
31. ICTV Taxonomy Proposals templates. Available online: https://talk.ictvonline.org/files/taxonomy-proposal-templates/ (accessed Feb 23, 2017).
32. Petrov, V. M.; Ratnayaka, S.; Nolan, J. M.; Miller, E. S.; Karam, J. D. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. *Virol. J.* **2010**, 7, 292, DOI 10.1186/1743-422X-7-292.

33. Roux, S.; Enault, F.; Ravet, V.; Pereira, O.; Sullivan, M. B. Genomic characteristics and environmental distributions of the uncultivated Far-T4 phages. *Front. Microbiol.* **2015**, 6, 1–13, DOI 10.3389/fmicb.2015.00199.
34. Krumsiek, J.; Arnold, R.; Rattei, T. Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **2007**, 23, 1026–1028, DOI 10.1093/bioinformatics/btm039.
35. Zafar, N.; Mazumder, R.; Seto, D. CoreGenes: A computational tool for identifying and cataloging "core" genes in a set of small genomes. *BMC Bioinformatics* **2002**, 3, 12, DOI 10.1186/1471-2105-3-12.
36. Mahadevan, P.; King, J. F.; Seto, D. CGUG: in silico proteome and genome parsing tool for the determination of "core" and unique genes in the analysis of genomes up to ca. 1.9 Mb. *BMC Res. Notes* **2009**, 2, 168, DOI 10.1186/1756-0500-2-168.
37. Page, A. J.; Cummins, C. A.; Hunt, M.; Wong, V. K.; Reuter, S.; Holden, M. T. G.; Fookes, M.; Falush, D.; Keane, J. A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, 31, 3691–3693, DOI 10.1093/bioinformatics/btv421.
38. Grazziotin, A. L.; Koonin, E. V; Kristensen, D. M. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res.* **2016**, 45, D491-D498, DOI 10.1093/nar/gkw975.
39. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, 7, 539, DOI 10.1038/msb.2011.75.
40. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, 32, 1792–1797, DOI 10.1093/nar/gkh340.
41. Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J.-F.; Guindon, S.; Lefort, V.; Lescot, M.; Claverie, J.-M.; Gascuel, O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **2008**, 36, W465-469, DOI 10.1093/nar/gkn180.
42. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, 30, 2725–2729, DOI 10.1093/molbev/mst197.
43. Price, M. N.; Dehal, P. S.; Arkin, A. P. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* **2010**, 5, e9490, 10.1371/journal.pone.0009490.
44. Darling, A. E.; Mau, B.; Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **2010**, 5, e11147, 10.1371/journal.pone.0011147.
45. Sullivan, M. J.; Petty, N. K.; Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **2011**, 27, 1009–1010, DOI 10.1093/bioinformatics/btr039.
46. Alikhan, N.-F.; Petty, N. K.; Ben Zakour, N. L.; Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **2011**, 12, 402, DOI 10.1186/1471-2164-12-402.