

1 **S2 Text. Generation of the Maximum-Likelihood phylogenies shown in**  
2 **this article.**

3 For the *Drexlerviridae* phylogeny shown in Fig 4B, the orthologs of a DNA helicase gene (*T1p47* in  
4 T1), the major capsid protein gene (*T1p47* in T1), the tail tape measure protein gene (*T1p38* in T1), the  
5 large terminase subunit gene (*T1p53* in T1), and another DNA helicase (*T1p09* in T1) were used. We aligned  
6 each set of orthologs, curated the alignments, concatenated them, and calculated a Maximum-Likelihood  
7 phylogeny as described above.

8 The phylogeny of *Drexlerviridae* and *Dhillonvirus* phages shown in S1A Fig is based on a  
9 concatenated alignment of amino acid sequences of major capsid protein (*T1p47* in T1) and large  
10 terminase subunit (*T1p53* in T1) of all phages that were included.

11 For the phylogenies of the three *Siphoviridae* genera shown in Fig 6B and Fig. 6D, each one whole-  
12 genome alignment was manually curated and then used to calculate a Maximum-Likelihood phylogeny.  
13 The lengths of these alignments were 21.8 kb (*Dhillonvirus*), 30.7 kb (*Nonagvirus*), and 33 kb (*Seuratvirus*).  
14 We present the phylogenies of *Nonagvirus* and *Seuratvirus* in Fig 6D on opposite sides of a single  
15 phylogeny because the two genera are well-known to be sister clades [1].

16 The phylogeny of *Markadamsvirinae* shown in Fig 7B was calculated based on a concatenation of  
17 curated nucleotide sequence alignments of the DNA polymerase gene (*T5.122* in T5), a DNA helicase gene  
18 (*T5.124* in T5), the major capsid protein gene (*T5.149* in T5), the DNA primase gene (*T5.108* in T5), another  
19 DNA helicase gene (*T5.119* in T5), the tail tape measure protein gene (*T5.140* in T5), and the large  
20 terminase subunit gene (*T5.155* in T5).

21 For the phylogeny of *Tevenvirinae* genera *Tequatrovirus* and *Mosigvirus* in Fig 8B, we used each  
22 one whole-genome alignment that was manually curated and then used to calculate Maximum-Likelihood  
23 phylogeny. The lengths of these alignments were 29.1 kb (*Tequatrovirus*) and 57.9 kb (*Mosigvirus*). The  
24 phylogeny of short tail fibers in Fig 8D was generated based on an amino acid sequence alignment of the  
25 orthologs in all included *Tevenvirinae* genomes (T4p157/Gp12 in T4).

26 The phylogeny of *Vequintavirinae* and relatives in Fig 9B was assembled from a phylogeny  
27 of *Vequintavirinae sensu stricto* (top) the related clusters of phages including phAPEC8 and phi92  
28 (bottom). We present these two phylogenies with a common root since these groups of phages are known  
29 to be closely related [1-3]. For the *Vequintavirinae sensu stricto*, the phylogeny was calculated based on  
30 a concatenation of curated nucleotide sequence alignments of three conserved loci of the phage core, the  
31 DNA packaging region (from the i-spanin gene (*rv5\_gp068* in rV5) until the first tail genes (*rv5\_gp054* in  
32 rV5), a locus comprising *rIIAB* and DNA replication functions (from *rIIA* (*rv5\_001* in rV5) to the DNA  
33 polymerase (*rv5\_gp223* in rV5), and a locus around the NTP reductase genes (from a *phoH*-like gene  
34 (*rv5\_gp115* in rV5) to the thymidylate synthase (*rv5\_gp102* in rV5). We curated the alignments,  
35 concatenated them, and calculated a Maximum-Likelihood phylogeny as described above. The phylogeny  
36 of phAPEC8-like and phi92-like phages was calculated based on a whole-genome alignment that was  
37 manually curated, resulting in a final length of 22.2 kb.

38 For the phylogeny of *Autographiviridae* phages in Fig 10B, we extracted the orthologs of the DNA  
39 polymerase gene (*T7p29* in T7), the DNA primase / helicase gene (*T7p22* in T7), the T3/T7 family RNA  
40 polymerase gene (*T7p07* in T7), and the large terminase subunit gene (*T7p57* in T7). We aligned each set  
41 of orthologs, curated the alignments, concatenated them, and calculated a Maximum-Likelihood  
42 phylogeny.

43           The phylogeny of *Enquatrovirus* phages and related *Podoviridae* in Fig 10E was calculated based  
44 on a concatenation of curated nucleotide sequence alignments of the DNA primase gene (*EPNV4\_gp43* in  
45 N4), the terminase large subunit gene (*EPNV4\_gp68* in N4), and the virion RNA polymerase gene  
46 (*EPNV4\_gp50* in N4).

47           For the phylogeny of *Felixounavirus* phages and related *Ounavirinae* in Fig 11B, we extracted the  
48 genes coding for a DNA ligase (*Felix01p163* in Felix O1), a DNA primase / helicase (*Felix01p188* in Felix O1),  
49 the major capsid protein (*Felix01p112* in Felix O1), and the tail tape measure protein (*Felix01p122* in Felix  
50 O1). Each set of orthologs was aligned and the alignments curated, concatenated, and used to calculate a  
51 Maximum-Likelihood phylogeny.

52

## 53 **References (S2 Text)**

- 54 1.       Korf IHE, Meier-Kolthoff JP, Adriaenssens EM, Kropinski AM, Nimtz M, Rohde M, et al. Still  
55 Something to Discover: Novel Insights into *Escherichia coli* Phage Diversity and Taxonomy. *Viruses*.  
56 2019;11(5). doi: 10.3390/v11050454. PubMed PMID: 31109012; PubMed Central PMCID:  
57 PMC6563267.
- 58 2.       Kropinski AM, Waddell T, Meng J, Franklin K, Ackermann HW, Ahmed R, et al. The host-range,  
59 genomics and proteomics of *Escherichia coli* O157:H7 bacteriophage rV5. *Virology*. 2013;10:76. Epub  
60 2013/03/19. doi: 10.1186/1743-422X-10-76. PubMed PMID: 23497209; PubMed Central PMCID:  
61 PMC3606486.
- 62 3.       Schwarzer D, Buettner FF, Browning C, Nazarov S, Rabsch W, Bethe A, et al. A multivalent  
63 adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and  
64 structural analysis. *Journal of virology*. 2012;86(19):10384-98. Epub 2012/07/13. doi: 10.1128/JVI.00801-  
65 12. PubMed PMID: 22787233; PubMed Central PMCID: PMC3457257.

66