

1 *Communication*

2 **How to name and classify your phage: an informal** 3 **guide**

4 **Evelien M. Adriaenssens** ^{1,2,*} and **J. Rodney Brister** ³

5 ¹ Microbiology Research Group, Institute of Integrative Biology, University of Liverpool, UK;
6 evelien.adriaenssens@liv.ac.uk

7 ² Vice Chair of the Bacterial and Archaeal Virus Subcommittee of the International Committee on the
8 Taxonomy of Viruses; evelien.adriaenssens@gmail.com

9 ³ National Center for Biotechnology Information, National Library of Medicine, National Institutes of
10 Health, Bethesda, MD 20892, USA; jamesbr@ncbi.nlm.nih.gov

11 * Correspondence: evelien.adriaenssens@gmail.com; Tel.: +44-151-795-4576

12 Academic Editor: name

13 Received: date; Accepted: date; Published: date

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

23 **Keywords:** Bacteriophages; phage taxonomy; phage classification; naming guide; classification
24 guide
25

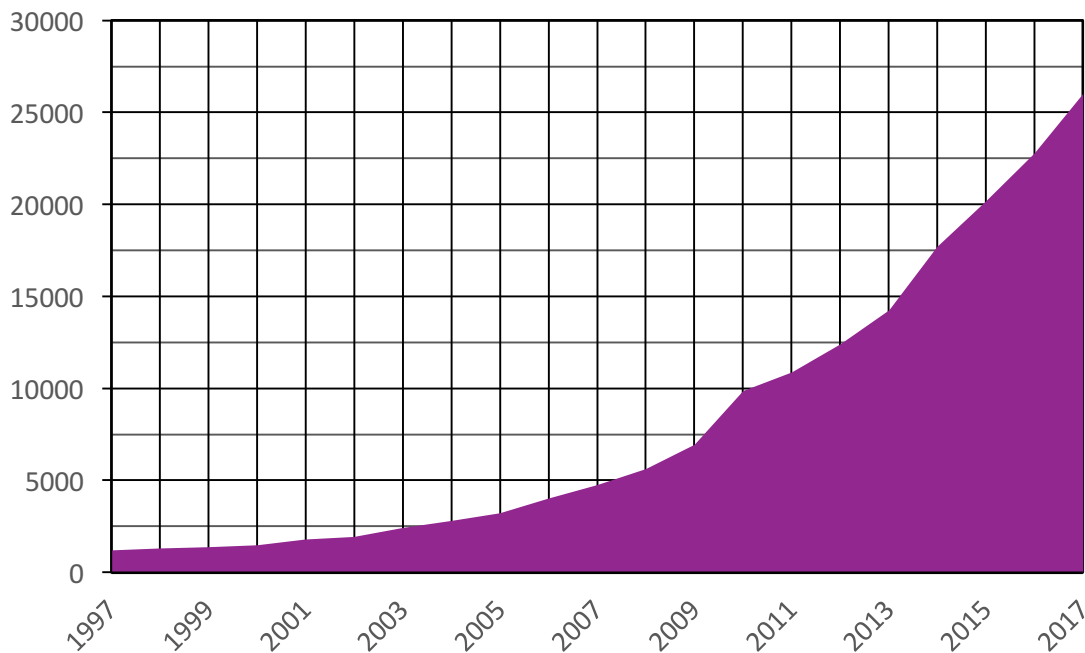
26 **1. Introduction**

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

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

41 For many years, the grouping of prokaryotic viruses in lower rank taxa such as genus and
42 subfamily, happened at a minimal pace. Taking the tailed phage families as an example, the 5th Report
43 of ICTV (1991) described one genus in each of the families *Myoviridae*, *Podoviridae* and *Siphoviridae* [5].
44 This increased to 16 genera spread over the three families by the 7th Report [6] and 18 genera by the
45 8th Report [7]. As nucleotide sequencing techniques improved the number of publically available

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



49

50 **Figure 1.** Number of bacteriophage nucleotide sequences deposited in INSDC databases. The number
51 of nucleotide sequences publicly available from INSDC databases was calculated by searching the
52 GenBank nucleotide database with the term “vhost bacteria[filter]” and plotting the number of
53 sequences available on January 1 of each year shown [9].

54 The availability of genome sequence data also gave rise to a range of potential classification or
55 grouping schemes, such as the Phage Proteomic Tree [10], phage network clusters [11], kmer-based
56 grouping [12], signature genes-based grouping [13] or whole genome nucleotide identity-based
57 grouping [14], which were not always compatible with the rules laid out in the ICTV Code and/or
58 the International Code of Virus Classification and Nomenclature (ICVCN). Since the 8th Report of
59 ICTV, both genome and proteome-based methods have been used by the BAVS to classify phages
60 into species, genera and subfamilies, resulting in 14 subfamilies, 204 genera and 873 species in the
61 2015 taxonomy release [15–20].

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

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

68 *2.1. Congratulations, you have isolated a bacteriophage!*

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

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

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

86 The current approach to bacteriophage naming is a tripartite construct consisting of the bacterial
87 host genus name, the word “phage,” and a unique identifier, for example “*Escherichia* phage T4.”
88 Since the first two components of this naming construct are not unique, the third component is critical
89 to the usability of the name. Leafing through a list of bacteriophage names, it is clear that there are
90 a number of approaches to providing unique identifiers in names. For example, one approach to
91 constructing unique identifiers includes information about phage morphology and host [24]. So the
92 name *Escherichia* phage vB_EcoM-VR20 denotes a virus of **Bacteria**, infecting *Escherichia coli*, with
93 myovirus morphology. One caveat to this approach is that one needs to employ electron microscopy
94 or computational methods to derive the correct morphotype. While there are few hard and fast rules
95 for these terms, please be careful when choosing one, because it is likely to be used as shorthand in a
96 variety of contexts for years to come.

97

98 Please use the following bacteriophage naming guidelines:

99

- 100 • Always use the complete host genus name, followed by a space, followed by the word
- 101 “phage,” followed by a space, followed by a unique identifier, e.g. *Escherichia* phage T4.
- 102 • Use only the isolation host genus in the name, rather than higher order taxa names - such
- 103 as *Enterobacteria*, *Pseudomonad*, or the generic *Bacteriophage* – or lower order taxa
- 104 names like *Staphylococcus aureus* DSM 1234.
- 105 • Do not combine the host genus and the word “phage” into a single word, for example,
- 106 *Mycobacteriophage*, *Mycophage*, etc.
- 107 • Do not use an existing unique identifier in the name.
- 108 • Do not use Greek letters in the unique identifier.
- 109 • Do not start the unique identifier with a numeral and do not use only use only a single
- 110 letter. Identifiers should include enough complexity to easily distinguish your
- 111 bacteriophage from others.
- 112 • Do not use hyphens, slashes or any type of special character like %\$@ etc. You may use
- 113 underscores to separate parts of the designation, for example vs_p123_233, but these
- 114 underscores cannot be carried over into official taxon names (see paragraph 2.4).
- 115 • Do not use controversial names/phases, profanity, names of prominent people, and
- 116 trademarked names/phrases as unique identifiers.
- 117 • Please do contact the friendly folks on the BAVS if you have any questions.
- 118

119 2.2 What is the relationship between bacteriophage isolate names and taxa names?

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

126 taxon names, inappropriate, or do not otherwise conform to ICTV taxa naming standards, and a
127 different taxon name must be chosen.
128

129 2.3 Now it is time to publish your phage sequence.

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

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

149

150 Please use the following guidelines when submitting to public databases:

151

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

164

165 2.4 Classifying bacteriophage.

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

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

176 called “genome neighbors” of the RefSeq genome [9,27]. This arrangement provides a compressed
177 dataset where each species is represented by one (or more) representative sequences – typically from
178 type isolates - as well as a method for retrieving a set of validated genomes from each viral species.
179

180 2.4.1 Does my phage represent a new species?

181 The first question you need to answer is basic one: “Does my newly sequenced phage belong to
182 an existing species?” The main species demarcation criterion for bacterial and archaeal viruses is
183 currently set at a genome sequence identity of 95%, meaning that two viruses belonging to the same
184 species differ from each other by less than 5% at the nucleotide level. This can be calculated by
185 comparing your sequence to existing phage genomes. There are several tools to do this (e.g. BLASTN
186 [28], PASC [29], Gegenees [30] or EMBOSS Stretcher), but each comparison needs to be checked for
187 genomic synteny. While it is common for larger dsDNA phages to differ in their genome
188 organization, isolates showing high levels of rearrangements can never belong to the same species.

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

194 We recommend that you alert the appropriate BAVS Study Group chair or the Subcommittee
195 Chair [2] who will advise you on how to proceed. This will generally involve filling out an ICTV
196 Taxonomy Proposal Submission Template (TaxoProp for short) which is available here from the ICTV
197 website [31]. The ICTV website includes examples of completed TaxoProps.
198

199 2.4.2 Is my phage a member of an existing genus?

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

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

216 While you can propose a new genus and species based upon a unique virus, the BAVS generally
217 recommends that genera be established when two or more related viruses have been deposited in
218 one of the INSDC databases.
219

220 2.4.3 What about subfamilies and families?

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

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

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

237 In a genome-based taxonomy, the tailed phage families *Myoviridae*, *Siphoviridae* and *Podoviridae*,
238 have become an artificial “ceiling” prohibiting the accurate description and grouping of the genomic
239 diversity present among their member phages. For example, T4-related phages, infecting a wide
240 range of host bacteria from different phyla, are characterized by the presence of a set of 30 conserved
241 (core) proteins [32], but also have more distant cousins in the Far-T4 group sharing only 10 core
242 proteins [33]. These phages are currently all classified within the family *Myoviridae*, but have the
243 genomic diversity to represent a new order. Another example involves the “lambdoid phages”,
244 comprising both siphoviruses (Escherichia phage Lambda) and podoviruses (Salmonella phage P22),
245 which cannot be grouped together in the same family at this time. The BAVS is therefore working on
246 a new system that would abolish these families in favor of genome/proteome-based family
247 descriptions.
248

249 2.4.4 My phage/virus does not fit in anywhere, what now?

250 In the very special circumstance that your new phage does not fit in with any known bacterial
251 or archaeal virus, genomically or morphologically, it is the first representative of a new family. In this
252 case, we strongly urge you to contact the BAVS Chair, or the chair of an appropriate Study Group, to
253 work together to define the demarcation criteria for this new family.
254

255 2.5 Proposed software to use

256 This is a non-exhaustive list with suggested software to use. The BAVS as a subcommittee is not
257 associated with the developers of the software described below.
258

- 259 • Nucleotide sequence comparison: NCBI BLASTn [28], Gegenees (uses BLASTn) [30],
260 PASC [29], Gepard dotplot [34].
- 261 • Comparison of protein groups, predicted proteomes, identification of signature genes:
262 CoreGenes 3.5 [35,36], Roary (core and accessory genome analysis) [37], prokaryotic
263 Virus Orthologous Groups resource (pVOGs) [38].
- 264 • Multiple alignment and/or phylogenetic analyses: Clustal Omega [39], MUSCLE [40],
265 phylogeny.fr [41], MEGA6 [42], FastTree [43].
- 266 • Visualization: progressiveMAUVE [44], Easyfig [45], BRIG [46].
267

268 3. Conclusion

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

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

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

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

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

281 **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
282 National Environmental Research Council of the UK. Research by J.R.B. was supported by the Intramural
283 Research Program of the National Institutes of Health, National Library of Medicine. The authors declare no
284 conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or
285 interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

286 References

- 287 1. International Committee on Taxonomy of Viruses (ICTV). Available online: <http://www.ictvonline.org/>
288 (accessed Feb 23, 2017).
- 289 2. Bacterial and Archaeal Viruses Subcommittee. Available online:
290 <https://talk.ictvonline.org/information/w/members/441/bacterial-and-archaeal-viruses-subcommittee>
291 (accessed Feb 23, 2017).
- 292 3. Bradley, D. E. Ultrastructure of bacteriophage and bacteriocins. *Bacteriol. Rev.* **1967**, *31*, 230–314. Available
293 online: <http://mmbbr.asm.org/cgi/content/short/31/4/230> (accessed on Feb 23, 2017).
- 294 4. Ackermann, H.-W. Classification of bacteriophages. In *The Bacteriophages*, 2nd ed.; Calendar, R., Ed.; Oxford
295 University Press: New York, NY, USA, 2006; pp. 8–17.
- 296 5. Francki, R.; Fauquet, C.; Knudson, D.; Brown, F. *Classification and nomenclature of viruses: Fifth report of the*
297 *International Committee on the Taxonomy of Viruses*; Springer: Wien, NY, USA, 1991.
- 298 6. Fauquet, M. C.; Mayo, A. M. The 7th ICTV Report. *Arch. Virol.* **2001**, *146*, 189–194, DOI
299 10.1007/s007050170203
- 300 7. Fauquet, C.; Mayo, M.; Maniloff, J.; Desselberger, U.; Ball, L. *Virus taxonomy: Eighth report of the International*
301 *Committee on Taxonomy of Viruses*; Elsevier Academic Press: San Diego, CA, USA, 2005.
- 302 8. Nelson, D. Phage taxonomy: we agree to disagree. *J. Bacteriol.* **2004**, *186*, 7029–7031, DOI
303 10.1128/JB.186.21.7029-7031.2004.
- 304 9. Brister, J. R.; Ako-Adjei, D.; Bao, Y.; Blinkova, O. NCBI viral genomes resource. *Nucleic Acids Res.* **2015**, *43*,
305 D571–D577, DOI 10.1093/nar/gku1207.
- 306 10. Rohwer, F.; Edwards, R. The Phage Proteomic Tree: a genome-based taxonomy for phage. *J. Bacteriol.* **2002**,
307 *184*, 4529–4535, DOI 10.1128/JB.184.16.4529.
- 308 11. Lima-Mendez, G.; Van Helden, J.; Toussaint, A.; Leplae, R. Reticulate representation of evolutionary and
309 functional relationships between phage genomes. *Mol. Biol. Evol.* **2008**, *25*, 762–777, DOI
310 0.1093/molbev/msn023.
- 311 12. Pride, D. T.; Wassenaar, T. M.; Ghose, C.; Blaser, M. J. Evidence of host-virus co-evolution in tetranucleotide
312 usage patterns of bacteriophages and eukaryotic viruses. *BMC Genomics* **2006**, *7*, 8, DOI 10.1186/1471-2164-
313 7-8.
- 314 13. Asare, P. T.; Jeong, T.; Ryu, S.; Klumpp, J.; Loessner, M. J.; Merrill, B. D.; Kim, K. Putative type 1 thymidylate
315 synthase and dihydrofolate reductase as signature genes of a novel bastille-like group of phages in the
316 subfamily Spounavirinae. *BMC Genomics* **2015**, *16*, 582, DOI 10.1186/s12864-015-1757-0.
- 317 14. Grose, J. H.; Casjens, S. R. Understanding the enormous diversity of bacteriophages: The tailed phages that
318 infect the bacterial family *Enterobacteriaceae*. *Virology* **2014**, *468–470*, 421–443, DOI
319 10.1016/j.virol.2014.08.024.
- 320 15. Adriaenssens, E. M.; Edwards, R.; Nash, J. H. E.; Mahadevan, P.; Seto, D.; Ackermann, H.-W.; Lavigne, R.;
321 Kropinski, A. M. Integration of genomic and proteomic analyses in the classification of the *Siphoviridae*
322 family. *Virology* **2015**, *477*, 144–154, DOI 10.1016/j.virol.2014.10.016.

- 323 16. Lavigne, R.; Seto, D.; Mahadevan, P.; Ackermann, H.-W.; Kropinski, A. M. Unifying classical and molecular
324 taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. *Res. Microbiol.* **2008**, *159*,
325 406–414, DOI 10.1016/j.resmic.2008.03.005.
- 326 17. Lavigne, R.; Darius, P.; Summer, E. J.; Seto, D.; Mahadevan, P.; Nilsson, A. S.; Ackermann, H.-W.;
327 Kropinski, A. M. Classification of *Myoviridae* bacteriophages using protein sequence similarity. *BMC*
328 *Microbiol.* **2009**, *9*, 224, DOI 10.1186/1471-2180-9-224.
- 329 18. Krupovic, M.; Dutilh, B. E.; Adriaenssens, E. M.; Wittmann, J.; Vogensen, F. K.; Sullivan, M. B.; Rumnieks,
330 J.; Prangishvili, D.; Lavigne, R.; Kropinski, A. M.; Klumpp, J.; Gillis, A.; Enault, F.; Edwards, R. A.; Duffy,
331 S.; Clokie, M. R. J.; Barylski, J.; Ackermann, H.-W.; Kuhn, J. H. Taxonomy of prokaryotic viruses: update
332 from the ICTV bacterial and archaeal viruses subcommittee. *Arch. Virol.* **2016**, *4*, 1095-1099, DOI
333 10.1007/s00705-015-2728-0.
- 334 19. Adriaenssens, E. M.; Krupovic, M.; Knezevic, P.; Ackermann, H.-W.; Barylski, J.; Brister, J. R.; Clokie, M. R.
335 C.; Duffy, S.; Dutilh, B. E.; Edwards, R. A.; Enault, F.; Jang, H. Bin; Klumpp, J.; Kropinski, A. M.; Lavigne,
336 R.; Poranen, M. M.; Prangishvili, D.; Rumnieks, J.; Sullivan, M. B.; Wittmann, J.; Oksanen, H. M.; Gillis, A.;
337 Kuhn, J. H. Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses
338 subcommittee. *Arch. Virol.* **2017**, in press, DOI 10.1007/s00705-016-3173-4.
- 339 20. Virus Taxonomy: 2015 Release. Available online: <http://talk.ictvonline.org/taxonomy/> (accessed Feb 23,
340 2017).
- 341 21. Karsch-Mizrachi, I.; Nakamura, Y.; Cochrane, G. The International Nucleotide Sequence Database
342 Collaboration. *Nucleic Acids Res.* **2012**, *40*, D33–D37, DOI 10.1093/nar/gkr1006.
- 343 22. Ackermann, H.-W.; Abedon, S. T. Bacteriophage names 2000. Available online:
344 <http://www.phage.org/names/2000/> (accessed Feb 23, 2017).
- 345 23. NCBI Nucleotide Database. Available online: <http://www.ncbi.nlm.nih.gov/nucleotide/> (accessed Feb 24,
346 2017).
- 347 24. Kropinski, A. M.; Prangishvili, D.; Lavigne, R. Position paper: the creation of a rational scheme for the
348 nomenclature of viruses of Bacteria and Archaea. *Environ. Microbiol.* **2009**, *11*, 2775–2777, DOI
349 10.1111/j.1462-2920.2009.01970.x.
- 350 25. ICTV Code. Available online: <http://talk.ictvonline.org/information/w/ictv-information/383/ictv-code>
351 (accessed Feb 23, 2017).
- 352 26. National Center for Biotechnology Information (US), The GenBank submissions handbook [Internet].
353 Available online: <https://www.ncbi.nlm.nih.gov/books/NBK51157/> (accessed Feb 23, 2017).
- 354 27. O’Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.;
355 Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin,
356 V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C. M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina,
357 W.; Joardar, V. S.; Kodali, V. K.; Li, W.; Maglott, D.; Masterson, P.; McGarvey, K. M.; Murphy, M. R.; O’Neill,
358 K.; Pujar, S.; Rangwala, S. H.; Rausch, D.; Riddick, L. D.; Schoch, C.; Shkeda, A.; Storz, S. S.; Sun, H.;
359 Thibaud-Nissen, F.; Tolstoy, I.; Tully, R. E.; Vatsan, A. R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M. J.;
360 Kimchi, A.; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T. D.; Pruitt, K. D. Reference sequence (RefSeq)
361 database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**,
362 *44*, D733–D745, DOI 10.1093/nar/gkv1189.
- 363 28. Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezuk, Y.; McGinnis, S.; Madden, T. L. NCBI BLAST: a better
364 web interface. *Nucleic Acids Res.* **2008**, *36*, 5–9, DOI 10.1093/nar/gkn201.
- 365 29. Bao, Y.; Chetvernin, V.; Tatusova, T. Improvements to pairwise sequence comparison (PASC): a genome-
366 based web tool for virus classification. *Arch. Virol.* **2014**, *159*, 3293–3304, DOI 10.1007/s00705-014-2197-x.
- 367 30. Ågren, J.; Sundström, A.; Häfström, T.; Segerman, B. Gegenees: Fragmented alignment of multiple
368 genomes for determining phylogenomic distances and genetic signatures unique for specified target
369 groups. *PLoS One* **2012**, *7*, e39107, DOI 10.1371/journal.pone.0039107.
- 370 31. ICTV Taxonomy Proposals templates. Available online: [https://talk.ictvonline.org/files/taxonomy-](https://talk.ictvonline.org/files/taxonomy-proposal-templates/)
371 [proposal-templates/](https://talk.ictvonline.org/files/taxonomy-proposal-templates/) (accessed Feb 23, 2017).
- 372 32. Petrov, V. M.; Ratnayaka, S.; Nolan, J. M.; Miller, E. S.; Karam, J. D. Genomes of the T4-related
373 bacteriophages as windows on microbial genome evolution. *Virol. J.* **2010**, *7*, 292, DOI 10.1186/1743-422X-
374 7-292.

- 375 33. Roux, S.; Enault, F.; Ravet, V.; Pereira, O.; Sullivan, M. B. Genomic characteristics and environmental
376 distributions of the uncultivated Far-T4 phages. *Front. Microbiol.* **2015**, *6*, 1–13, DOI
377 10.3389/fmicb.2015.00199.
- 378 34. Krumsiek, J.; Arnold, R.; Rattei, T. Gepard: A rapid and sensitive tool for creating dotplots on genome scale.
379 *Bioinformatics* **2007**, *23*, 1026–1028, DOI 10.1093/bioinformatics/btm039.
- 380 35. Zafar, N.; Mazumder, R.; Seto, D. CoreGenes: A computational tool for identifying and cataloging “core”
381 genes in a set of small genomes. *BMC Bioinformatics* **2002**, *3*, 12, DOI 10.1186/1471-2105-3-12.
- 382 36. Mahadevan, P.; King, J. F.; Seto, D. CGUG: in silico proteome and genome parsing tool for the
383 determination of “core” and unique genes in the analysis of genomes up to ca. 1.9 Mb. *BMC Res. Notes* **2009**,
384 *2*, 168, DOI 10.1186/1756-0500-2-168.
- 385 37. Page, A. J.; Cummins, C. A.; Hunt, M.; Wong, V. K.; Reuter, S.; Holden, M. T. G.; Fookes, M.; Falush, D.;
386 Keane, J. A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*,
387 3691–3693, DOI 10.1093/bioinformatics/btv421.
- 388 38. Grazziotin, A. L.; Koonin, E. V.; Kristensen, D. M. Prokaryotic Virus Orthologous Groups (pVOGs): a
389 resource for comparative genomics and protein family annotation. *Nucleic Acids Res.* **2016**, *45*, D491–D498,
390 DOI 10.1093/nar/gkw975.
- 391 39. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.;
392 Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, scalable generation of high-quality protein multiple
393 sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539, DOI 10.1038/msb.2011.75.
- 394 40. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids*
395 *Res.* **2004**, *32*, 1792–1797, DOI 10.1093/nar/gkh340.
- 396 41. Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J.-F.; Guindon, S.; Lefort,
397 V.; Lescot, M.; Claverie, J.-M.; Gascuel, O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist.
398 *Nucleic Acids Res.* **2008**, *36*, W465–469, DOI 10.1093/nar/gkn180.
- 399 42. Tamura, K.; Stecher, G.; Peterson, D.; Filipinski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics
400 analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729, DOI 10.1093/molbev/mst197.
- 401 43. Price, M. N.; Dehal, P. S.; Arkin, A. P. FastTree 2 - Approximately maximum-likelihood trees for large
402 alignments. *PLoS One* **2010**, *5*, e9490, 10.1371/journal.pone.0009490.
- 403 44. Darling, A. E.; Mau, B.; Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss
404 and rearrangement. *PLoS One* **2010**, *5*, e11147, 10.1371/journal.pone.0011147.
- 405 45. Sullivan, M. J.; Petty, N. K.; Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **2011**, *27*,
406 1009–1010, DOI 10.1093/bioinformatics/btr039.
- 407 46. Alikhan, N.-F.; Petty, N. K.; Ben Zakour, N. L.; Beatson, S. A. BLAST Ring Image Generator (BRIG): simple
408 prokaryote genome comparisons. *BMC Genomics* **2011**, *12*, 402, DOI 10.1186/1471-2164-12-402.
- 409

