

Brief Report

# From Orphan Phage to a Proposed New Family—the Diversity of N4-Like Viruses

Johannes Wittmann <sup>1,\*</sup>, Dann Turner <sup>2</sup>, Andrew D. Millard <sup>3</sup>, Padmanabhan Mahadevan <sup>4</sup>, Andrew M. Kropinski <sup>5,6</sup> and Evelien M. Adriaenssens <sup>7</sup>

<sup>1</sup> Leibniz Institute DSMZ–German Collection of Microorganisms and Cell Cultures, 38124 Braunschweig, Germany

<sup>2</sup> Department of Applied Sciences, University of the West of England, Bristol BS16 1QY, UK; dann2.turner@uwe.ac.uk

<sup>3</sup> Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH UK; adm39@le.ac.uk

<sup>4</sup> Department of Biology, University of Tampa, Tampa, FL 33606, USA; pmahadevan1@gmail.com

<sup>5</sup> Department of Food Science, University of Guelph, Guelph, ON N1G 2W1, Canada; phage.canada@gmail.com

<sup>6</sup> Department of Pathobiology, University of Guelph, Guelph, ON N1G 2W1, Canada

<sup>7</sup> Quadram Institute Bioscience, Norwich Research Park, Norwich NR4 7UQ, UK; evelien.adriaenssens@quadram.ac.uk

\* Correspondence: Johannes.Wittmann@dsmz.de; Tel.: +49-531-2616-214

Received: 8 September 2020; Accepted: 29 September 2020; Published: 30 September 2020

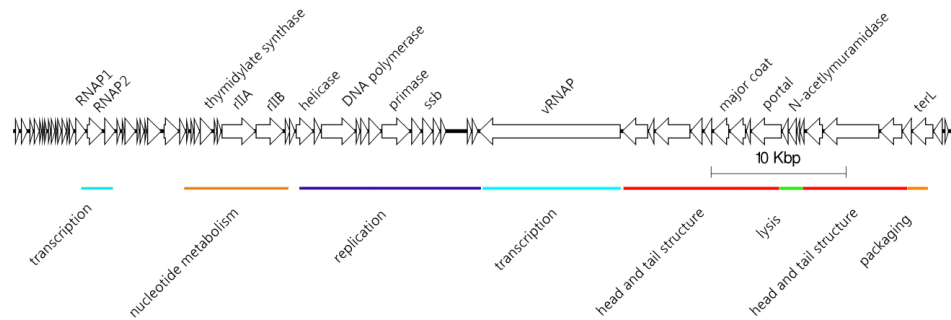
**Abstract:** Escherichia phage N4 was isolated in 1966 in Italy and has remained a genomic orphan for a long time. It encodes an extremely large virion-associated RNA polymerase unique for bacterial viruses that became characteristic for this group. In recent years, due to new and relatively inexpensive sequencing techniques the number of publicly available phage genome sequences expanded rapidly. This revealed new members of the N4-like phage group, from 33 members in 2015 to 115 N4-like viruses in 2020. Using new technologies and methods for classification, the Bacterial and Archaeal Viruses Subcommittee of the International Committee on Taxonomy of Viruses (ICTV) has moved the classification and taxonomy of bacterial viruses from mere morphological approaches to genomic and proteomic methods. The analysis of 115 N4-like genomes resulted in a huge reassessment of this group and the proposal of a new family “*Schitoviridae*”, including eight subfamilies and numerous new genera.

**Keywords:** N4; *Schitoviridae*; virus taxonomy; ICTV; bacteriophages; bacterial viruses

## 1. Introduction

Escherichia phage N4 is a virulent phage that was originally isolated by Gian Carlo Schito from sewers in Genoa (Italy) in 1966 [1]. TEM analysis revealed a 70-nm-diameter capsid and a short tail, the characteristic features of a podovirus. Its genome consists of double-stranded DNA and has a length of 70,153 bp with about 400 bp direct repeats and short 3'-noncohesive extensions [2]. Analysis of the N4 genome and its replication revealed unique characteristics. Apart from a phenomenon called lysis-inhibition [3] that causes delayed lysis and a subsequently increased burst size upon infection, further analysis revealed a gene for a large virion-associated RNA polymerase (vRNAP) that became a characteristic for N4-like genomes and the use of in total three DNA-dependent RNA polymerases for transcription that were subject to different scientific questions and thus were

intensively studied (Figure 1) [4]. The vRNAP is injected into the host cell along with DNA [5]. N4 genome is transcribed in three different temporal stages [6].



**Figure 1.** Genome structure of Escherichia phage N4 (70,153 bp) visualized by EasyFig [7].

From a taxonomic perspective, phage N4 has a long history; from the first proposal in 1987 to establish a species *Escherichia phage N4* [8] and its subsequent renaming to *Escherichia virus N4* in 2015, it has persisted as a genomic orphan. At the time of writing (Master Species List #35, ratified March 2020), its genus is now called *Enquatrovirus* and consists of only one species, representing four isolates.

Since 2015, following a comprehensive analysis of at that time 33 N4-like genomes [9], the number of publicly available N4-like phage genomes has nearly tripled [10]. The last report of the Bacterial and Archaeal Viruses Subcommittee [11] presented the new taxonomic classifications and reassessments that were achieved in 2018 and 2019 and listed a new order (*Tubulavirales*), ten new families, 22 new sub-families, 424 new genera and 964 new species, which still represent only a fraction of the genomes currently available. However, it has to be taken into account that ICTV does not classify viral strains or variants, i.e., those phage isolates with genomes that show  $\geq 95\%$  DNA sequence identity with an exemplar isolate of a species [12]. With regard to N4-like viruses, i.e., viruses encoding the vRNAP, only a rather small number of those have been officially classified by the ICTV so far. Currently, they are classified in 10 genera (*Baltimorevirus*, *Enquatrovirus*, *Gamaleyavirus*, *Ithacavirus*, *Johnsonvirus*, *Jwalphavirus*, *Litunavirus*, *Luzseptimavirus*, *Mukerjeevirus* and *Shizishanvirus*). This study provides further insight into the diversity and taxonomy of N4-like viruses using different approaches like genome-based phylogeny for deeper classification.

## 2. Results

### 2.1. Description of N4-Like Viruses

We downloaded 115 genomes from the NCBI databases (INSDC) [13,14] related to the N4-like group of viruses (Table 1). All N4-like phages and members of the proposed new family share the following characteristics:

- Podovirus morphology
- Genome size of 59–80 kb
- Linear genome with defined ends (terminal repeats expected)
- Presence of three RNA polymerase genes, including a large (~3500 aa) virion-associated RNA polymerase (vRNAP)

**Table 1.** List of N4-like genomes proposed to belong to the new family “Schitoviridae” available from INSDC databases.

Phage	Subfamily	Genus	Reference
Achromobacter phage JWAlpha	“Rothmandenesvirinae”	Jwalphavirus	[15]
Achromobacter phage JWDelta	“Rothmandenesvirinae”	Jwalphavirus	[15]
Achromobacter phage phiAxp-3	“Rothmandenesvirinae”	“Dongdastovirus”	[16]
Achromobacter phage vB_AxyP_19-32_Axy04	“Rothmandenesvirinae”	“Dongdastovirus”	[17]
Achromobacter phage vB_AxyP_19-32_Axy10	“Rothmandenesvirinae”	“Pourcelvirus”	[17]
Achromobacter phage vB_AxyP_19-32_Axy11	“Rothmandenesvirinae”	“Pourcelvirus”	[17]
Achromobacter phage vB_AxyP_19-32_Axy12	“Rothmandenesvirinae”	“Dongdastovirus”	[17]
Achromobacter phage vB_AxyP_19-32_Axy13	“Rothmandenesvirinae”	“Inbricusovirus”	[17]
Achromobacter phage vB_AxyP_19-32_Axy22	“Rothmandenesvirinae”	“Pourcelvirus”	[17]
Achromobacter phage vB_AxyP_19-32_Axy24	“Rothmandenesvirinae”	“Dongdastovirus”	[17]
Acinetobacter phage Presley		“Presleyvirus”	[18]
Acinetobacter phage VB_ApiP_XC38			[19]
Delftia phage RG-2014		“Dendoorenavirus”	[20]
Dinoroseobacter phage DFL12phi1	“Rhodovirinae”	Baltimorevirus	
Dinoroseobacter phage DS-1410Ws-06	“Rhodovirinae”	“Sanyabayovirus”	[21]
Dinoroseobacter phage vBDshPR2C	“Rhodovirinae”	Baltimorevirus	[22]
Enterobacter phage EcP1		“Eceepunavirus”	
Enterobacteria phage N4	“Enquatrovirinae”	Enquatrovirus	[3]
Erwinia phage Ea9-2	“Erskinevirinae”	Johnsonvirus	
Erwinia phage phiEaP-8	“Erskinevirinae”	Yonginivirus	[23]
Erwinia phage vB_EamP_Frozen	“Erskinevirinae”	Johnsonvirus	[24]
Erwinia phage vB_EamP_Gutmeister	“Erskinevirinae”	Johnsonvirus	[24]
Erwinia phage vB_EamP_Rexella	“Erskinevirinae”	Johnsonvirus	[24]
Erwinia phage vB_EamP-S6		“Waedenswilovirus”	[25]
Escherichia phage Bp4	“Enquatrovirinae”	Gamaleyavirus	[26]
Escherichia phage EC1-UPM	“Enquatrovirinae”	Gamaleyavirus	[27]
Escherichia phage ECBP1	“Enquatrovirinae”	Gamaleyavirus	[28]
Escherichia phage IME11	“Enquatrovirinae”	Gamaleyavirus	[29]
Escherichia phage OLB145	“Enquatrovirinae”	Enquatrovirus	
Escherichia phage PD38	“Enquatrovirinae”	Gamaleyavirus	
Escherichia phage PGN829.1	“Enquatrovirinae”	Gamaleyavirus	
Escherichia phage phi G17	“Enquatrovirinae”	Gamaleyavirus	[30]
Escherichia phage PMBT57	“Enquatrovirinae”	Enquatrovirus	
Escherichia phage Pollock	“Humphriesvirinae”	“Pollockvirus”	[31]
Escherichia phage St11Ph5	“Enquatrovirinae”	Gamaleyavirus	[32]
Escherichia phage vB_EcoP_3HA13	“Enquatrovirinae”	Enquatrovirus	
Escherichia phage vB_EcoP_G7C	“Enquatrovirinae”	Gamaleyavirus	[33]
Escherichia phage vB_EcoP_PhAPEC5	“Enquatrovirinae”	Gamaleyavirus	[34]
Escherichia phage vB_EcoP_PhAPEC7	“Enquatrovirinae”	Gamaleyavirus	[34]
Klebsiella phage KP8	“Enquatrovirinae”	“Kaypocavirus”	[35]
Klebsiella phage KpCHEMY26	“Humphriesvirinae”	“Pylasvirus”	[36]
Klebsiella phage Pylas	“Humphriesvirinae”	“Pylasvirus”	[37]
Pectobacterium phage Nepra		“Cbunavirus”	
Pectobacterium phage phiA41		“Cbunavirus”	[38]
Pectobacterium phage vB_PatP_CB1		“Cbunavirus”	[39]
Pectobacterium phage vB_PatP_CB3		“Cbunavirus”	[39]
Pectobacterium phage vB_PatP_CB4		“Cbunavirus”	[39]
Pseudoalteromonas phage pYD6-A	“Fuhrmanvirinae”	“Mazuvirus”	
Pseudomonas phage 98PfluR60PP		“Littlefixvirus”	[40]
Pseudomonas phage DL64	“Migulavirinae”	Litunavirus	[41]
Pseudomonas phage inbricus	“Rothmandenesvirinae”	“Inbricusovirus”	
Pseudomonas phage KPP21	“Migulavirinae”	Luzseptimavirus	[42]
Pseudomonas phage LIT1	“Migulavirinae”	Litunavirus	[43]
Pseudomonas phage Littlefix		“Littlefixvirus”	

Pseudomonas phage LP14	"Migulavirinae"	Litunavirus	[44]
Pseudomonas phage LUZ7	"Migulavirinae"	Luzseptimavirus	[43]
Pseudomonas phage LY218	"Migulavirinae"	Litunavirus	
Pseudomonas phage Pa2	"Migulavirinae"	Litunavirus	
Pseudomonas phage PA26	"Migulavirinae"	Litunavirus	[45]
Pseudomonas phage PEV2	"Migulavirinae"	Litunavirus	[43]
Pseudomonas phage pHcDa		Shizishanvirus	
Pseudomonas phage phi176	"Migulavirinae"	Litunavirus	[46]
Pseudomonas phage RWG	"Migulavirinae"	Litunavirus	[46]
Pseudomonas phage vB_Pae1396P-5	"Migulavirinae"	Litunavirus	
Pseudomonas phage vB_Pae575P-3	"Migulavirinae"	Litunavirus	
Pseudomonas phage vB_PaeP_C2-10_Ab09	"Migulavirinae"	Litunavirus	[47]
Pseudomonas phage vB_PaeP_DEV	"Migulavirinae"	Litunavirus	[48]
Pseudomonas phage vB_PaeP_MAG4	"Migulavirinae"	Litunavirus	[49]
Pseudomonas phage vB_PaeP_PYO2	"Migulavirinae"	Litunavirus	[48]
Pseudomonas phage YH30	"Migulavirinae"	Litunavirus	[50]
Pseudomonas phage YH6	"Migulavirinae"	Litunavirus	[51]
Pseudomonas phage ZC03		"Zicotriavirus"	[52]
Pseudomonas phage ZC08		"Zicotriavirus"	[52]
Pseudomonas phage Zuri		"Zurivirus"	
Roseobacter phage RD-1410W1-01	"Rhodovirinae"	"Aquinivirus"	[21]
Roseobacter phage RD-1410Ws-07	"Rhodovirinae"	"Sanyabayvirus"	[21]
Roseovarius Plymouth podovirus 1	"Rhodovirinae"	"Plymouthvirus"	[53]
Roseovarius sp. 217 phage 1	"Rhodovirinae"	"Plymouthvirus"	[53]
Ruegeria phage vB_RpoP-V12	"Rhodovirinae"	"Aorunvirus"	
Ruegeria phage vB_RpoP-V13	"Rhodovirinae"	"Pomeroyivirus"	
Ruegeria phage vB_RpoP-V14	"Rhodovirinae"	"Aorunvirus"	
Ruegeria phage vB_RpoP-V17	"Rhodovirinae"	"Aorunvirus"	
Ruegeria phage vB_RpoP-V21	"Rhodovirinae"	"Aorunvirus"	
Salmonella phage FSL SP-058	"Humphriesvirinae"	"Ithacavirus"	[54]
Salmonella phage FSL SP-076	"Humphriesvirinae"	"Ithacavirus"	[54]
Shigella phage pSb-1	"Enquatrovirinae"	Gamaleyavirus	[55]
Silicibacter phage DSS3phi2	"Rhodovirinae"	"Aorunvirus"	[56]
Sinorhizobium phage ort11		"Huelvavirus"	[57]
Stenotrophomonas phage Pokken		"Pokkenvirus"	[58]
Sulfitobacter phage EE36phi1	"Rhodovirinae"	"Aorunvirus"	[56]
Sulfitobacter phage phiCB2047-B	"Rhodovirinae"	"Raunefjordvirus"	[59]
Vibrio phage 1.025.O_10N.222.46.B6	"Pontosvirinae"	"Nahantovirus"	
Vibrio phage 1.026.O_10N.222.49.C7	"Pontosvirinae"	"Nahantovirus"	
Vibrio phage 1.097.O_10N.286.49.B3	"Pontosvirinae"	"Dorisvirus"	
Vibrio phage 1.150.O_10N.222.46.A6	"Pontosvirinae"	"Nahantovirus"	
Vibrio phage 1.152.O_10N.222.46.E1	"Pontosvirinae"	"Nahantovirus"	
Vibrio phage 1.169.O_10N.261.52.B1		Mukerjeevirus	
Vibrio phage 1.188.A_10N.286.51.A6		Mukerjeevirus	
Vibrio phage 1.224.A_10N.261.48.B1		Mukerjeevirus	
Vibrio phage 1.261.O_10N.286.51.A7		Mukerjeevirus	
Vibrio phage 2.130.O_10N.222.46.C2	"Pontosvirinae"	"Nahantovirus"	
Vibrio phage JA-1		"Pacivirus"	[60]
Vibrio phage JSF3		"Pacivirus"	[61]
Vibrio phage phi 1		"Pacivirus"	[62]
Vibrio phage phi50-12			
Vibrio phage pVa5	"Pontosvirinae"	"Galateavirus"	[63]
Vibrio phage pVco-5			[64]
Vibrio phage VBP32	"Fuhrmanvirinae"	"Stoningtonvirus"	
Vibrio phage VBP47	"Fuhrmanvirinae"	"Stoningtonvirus"	
Vibrio phage VCO139		"Pacivirus"	[60]
Vibrio virus vB_VspP_SBP1			
Xanthomonas phage RiverRider		"Riverridervirus"	[65]

**From metagenomes**  
*Podoviridae* sp. isolate ctda\_1  
*Podoviridae* sp. ctLUJ1  
*Siphoviridae* sp. isolate 355

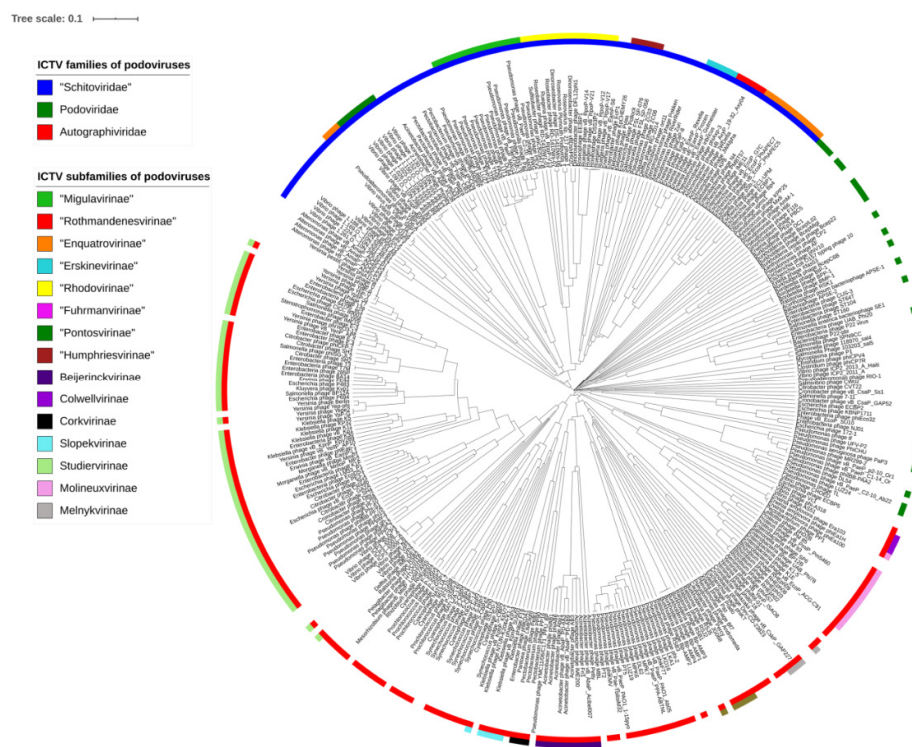
---

*"Enquatrovirinae"*      *Gamaleyavirus*

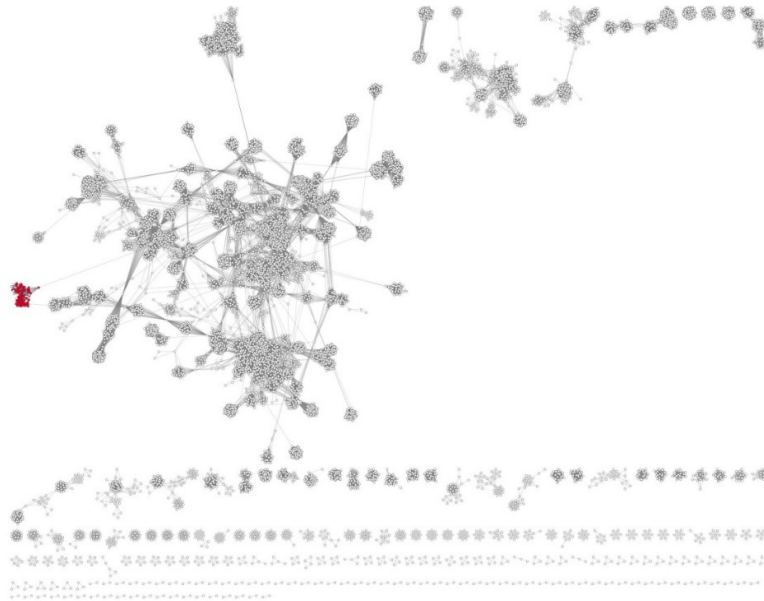
So far, only N4-like phages infectious for Gram-negative host bacteria belonging to the  $\alpha$ -proteobacteria such as *Roseobacter* [21],  $\beta$ -proteobacteria such as *Achromobacter* [3] and  $\gamma$ -proteobacteria such as *Pseudomonas* [43] from different habitats have been described. From the morphological perspective, N4-like phages show characteristic features of podoviruses, capsid sizes range from 50 [35] to 85 nm [63] with short non-contractile tails.

## 2.2. Proposal of a New Family

To analyze the similarities or relationship, respectively, between N4-like viruses and other podoviruses, we used ViPTree (<https://www.genome.jp/viptree/>; [66]) which is originally based on the Phage Proteomic Tree [67]. The results showed that the group of N4-like is clearly monophyletic and forms a distinct clade (Figure 2). The distinct clustering of the newly proposed family was confirmed with a gene-sharing network analysis using vConTACT2 (Figure 3), where the N4-like viruses cluster clearly separates from all other dsDNA bacterial viruses. In fact, the deep branch lengths in the ViPTree and limited connectedness in the gene-sharing network show that there are no unifying genomic features among all members of the *Podoviridae* to justify the current membership.



**Figure 2.** ViPTree analysis of N4-like viruses with related podoviruses. Results were visualized with iTol. Viruses were assigned and marked according to the official ICTV classification with the outer and inner rings representing classification at the subfamily and family level, respectively. Non-marked viruses have not been classified yet.



**Figure 3.** VConTACT2 network analysis. Members of the proposed “*Schitoviridae*” family are marked in red.

Panproteome analysis revealed that seventeen N4-like proteins are conserved in this proposed family of phages: RNAP 1 (EPNV4\_gp15), RNAP 2 (EPNV4\_gp16), vRNAP (EPNV4\_gp50), EPNV4\_gp24, EPNV4\_gp25, DNA polymerase (EPNV4\_gp39), EPNV4\_gp42, DNA primase (EPNV4\_gp43), EPNV4\_gp44, EPNV4\_gp52, EPNV4\_gp54, major capsid protein (EPNV4\_gp56), tape measure protein (EPNV4\_gp57), portal protein (EPNV4\_gp59), EPNV4\_gp67, large terminase subunit (EPNV4\_gp68), and EPNV4\_gp69 (Table 2).

**Table 2.** Panproteome analysis of N4-like viruses using three different approaches.

#	N4_Product	N4 Locus Tag	N4 Protein Accession	PIRATE	Proteinortho_30	CoreGenes 5.0
1	RNAP 1	EPNV4_gp15	YP_950493.1	Y	Y	Y
2	RNAP 2	EPNV4_gp16	YP_950494.1	N	Y	Y
3	AAA+ ATPase	EPNV4_gp24	YP_950502.1	*	Y	Y
4	gp25	EPNV4_gp25	YP_950503.1	N	N	Y
5	DNA polymerase	EPNV4_gp39	YP_950517.1	Y	Y	Y
6	gp42	EPNV4_gp42	YP_950520.1	Y	Y	Y
7	DNA primase	EPNV4_gp43	YP_950521.1	Y	Y	Y
8	gp44	EPNV4_gp44	YP_950522.1	Y	Y	Y
9	vRNAP	EPNV4_gp50	YP_950528.1	N	N	Y
10	16.5 kDa protein	EPNV4_gp52	YP_950530.1	Y	Y	Y
11	gp54	EPNV4_gp54	YP_950532.1	N	N	Y
12	Major capsid protein	EPNV4_gp56	YP_950534.1	Y	Y	Y
13	gp57 (tape measure)	EPNV4_gp57	YP_950535.1	N	*	Y
14	94 kDa protein (portal vertex protein)	EPNV4_gp59	YP_950537.1	Y	Y	Y
15	30 kDa protein	EPNV4_gp67	YP_950545.1	N	N	Y
16	Terminase, large subunit	EPNV4_gp68	YP_950546.1	Y	Y	Y
17	gp69	EPNV4_gp69	YP_950547.1	Y	N	Y

\* Indicates presence in 113/114 genomes.

Based on the different analyses, we propose a new family “*Schitoviridae*” in honor of Gian Carlo Schito who isolated *Escherichia* phage N4, the first isolated species of this group.

### 2.3. Proposal of New Subfamilies and Genera

Results of an all-by-all pairwise nucleotide identity analysis or intergenomic similarity analysis with VIRIDIC gave strong evidence for the proposal of eight new subfamilies and 30 genera which were confirmed by phylogenetic analysis of the terminase large subunit and vRNA polymerase genes, i.e., all proposed taxa are monophyletic in these marker gene trees (Supplementary Figure S1 and S2). In line with previously established taxa, we used 95% and 70% nucleotide sequence identity over the length of the genome as species and genus demarcation criteria, respectively [11,12,68,69]. At the subfamily level, members of the same subfamily share at least 40% intergenomic distance as calculated with VIRIDIC, with members of different subfamilies sharing little to no nucleotide identity [70].

The proposed subfamily “*Migulavirinae*” consists of two previously ratified genera (eight species), *Litunavirus* and *Luzseptimavirus*, representing phages with *Pseudomonas aeruginosa* as their host. The subfamily “*Enquatrovirinae*” contains three genera (14 species), *Gamaleyavirus*, *Enquatrovirus* and the newly proposed genus “*Kaypocavirus*” and includes phages infecting members of the *Enterobacteriaceae* like *E. coli*, *Shigella boydii* or *Klebsiella pneumoniae*. N4-like viruses infecting *Achromobacter xylosoxidans* were grouped into four genera (eight species) in the proposed subfamily “*Rothmandenesvirinae*” in honour of Lucia Rothman-Denes, who worked on N4 and its RNA polymerases. The subfamily “*Erskinevirinae*” was named after John M. Erskine who in the early 1970s was one of the first people to isolate phages against *Erwinia*. It consists of two genera, “*Yonginivirus*” and *Johnsonvirus*, with three species and represents most of the N4-like viruses against *Erwinia*. The relatively large subfamily “*Rhodovirinae*” consists of seven genera, “*Aorunvirus*”, “*Raunefjordvirus*”, “*Aoquinivirus*”, “*Pomeroyivirus*”, “*Sanyabayvirus*”, “*Plymouthvirus*” and *Baltimorevirus*, and contains aquatic viruses infecting members of the *Rhodobacteraceae*. Two further proposed subfamilies (five genera), “*Fuhrmanvirinae*” (named after American oceanographer and marine biologist Jed Alan Fuhrman) and “*Pontosvirinae*”, mainly consist of phages against marine *Vibrio* species. The “*Humphriesvirinae*” subfamily in honour of James C. Humphries (1914–1992), who was the first to isolate a *Klebsiella* phage, comprises five genera with viruses infecting different genera of the *Enterobacteriaceae* like *Escherichia*, *Klebsiella* or *Salmonella*.

### 3. Discussion

The constantly rising number of sequences provides the scientific community with valuable data to work with to answer various scientific questions. However, the taxonomic classification of phage genomes has not kept pace which has led to the presence of large numbers of unclassified genomes in the INSDC. While the ICTV makes a huge effort to manage this problem and improvements have been made on the genus and subfamily level (2019: 103 proposals, 2020: 188 proposals submitted [68]), it is clear that at the family level that concerted efforts, both by the ICTV and the wider community of phage biologists are required to address the issue of family-level classification. The creation of the family *Herelleviridae* from the subfamily *Spounavirinae* and related phages [69], provided the blueprint for the creation of new families of tailed phages, and the start to the dismantling of the morphology-based families *Myoviridae*, *Siphoviridae* and *Podoviridae*. Following from that example, we used some of the methods trialed and tested for the creation of a new family (Phage Proteomic Tree, vConTACT2) and the delineation of its internal structure (genome-distance comparisons, phylogenetic analysis of signature genes) to define the new family “*Schitoviridae*” of N4-like phages, to be removed from the family *Podoviridae*.

## 4. Materials and Methods

### 4.1. vConTACT2 Analysis

To create the gene-sharing network, a total of 16,050 phage contigs (<http://millardlab.org/>) [71] were reannotated using Prodigal v2.6.3 and clustered using vConTACT.2.0 [72] and the ProkaryoticViralRefSeq database v94. The resulting network was visualised and annotated using Cytoscape v3.8.0.

### 4.2. Panproteome Analysis

To identify conserved proteins present in bacteriophages comprising the proposed family, all genomes were reannotated using Prokka v1.14.5 [73] and predicted CDS mapped against the VOG hmm database using hmmscan. GFF3 files or protein FASTA files were used as input for Proteinortho v6 [74] and PIRATE v1.0.4 [75], respectively.

For panproteome construction with PIRATE the settings used were 30 and 35% identity threshold, cdhit lowest percentage id of 95 and e-value for BLAST hit filtering of 1E-5. For Proteinortho, the search options were adjusted so that the minimum percent identity and coverage of the best blast hits were 30% and 50%, respectively. All other parameters were left as default.

The CoreGenes5.0 webserver (<https://coregenes.ngrok.io/>) was used with the OrthoMCL option with E-value of 1e-5. CoreGenes5.0 uses the GET\_HOMOLOGUES package to implement the ortholog clustering [76,77]. We considered signature genes to be gene products present in all members of the proposed family where there was consensus between two or more of the analyses.

### 4.3. VIRIDIC Analysis

The Bacterial and Archaeal Viruses Subcommittee uses nucleotide based sequence similarities as a crucial feature for taxonomic classification of viruses at the ranks of species and genus. We therefore employed the online tool VIRIDIC (Virus Intergenomic Distance Calculator, <http://rhea.icbm.uni-oldenburg.de/VIRIDIC/>) [70] for the calculation of pairwise intergenomic similarities amongst the phage genomes of this study. We have chosen 95% DNA sequence identity as the criterion for demarcation of species in genera. Each of the proposed species differs from the others with more than 5% at the DNA level. For the demarcation of genera and subfamilies, we have chosen 70% and 40% DNA sequence identity, respectively. Based on this analysis, new genera and subfamilies were identified (Supp. Table 1).

**Supplementary Materials:** The following are available online at [www.mdpi.com/2079-6382/9/10/663/s1](http://www.mdpi.com/2079-6382/9/10/663/s1), Figure S1: Phylogenetic analysis using the terminase protein sequences of N4-like phages, respectively. The amino acid sequences were compared using MUSCLE with MEGA7 [78]. The tree was constructed using the maximum likelihood algorithm. The percentages of replicate trees were assessed with the bootstrap test (100).; Figure S2: Phylogenetic analysis using the vRNA polymerase protein sequences of N4-like phages, respectively. The amino acid sequences were compared using MUSCLE with MEGA7 [78]. The tree was constructed using the maximum likelihood algorithm. The percentages of replicate trees were assessed with the bootstrap test (100).; Figure S3 ViPTree analysis of N4-like viruses with related podoviruses, Table S1: VIRIDIC analysis of N4-like phages.

**Author Contributions:** Conceptualization, J.W. and E.M.A.; software, P.M.; formal analysis, J.W., D.T., A.M.K., E.M.A.; writing—original draft preparation, J.W.; writing—review and editing, J.W., D.T., A.D.M., A.M.K., E.M.A.; visualization, J.W.; supervision, J.W. and E.M.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** Evelien M. Adriaenssens gratefully acknowledges the support of the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded by the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1 and its constituent project BBS/E/F/000PR10353. Computational analysis was carried out on infrastructure provided by CLIMB infrastructure, Medical Research Council (MRC) grant MR/L015080/1.

**Conflicts of Interest:** With the exception of P.M. all authors are members of the Bacterial and Archaeal Viruses Subcommittee of the International Committee on Taxonomy of Viruses (ICTV).



## References

- Schito, G.C.; Rialdi, G.; Pesce, A. Biophysical properties of N4 coliphage. *Biochim. Biophys. Acta.* **1966**, *129*, 482–490.
- Ohmori, H.; Haynes, L.L.; Rothman-Denes, L.B. Structure of the ends of the coliphage N4 genome. *J. Mol. Biol.* **1988**, *202*, 1–10.
- Paddison, P.; Abedon, S.T.; Dressman, H.K.; Gailbreath, K.; Tracy, J.; Mosser, E.; Neitzel, J.; Guttman, B.; Kutter, E. The roles of the bacteriophage T4 r genes in lysis inhibition and fine-structure genetics: A new perspective. *Genetics* **1998**, *148*, 1539–1550.
- Kazmierczak, K.M.; Rothman-Denes, L.B. Bacteriophage N4. in: the bacteriophages, 2nd edn. Calendar, R. (ed), Oxford University Press, New York, 2006, pp. 302–314.
- Falco, S.C.; Laan, K.V.; Rothman-Denes, L.B. Virion-associated RNA polymerase required for bacteriophage N4 development. *Proc. Natl. Acad. Sci.* **1977**, *74*, 520–523.
- Zivin, R.; Zehring, W.; Rothman-Denes, L.B. Transcriptional map of bacteriophage N4: Location and polarity of N4 RNAs. *J. Mol. Biol.* **1988**, *152*, 335–356.
- Sullivan, M.J.; Petty, N.K.; Beatson, S.A. Easyfig: A genome comparison visualizer. *Bioinformatics* **2011**, *27*, 1009–1010.
- Ackermann, H.W.; DuBow, M. Natural groups of bacteriophages. In: Viruses of prokaryotes. CRC Press, Boca Raton, **1987**, pp. 85–100.
- Wittmann, J.; Klumpp, J.; Moreno Switt, A.I.; Yagubi, A.; Ackermann, H.W.; Wiedemann, M.; Svircev, A.; Nash, H.E.; Kropinski, A.M. Taxonomic reassessment of N4-like viruses using comparative genomics and proteomics suggests a new subfamily - “*Enquartavirinae*”. *Arch. Virol.* **2015**, *160*, 3053–3062.
- Kropinski, A.M. Bacteriophage research—what we have learnt and what still needs to be addressed. *Res. Microbiol.* **2018**, *169*, 481–487.
- Adriaenssens, E.M.; Sullivan, M.B.; Knezevic, P.; van Zyl, L.J.; Sarkar, B.L.; Dutilh, B.E.; Alfenas-Zerbini, P.; Lobočka, M.; Tong, Y.; Brister, J.R.; et al. Taxonomy of prokaryotic viruses: 2018–2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Arch. Virol.* **2020**, *165*, 1253–1260.
- Adriaenssens, E.; Brister, J.R. How to name and classify your phage: An informal guide. *Viruses* **2017**, *9*, 70.
- Karsch-Mizrachi, I.; Nakamura, Y.; Cochrane, G. The international nucleotide sequence database collaboration. *Nucleic Acids Res.* **2012**, *40*, 33–37.
- Brister, J. R.; Ako-adjei D.; Bao, Y.; Blinkova, O. NCBI viral genomes resource. *Nucleic. Acids. Res.* **2015**, *43*, 571–577.
- Wittmann, J.; Dreiseikelmann, B.; Rohde, M.; Meier-Kolthoff, J.P.; Bunk, B.; Rohde, C. First genome sequences of *Achromobacter* phages reveal new members of the N4 family. *Virol. J.* **2014**, *27*, 11–14.
- Ma, Y.; Li, E.; Qi, Z.; Li, H.; Wei, X.; Lin, W.; Zhao, R.; Jiang, A.; Yang, H.; Yin, Z.; et al. Isolation and molecular characterisation of *Achromobacter* phage phiAxp-3, an N4-like bacteriophage. *Sci. Rep.* **2016**, *6*, 24776.
- Essoh, C.; Vernadet, J.P.; Vergnaud, G.; Coulibaly, A.; Kakou-N’Douba, A.; N’Guetta, A.S.P.; Ouassa, T.; Pourcel, C. Characterization of sixteen *Achromobacter xylosoxidans* phages from Abidjan, Côte d’Ivoire, isolated on a single clinical strain. *Arch. Virol.* **2020**, *165*, 725–730.
- Farmer, N.G.; Wood, T.L.; Chamakura, K.R.; Kutty Everett, G.F. Complete genome of *Acinetobacter baumannii* N4-Like podophage Presley. *Genome Announc.* **2013**, *1*, e00852–13.
- Cheng, M.; Luo, M.; Xi, H.; Zhao, Y.; Le, S.; Chen, L.K.; Tan, D.; Guan, Y.; Wang, T.; Han, W.; et al. The characteristics and genome analysis of vB\_ApiP\_XC38, a novel phage infecting *Acinetobacter pittii*. *Virus Genes* **2020**, *56*, 498–507.
- Bhattacharjee, A.S.; Motlagh, A.M.; Gilcrease, E.B.; Imdadul Islam, M.I.; Casjens, S.R.; Goel, R. Complete genome sequence of lytic bacteriophage RG-2014 that infects the multidrug resistant bacterium *Delftia tsuruhatensis* ARB-1. *Stand. Genom. Sci.* **2017**, *12*, 82.
- Li, B.; Zhang, S.; Long, L.; Huang, S. Characterization and complete genome sequences of three N4-like roseobacter phages isolated from the South China sea. *Curr. Microbiol.* **2016**, *73*, 409–418.
- Cai, L.; Yang, Y.; Jiao, N.; Zhang, R. Complete genome sequence of vB\_DshP-R2C, a N4-like lytic roseophage. *Mar. Genom.* **2015**, *22*, 15–17.
- Park, J.; Lee, G.M.; Kim, D.; Park, D.H.; Oh, C.S. Characterization of the lytic bacteriophage phiEaP-8 effective against both *Erwinia amylovora* and *Erwinia pyrifoliae* causing severe diseases in apple and pear. *Plant. Pathol. J.* **2018**, *34*, 445–450.

24. Esplin, I.N.D.; Berg, J.A.; Sharma, R.; Allen, R.C.; Arens, D.K.; Ashcroft, C.R.; Bairett, S.R.; Beatty, N.J.; Bickmore, M.; Bloomfield, T.J.; et al. Genome sequences of 19 novel *Erwinia amylovora* bacteriophages. *Genome Announc.* **2017**, *5*, e00931–17.
25. Born, Y.; Fieseler, L.; Marazzi, J.; Lurz, R.; Duffy, B.; Loessner, M.J. Novel virulent and broad-host-range *Erwinia amylovora* bacteriophages reveal a high degree of mosaicism and a relationship to *Enterobacteriaceae* phages. *Appl. Env. Microbiol.* **2011**, *77*, 5945–5954.
26. Zhang, C.; Ma, Y.; Wang, T.; Sun, H.; Lu, G.; Ren, H. Characterization and complete genome sequence of vB\_EcoP-Bp4, a novel polyvalent N4-like bacteriophage that infects chicken pathogenic *Escherichia coli*. *Virol. Sin.* **2016**, *31*, 353–356.
27. Gan, H.M.; Sieo, C.C.; Tang, S.G.H.; Omar, A.R.; Ho, Y.W. The complete genome sequence of EC1-UPM, a novel N4-like bacteriophage that infects *Escherichia coli* O78:K80. *Virol. J.* **2013**, *10*, 308.
28. Nho, S.W.; Ha, M.A.; Kim, K.S.; Kim, T.H.; Jang, H.B.; Cha, I.S.; Park, S.B.; Kim, Y.K.; Jung, T.S. Complete genome sequence of the bacteriophages ECBP1 and ECBP2 isolated from two different *Escherichia coli* strains. *J. Virol.* **2012**, *86*, 12439–12440.
29. Fan, H.; Fan, H.; An, X.; Huang, Y.; Zhang, Z.; Mi, Z.; Tong, Y. Complete genome sequence of IME11, a new N4-like bacteriophage. *J. Virol.* **2012**, *86*, 13861.
30. Akindolire, M.A.; Aremu, B.R.; Ateba, C.N. Complete genome sequence of *Escherichia coli* O157:H7 phage PhiG17. *Microbiol. Resour. Announc.* **2019**, *8*, e01296–18.
31. Patel, R.S.; Lessor, L.E.; Hernandez, A.C.; Kutty Everett, G.F. Complete genome sequence of enterotoxigenic *Escherichia coli* N4-like podophage pollock. *Genome Announc.* **2015**, *3*, e01431–14.
32. Golomidova, A.K.; Kulikov, E.E.; Babenko, V.V.; Kostryukova, E.S.; Letarov, A.V. Complete genome sequence of bacteriophage St11Ph5, which infects uropathogenic *Escherichia coli* strain up11. *Genome Announc.* **2018**, *6*, e01371–17.
33. Kulikov, E.; Kropinski, A.M.; Golomidova, A.; Lingohr, E.; Govorun, V.; Serebryakova, M.; Prokhorov, N.; Letarova, M.; Manykin, A.; Strotskaya, A.; et al. Isolation and characterization of a novel indigenous intestinal N4-related coliphage vB\_EcoP\_G7C. *Virology* **2012**, *426*, 93–99.
34. Tsonos, J.; Oosterik, L.H.; Tuntufye, H.N.; Klumpp, J.; Butaye, P.; De Greve, H.; Hernalsteens, J.P.; Lavigne, R.; Goddeeris, B.M. A cocktail of in vitro efficient phages is not a guarantee for in vivo therapeutic results against avian colibacillosis. *Vet. Microbiol.* **2014**, *171*, 470–479.
35. Morozova, V.; Babkin, I.; Kozlova, Y.; Baykov, I.; Bokovaya, O.; Tikunov, A.; Ushakova, T.; Bardasheva, A.; Ryabchikova, E.; Zelentsova, E.; et al. Isolation and characterization of a novel *Klebsiella pneumoniae* N4-like bacteriophage KP8. *Viruses* **2019**, *11*, 1115.
36. Yerushalmy, O.; Copenhagen-Glazer, S.; Nir-Paz, R.; Tuomala, H.; Skurnik, M.; Kiljunen, S.; Hazan, R. Complete genome sequences of two *Klebsiella pneumoniae* phages isolated as part of an international effort. *Microbiol. Resour. Announc.* **2019**, *8*, e00843–e19.
37. Powell, J.E.; Lessor, L.; O’Leary, C.; Gill, J.; Liu, M. Complete genome sequence of *Klebsiella pneumoniae* podophage pylas. *Microbiol. Resour. Announc.* **2019**, *8*, e01287–19.
38. Smolarska, A.; Rabalski, L.; Narajczyk, M.; Czajkowski, R. Isolation and phenotypic and morphological characterization of the first *Podoviridae* lytic bacteriophages  $\phi$ A38 and  $\phi$ A41 infecting *Pectobacterium parmentieri* (former *Pectobacterium wasabiae*). *Eur. J. Plant. Pathol.* **2018**, *150*, 413–425.
39. Buttimer, C.; Hendrix, H.; Lucid, A.; Neve, H.; Noben, J.P.; Franz, C.; O’Mahony, J.; Lavigne, R.; Coffey, A. Novel N4-Like Bacteriophages of *Pectobacterium atrosepticum*. *Pharm. (Basel)* **2018**, *11*, 45.
40. Kazimierczak, J.; Wójcik, E.A.; Witaszewska, J.; Guziński, A.; Górecka, E.; Stańczyk, M.; Kaczorek, E.; Siwicki, A.K.; Dastyk, J. Complete genome sequences of *Aeromonas* and *Pseudomonas* phages as a supportive tool for development of antibacterial treatment in aquaculture. *Virol. J.* **2019**, *16*, 4.
41. Alves, D.R.; Perez-Esteban, P.; Kot, W.; Bean, J.E.; Arnot, T.; Hansen, L.H.; Enright, M.C.; Jenkins, A.T. A novel bacteriophage cocktail reduces and disperses *Pseudomonas aeruginosa* biofilms under static and flow conditions. *Microb Biotechnol* **2016**, *9*, 61–74.
42. Shigehisa, R.; Uchiyama, J.; Kato, S.I.; Takemura-Uchiyama, I.; Yamaguchi, K.; Miyata, R.; Ujihara, T.; Sakaguchi, Y.; Okamoto, N.; Shimakura, H.; et al. Characterization of *Pseudomonas aeruginosa* phage KPP21 belonging to family *Podoviridae* genus N4-like viruses isolated in Japan. *Microbiol. Immunol.* **2016**, *60*, 64–67.
43. Ceysens, P.J.; Brabban, A.; Rogge, L.; Lewis, M.S.; Pickard, D.; Goulding, D.; Dougan, G.; Noben, J.P.; Kropinski, A.; Kutter, E.; et al. Molecular and physiological analysis of three *Pseudomonas aeruginosa* phages belonging to the “N4-like viruses”. *Virology* **2010**, *405*, 26–30.

44. Shi, X.; Zhao, F.; Sun, H.; Yu, X.; Zhang, C.; Liu, W.; Pan, Q.; Ren, H. Characterization and complete genome analysis of *Pseudomonas aeruginosa* bacteriophage vB\_PaeP\_LP14 belonging to genus *Litunavirus*. *Curr. Microbiol.* **2020**, *77*, 2465–2474.
45. Kim, M.S.; Cha, K.E.; Myung, H. Complete genome of *Pseudomonas aeruginosa* phage PA26. *J. Virol.* **2012**, *86*, 10244.
46. Burrowes, B.H.; Molineux, I.J.; Fralick, J.A. Directed in vitro evolution of therapeutic bacteriophages: The appelmans protocol. *Viruses* **2019**, *11*, 241.
47. Essoh, C.; Latino, L.; Midoux, C.; Blouin, Y.; Loukou, G.; Nguetta, S.P.A.; Lathro, S.; Cablanmian, A.; Kouassi, A.K.; Vergnaud, G.; et al. Investigation of a large collection of *Pseudomonas aeruginosa* bacteriophages collected from a single environmental source in Abidjan, Côte d’Ivoire. *Plos One* **2015**, *10*, e0130548.
48. Forti, F.; Roach, D.R.; Cafora, M.; Pasini, M.E.; Horner, D.S.; Fiscarelli, E.V.; Rossitto, M.; Cariani, L.; Briani, F.; Debarbieux, L.; et al. Design of a broad-range bacteriophage cocktail that reduces *Pseudomonas aeruginosa* biofilms and treats acute infections in two animal models. *Antimicrob. Agents Chemother.* **2018**, *62*, e02573–17.
49. Kwiatek, M.; Parasion, S.; Rutyna, P.; Mizak, L.; Gryko, R.; Niemcewicz, M.; Olender, A.; Łobocka, M. Isolation of bacteriophages and their application to control *Pseudomonas aeruginosa* in planktonic and biofilm models. *Res. Microbiol.* **2017**, *168*, 194–207.
50. Gu, J.; Li, X.; Yang, M.; Du, C.; Cui, Z.; Gong, P.; Xia, F.; Song, J.; Zhang, L.; Li, J.; et al. Therapeutic effect of *Pseudomonas aeruginosa* phage YH30 on mink hemorrhagic pneumonia. *Vet. Microbiol.* **2016**, *190*, 5–11.
51. Yang, M.; Du, C.; Gong, P.; Xia, F.; Sun, C.; Feng, X.; Lei, L.; Song, J.; Zhang, L.; Wang, B.; et al. Therapeutic effect of the YH6 phage in a murine hemorrhagic pneumonia model. *Res. Microbiol.* **2015**, *166*, 633–643.
52. Amgarten, D.; Martins, L.F.; Lombardi, K.C.; Principal Antunes, L.; Silva de Souza, A.P.; Gonçalves Nicastrro, G.; Kitajima, E.W.; Quaggio, R.B.; Upton, C.; Setubal, J.C.; et al. Three novel *Pseudomonas* phages isolated from composting provide insights into the evolution and diversity of tailed phages. *BMC Genom.* **2017**, *18*, 346.
53. Chan, J.Z.M.; Millard, A.D.; Mann, N.H.; Schäfer, H. Comparative genomics defines the core genome of the growing N4-like phage genus and identifies N4-like roseophage specific genes. *Front. Microbiol.* **2014**, *5*, 506.
54. Moreno Switt, A.I.; Orsi, R.H.; den Bakker, H.C.; Vongkamjan, K.; Altier, C.; Wiedmann, M. Genomic characterization provides new insight into *Salmonella* phage diversity. *BMC Genom.* **2013**, *14*, 481.
55. Jun, J.W.; Yun, S.K.; Kim, H.J.; Chai, J.Y.; Park, S.C. Characterization and complete genome sequence of a novel N4-like bacteriophage, pSb-1 infecting *Shigella boydii*. *Res. Microbiol.* **2014**, *165*, 671–678.
56. Zhao, Y.; Wang, K.; Jiao, N.; Chen, F. Genome sequences of two novel phages infecting marine roseobacters. *Environ. Microbiol.* **2009**, *11*, 2055–2064.
57. Cubo, M.T.; Alías-Villegas, C.; Balsanelli, E.; Mesa, D.; de Souza, E.; Espuny, M.R. Diversity of *Sinorhizobium (Ensifer) meliloti* bacteriophages in the rhizosphere of *Medicago marina*: myoviruses, filamentous and N4-like podovirus. *Front. Microbiol.* **2020**, *11*, 22.
58. Hayden, A.; Martinez, N.; Moreland, R.; Liu, M.; Gonzalez, C.F.; Gill, J.J.; Ramsey, J. Complete genome sequence of *Stenotrophomonas* phage pokken. *Microbiol. Resour. Announc.* **2019**, *8*, e01095–19.
59. Ankrah, N.Y.D.; Budinoff, C.R.; Wilson, W.H.; Wilhelm, S.W.; Buchan, A. Genome sequence of the sulfitebacter sp. strain 2047–infecting lytic phage Φ CB2047–B. *Genome Announc.* **2014**, *2*, e00945–13.
60. Fouts, D.E.; Klumpp, J.; Bishop-Lilly, K.A.; Rajavel, M.; Willner, K.M.; Butani, A.; Henry, M.; Biswas, B.; Li, M.; Albert, M.J.; et al. Whole genome sequencing and comparative genomic analyses of two *Vibrio cholerae* O139 Bengal-specific podoviruses to other N4-like phages reveal extensive genetic diversity. *Virol. J.* **2013**, *10*, 165.
61. Naser, I.B.; Hoque, M.M.; Abdullah, A.; Bari, N.S.M.; Ghosh, A.N.; Faruque, S.M. Environmental bacteriophages active on biofilms and planktonic forms of toxigenic *Vibrio cholerae*: Potential relevance in cholera epidemiology. *PLoS ONE* **2017**, *12*, e0180838.
62. Bhandare, S.; Colom, J.; Baig, A.; Ritchie, J.M.; Bukhari, H.; Shah, M.A.; Sarkar, B.L.; Su, J.; Wren, B.; Barrow, P.; et al. Reviving phage therapy for the treatment of cholera. *J. Infect. Dis.* **2019**, *219*, 786–794.
63. Katharios, P.; Kalatzis, P.G.; Kokkari, C.; Sarropoulou, E.; Middelboe, M. Isolation and characterization of a N4-like lytic bacteriophage infecting *Vibrio splendidus*, a pathogen of fish and bivalves. *PLoS ONE* **2017**, *12*, e0190083.

64. Kim, H.J.; Jun, J.W.; Giri, S.S.; Chi, C.; Yun, S.; Kim, S.G.; Kim, S.W.; Kang, J.W.; Han, S.J.; Park, S.C. Complete Genome Sequence of a Bacteriophage, pVco-5, That infects *Vibrio coralliilyticus*, which causes bacillary necrosis in Pacific Oyster (*Crassostrea gigas*) Larvae. *Genome Announc.* **2018**, *6*, e01143–17.
65. Miller, M.; Deiulio, A.; Holland, C.; Douthitt, C.; McMahon, J.; Wiersma-Koch, H.; Turechek, W.W.; D'Elia, T. Complete genome sequence of Xanthomonas phage RiverRider, a novel N4-like bacteriophage that infects the strawberry pathogen *Xanthomonas fragariae*. *Arch. Virol.* **2020**, *165*, 1481–1484.
66. Nishimura, Y.; Yoshida, T.; Kuronishi, M.; Uehara, H.; Ogata, H.; Goto, S. ViPTree: The viral proteomic tree server. *Bioinformatics* **2017**, *33*, 2379–2380.
67. Rohwer, F.; Edwards, R. The phage proteomic tree: A genome-based taxonomy for phage. *J. Bacteriol.* **2002**, *184*, 4529–4535.
68. Walker, P.J.; Siddell, S.G.; Lefkowitz, E.J.; Mushegian, A.R.; Adriaenssens, E.M.; Dempsey, D.M.; Dutilh, B.E.; Harrach, B.; Harrison, R.L.; Hendrickson, R.C.; et al. Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses. *Arch. Virol.* **2020**.
69. Barylski, J.; Enault, F.; Dutilh, B.E.; Schuller, M.B.P.; Edwards, R.A.; Gillis, A.; Klumpp, J.; Knezevic, P.; Krupovic, M.; Kuhn, J.H.; et al. Analysis of spounaviruses as a case study for the overdue reclassification of tailed phages. *Syst. Biol.* **2020**, *69*, 110–123.
70. Preprint: Moraru, C.; Varsani, A.; Kropinski, A.M. VIRIDIC—a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. **2020** bioRxiv doi:10.1101/2020.07.05.188268.
71. Michniewski, S.; Redgwell, T.; Grigonyte, A.; Rihtman, B.; Aguilo-Ferretjans, M.; Christie-Oleza, J.; Jameson, E.; Scanlan, D.J.; Millard, A.D. Riding the wave of genomics to investigate aquatic coliphage diversity and activity. *Env. . Microbiol.* **2019**, *21*, 2112–2128.
72. Bin Jang, H.; Bolduc, B.; Zablocki, O.; Kuhn, J.H.; Roux, S.; Adriaenssens, E.M.; Brister, R.; Kropinski, A.M.; Krupovic, M.; Lavigne, R.; et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **2019**, *37*, 632–639.
73. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069.
74. Lechner, M.; Findeisz, S.; Steiner, L.; Marz, M.; Stadler, P.F.; Prohaska, S.J. Proteinortho: Detection of (co-) orthologs in large-scale analysis. *BMC Bioinform.* **2011**, *12*, 124.
75. Bayliss, S.C.; Thorpe, H.A.; Coyle, N.M.; Sheppard, S.K.; Feil, E.J. PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *Gigascience* **2019**, *8*, giz119.
76. Contreras-Moreira, B.; Vinuesa, P. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Env. . Microbiol.* **2013**, *79*, 7696–7701.
77. Li, L.; Stoeckert, C.J.; Jr Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **2003**, *13*, 2178–2189.
78. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874.

